RG-S6250-48XS8CQ, RG-S6500 Series Switches

S6500_RGOS 12.5(1)B0506

S6500-X86_RGOS 12.5(1)B0506

Configuration Guide

Applicable products: RG-S6510-48VS8CQ, RG-S6250-48XS8CQ, RG-S6510-4C, RG-S6520-64CQ, RG-S6510-32CQ

There may be descriptions available for other products provided only for reference purposes.

Document Version: V1.0
Date: December 13, 2023
Copyright © 2023 Ruijie Networks

Copyright

Copyright © 2023 Ruijie Networks

All rights are reserved in this document and this statement.

Any reproduction, excerption, backup, modification, transmission, translation or commercial use of this document or any portion of this document, in any form or by any means, without the prior written consent of Ruijie Networks is prohibited.

Trademark  and  are owned by Ruijie Networks.

All other trademarks or registered trademarks mentioned in this document are owned by their respective owners.

Disclaimer

The products, services, or features you purchase are subject to commercial contracts and terms. Some or all of the products, services or features described in this document may not be within the scope of your purchase or use. Unless otherwise agreed in the contract, Ruijie Networks does not make any express or implied statement or guarantee for the content of this document.

Due to product version upgrades or other reasons, the content of this document will be updated from time to time. Ruijie Networks reserves the right to modify the content of the document without any notice or prompt.

This manual is for reference only. Ruijie Networks endeavors to ensure content accuracy and will not shoulder any responsibility for losses and damages caused due to content omissions, inaccuracies or errors.

Preface

Intended Audience

This document is intended for:

- Network engineers
- Technical support and servicing engineers
- Network administrators

Technical Support

- Ruijie Networks Website: https://www.ruijienetworks.com/
- Technical Support Website: https://ruijienetworks.com/support
- Case Portal: https://caseportal.ruijienetworks.com
- Community: https://community.ruijienetworks.com
- Technical Support Email: service_rj@ruijienetworks.com
- Live Chat: https://www.ruijienetworks.com/rita

Conventions

1. Conversions

Convention	Description
Bold font	Commands, command options, and keywords are in **bold** font.
Italic font	Arguments for which you supply values are in *italic* font.
[ ]	Elements in square brackets are optional.
{ x | y | z }	Alternative keywords are grouped in braces and separated by vertical bars.
[x | y | z]	Optional alternative keywords are grouped in brackets and separated by vertical bars.
&<1-n>	The argument before the sign (&) can be input for consecutive 1- n times.
//	Double slashes at the beginning of a line of code indicate a comment line.

2. Signs

The signs used in this document are described as follows:

Warning

An alert that calls attention to important rules and information that if not understood or followed can result in data loss or equipment damage.

Caution

An alert that calls attention to essential information that if not understood or followed can result in function failure or performance degradation.

Note

An alert that contains additional or supplementary information that if not understood or followed will not lead to serious consequences.

Specification

An alert that contains a description of product or version support.

3. Note

The manual offers configuration information (including model, port type and command line interface) for indicative purpose only. In case of any discrepancy or inconsistency between the manual and the actual version, the actual version prevails.

RG-S6250-48XS8CQ, RG-S6500 Series Switches

RGOS 12.5(1)B0506

Configuration Guide

---

1. Basic Configuration
2. Alarm Management
3. Virtualization
4. Interface
5. Ethernet Switching
6. IP Service
7. IP Routing
8. Multicast
9. ACL and QoS
10. Security
11. Reliability
12. Network Control and Monitoring
13. Data Center

Basic Configuration

1. Configuring CLI
2. Configuring ZAM
3. Configuring Basic Management
4. Configuring RBAC
5. Configuring Line
6. Configuring File System
7. Configuring USB
8. Configuring Syslog
9. Configuring Software Upgrade
10. Configuring Rboot
11. Configuring Uboot
12. Configuring Time Range
13. Configuring UFT Mode Management
14. Configuring Supervisor Module Redundancy
15. Configuring Hot Swapping
16. Configuring Process Restarting
17. Configuring Python
18. Configuring License Management
19. Configuring Hung Task Checking
20. Configuring TCAM Mode

21. Configuring ECC

Contents

1 Configuring CLI 1

1.1 Introduction 1

1.1.1 Accessing the CLI 1

1.1.2 Command Modes 1

1.1.3 System Help 3

1.1.4 Abbreviated Commands 5

1.1.5 Unset and Reset Options of Commands 5

1.1.6 Prompts for Incorrect Commands 5

1.1.7 History Commands 6

1.1.8 Featured Editing 6

1.1.9 Searching and Filtering of the Display Command Output 7

1.1.10 Command Alias 9

1.1.11 Configuring an Alias for a Command Keyword 10

1.1.12 Character Set Encoding 10

1.1.13 Intelligent Switching of Configuration Modes 10

1.2 Restrictions and Guidelines 11

1.3 Configuration Task Summary 11

1.4 Operating the Python Script 11

1.5 Configuring a Command Alias 12

1.6 Configuring the Privilege Level of a Command 12

1.7 Configuring the Character Set Encoding Format 12

1.8 Disabling Intelligent Switching of Configuration Modes 13

1.9 Configuring an Alias for a Command Keyword 13

1.10 Loading Command Keyword Aliases in Batches 15

1.11 Monitoring 15

1 Configuring CLI

1.1 Introduction

The command line interface (CLI) is a window for text instruction interaction between users and network devices. Users can enter commands in the CLI to configure and manage network devices.

1.1.1 Accessing the CLI

Before using the CLI, you need to connect a terminal or PC to a network device. After you start the network device and initialize the hardware and software of the network device, you can use the CLI. If you use the network device for the first time, you can connect it only through the console port. This is called out-of-band management. After performing relevant configuration, you can connect and manage the network device by telnetting to the virtual terminal.

1.1.2 Command Modes

Since the device has many commands, you can classify them by function to facilitate their use. The CLI provides several commands modes, and all commands are registered in one or several command modes. You must enter the mode of a command before using this command. Command modes are related with each other but differ from each other.

As soon as a new session is set up between a user and the management interface of a network device, you enter user EXEC mode. In this mode, you can use a few commands and the command functions are limited, such as the **display** command. Execution results of commands in user EXEC mode are not saved.

To use more commands, enter privileged EXEC mode. Generally, you must enter a password to enter privileged EXEC mode. In privileged EXEC mode, you can use all commands registered in this command mode, and further enter global configuration mode.

Using commands in configuration modes (such as global configuration mode and interface configuration mode) affects configurations in use. If you save the configurations, these commands are saved and run when the system restarts. To enter another configuration mode, enter global configuration mode. If you perform configuration in global configuration mode, you can enter various configuration sub-modes, such as interface configuration mode.

Table 1-1 Description of the Command Modes (suppose that the name of the network device is "Device")

Command Mode	Access Method	Prompt	Exit or Entering Next Mode	Function Description of the Mode
User EXEC mode	A user enters user EXEC mode by default when accessing a network device.	Device>	● Run the **exit** command to exit this mode. ● Run the **privilege** command to enter privileged EXEC mode.	Use this command mode to conduct basic tests or display system information.
Privileged EXEC mode	In User EXEC mode, run the **privilege** command to enter privileged EXEC mode.	Device#	● Run the **disable** command to return to user EXEC mode. ● Run the **config-system** command to enter global configuration mode.	Use this command mode to check whether the configurations take effect. This mode is password-protected.
Global configuration mode	In Privileged EXEC mode, run the **config-system**command to enter global configuration mode.	Device(cfg)#	● Run the **exit** or **end** command, or press **Ctrl+C** to return to privileged EXEC mode. ● Run the **interface** command to enter interface configuration mode. In the **interface** command, you must specify the interface configuration sub-mode you want to enter. ● Run the **vlan** *vlan-id* command to enter VLAN configuration mode.	Use commands in this mode to configure global parameters that affect the entire network device.

Command Mode	Access Method	Prompt	Exit or Entering Next Mode	Function Description of the Mode
Interface configuration mode	In global configuration mode, run the **interface gigabitethernet 0/1** command to enter interface configuration mode.	Device(cfg-if-gigabitethernet 0/1)#	● Run the **end** command or press **Ctrl+C** to return to privileged EXEC mode. ● Run the **exit** command to return to global configuration mode. In the **interface** command, you must specify the interface configuration sub-mode you want to enter.	Use this command mode to configure various interfaces of the network device.
Config-vlan (VLAN configuration mode)	In global configuration mode, run the **vlan** *vlan-id* command to enter VLAN configuration mode.	Device(cfg-vlan)#	● Run the **end** command or press **Ctrl+C** to return to privileged EXEC mode. ● Run the **exit** command to return to global configuration mode.	Use this command mode to configure VLAN parameters.

1.1.3 System Help

When entering commands in the CLI, you can obtain help information by using the following methods:

- At the command prompt in any mode, enter a question mark (?) to list the commands supported in current command mode and command description.

```
Device> ?
Exec commands:
  <1-99>          Session number to resume
  disable         Turn off privileged cmd
  disconnect      Disconnect an existing network connection
  privilege       Turn on privileged commands
  exit            Exit from the EXEC
  help            Description of the interactive help system
  lock            Lock the terminal
  ping            Send echo messages
  display         Show running system information
  telnet          Open a telnet connection
  traceroute      Trace route to destination
```

- Enter a space and a question mark (?) after a keyword of a command to list the next keyword or variable

associated with the keyword.

```
Device(cfg)# interface ?
  Aggregateport       Aggregate port interface
  Dialer              Dialer interface
  GigabitEthernet     Gigabit Ethernet interface
  Loopback            Loopback interface
  Multilink           Multilink-group interface
  Null                Null interface
  Tunnel              Tunnel interface
  Virtual-ppp         Virtual PPP interface
  Virtual-template    Virtual Template interface
  Vlan                Vlan interface
  range               Interface range command
```

Note

If a keyword is followed by a parameter value, the value range and description of this parameter are displayed as follows:

```
Device(cfg)# interface vlan ?
  <1-4094>  Vlan port number
```

- Enter a question mark (?) after an incomplete string of a command keyword to list all command keywords starting with the string.

```
Device# d?
debugging  delete  diagnostic  dir  disable  remove
```

- After an incomplete command keyword is entered, if the suffix of this keyword is unique, press the **Tab** key to display the complete keyword.

```
Device# display conf<Tab>
Device# display configuration
```

- In any command mode, you can run the **help** command to obtain brief description of the help system.

```
Device(cfg)# help
Help may be requested at any point in a command by entering
a question mark '?'. If nothing matches, the help list will
be empty and you must backup until entering a '?' displays the
available options.
Two styles of help are provided:
1. Full help is available when you are ready to enter a
   command argument (e.g. 'display ?') and describes each possible
   argument.
2. Partial help is provided when an abbreviated argument is entered
   and you want to know what arguments match the type
   (e.g. 'display pr?'.)
```

1.1.4 Abbreviated Commands

If a command is long, you can enter a part of the command keyword characters that are used to identify the command keyword.

For example, the **interface gigabitethernet 0/1** command for entering the configuration mode of interface GigabitEthernet 0/1 is abbreviated as follows:

```
Device(cfg)# int g0/1
Device(cfg-if-GigabitEthernet 0/1)#
```

1.1.5 Unset and Reset Options of Commands

Most commands have the **unset** option. Generally, the **unset** option is used to disable a feature or function, or perform an operation opposite to a command. For example, run the **unset shutdown** command on an interface to perform the operation opposite to the **shutdown** command, that is, enable the interface. The keyword without the **unset** option is used to enable a disabled feature or a feature that is disabled by default.

Most configuration commands have the **cmd-reset** option. The **cmd-reset** option is used to restore settings of a command to default values. Default values of most commands are used to disable this function. Therefore, in most cases, the function of the **cmd-reset** option is the same as that of the **unset** option. For some commands, however, the default values are used to enable this function. In this case, the function of the **cmd-reset** option is opposite to that of the **unset** option. At this time, the **cmd-reset** option is used to enable the function of this command and set the variables to default values.

Note

For specific function of the **unset** or **cmd-reset** option of each command, see the command reference.

1.1.6 Prompts for Incorrect Commands

If an incorrect command is entered, an error prompt is displayed.

Table 1-2 Common CLI Errors

Error	Description	How to Obtain Help
% Ambiguous command: "display c"	The characters entered are too few to identify a unique command.	Enter the command again, and enter a question mark (?) after the ambiguous word. All the possible keywords are displayed.
% Incomplete command.	The mandatory keyword or variable is not entered in the command.	Enter the command again, and enter a space and a question mark (?). All the possible keywords or variables are displayed.
% Invalid input detected at '^' marker.	An incorrect command is entered. The symbol (^) indicates the position of the wrong word.	At the current command mode prompt, enter a question mark (?). All the command keywords allowed in this command mode are displayed.

1.1.7 History Commands

The system automatically saves commands that are entered recently. You can use shortcut keys to query or call history commands.

Table 1-3 Operation Methods

Operation	Result
Ctrl+P or the Up key	The previous command in the history command list is displayed. Starting from the latest record, you can repeatedly perform this operation to query earlier records.
Ctrl+N or the Down key	After pressing **Ctrl+P** or the Up key, you can perform this operation to return to a command that is more recently executed in the history command list. You can repeatedly perform this operation to query more recently executed commands.

Specification

The standard terminals, such as the VT100 series, support the direction keys.

1.1.8 Featured Editing

When editing commands, you can use the keys or shortcut keys listed in the following table:

Table 1-4 Description of Shortcut Keys

Function	Key or Shortcut Key	Description
Move the cursor in a line.	Left key or **Ctrl+B**	Move the cursor to the left by one character.
	Right key or **Ctrl+F**	Move the cursor to the right by one character.
	Ctrl+A	Move the cursor to the head of the command line.
	Ctrl+E	Move the cursor to the end of the command line.
Delete an entered character.	Backspace key	Delete one character to the left of the cursor.
	Delete key	Delete one character to the right of the cursor.
	Ctrl+K	Delete all characters from the cursor to the end of the command line.
	Ctrl+W	Delete all characters between the cursor and the first space on the left of cursor.
	Ctrl+U	Delete all characters from the cursor to the head of the command line.

Execute the command	**Return**, **Ctrl+M** **Ctrl+J**	After entering the complete command, press the key to execute the command and start a new line.
	Ctrl+Z	After entering the complete command, press the key to execute the command and return to the EXEC mode.
Cancel command execution	**Ctrl+C**	After entering the complete command, press the key to cancel command execution and return to the EXEC mode.
Special characters	**Ctrl+V**	After pressing the key, special characters such as ? and TAB can be input as the special character string.
Move the output by one line or one page.	Enter key	When the output is displayed, press the **Enter** key to move the output one line upward and display the next line. Use this key only when the output does not end yet.
	Space key	When the output is displayed, press the **Space** key to page down and display the next page. Only when the output does not end, use this key.

When the cursor is close to the right boundary, the entire command line moves to the left by 20 characters, and the hidden front part is replaced with a dollar ($) sign. You can use related keys or shortcut keys to move the cursor to the characters in the front or return to the head of the command line.

For example, the entire **access-list** command may be wider than the screen. When the cursor is first close to the end of the command line, the entire command line moves to the left by 20 characters, and the hidden front part is replaced with a dollar sign ($). Each time the cursor is close to the right boundary, the entire command line moves to the left by 20 characters.

```
access-list 199 permit ip host 192.168.180.220 host
$ost 192.168.180.220 host 202.101.99.12
$0.220 host 202.101.99.12 timerange tr
```

Press **Ctrl+A** to return to the head of the command line. At this time, the hidden tail part of the command line is replaced with a dollar sign ($).

```
access-list 199 permit ip host 192.168.180.220 host 202.101.99.$
```

Note

The default line width of terminals is 80 characters.

1.1.9 Searching and Filtering of the Display Command Output

- To search specified content in the output of the **display** command, run the following command:

Table 1-5 Searching for Specified Content in the Output of the display Command

Command	Purpose
display any-command | [**regexp**] **begin** regular-expression	Searches for specified content in the output of the **display** command, and outputs the first line containing the content and all information that follows this line.

- To filter specified content in the output of the **display** command, run the following commands:

Table 1-6 Filtering Specified Content in the Output of the display Command

Command	Purpose
display any-command | [**regexp**] **exclude** regular-expression	Filters the output of the **display** command to filter out lines containing the specified content, and outputs other information.
display *any-command* | [**regexp**] **include** *regular-expression*	Filters the output of the **display** command to output only the lines containing the specified content, and filters out other information.

Note
- The **display** command can be executed in any mode.
- Searched content is case-sensitive.

When the **regexp** keyword is specified, a regular expression can be used for filtering content. The following table describes the usages of special characters in a regular expression.

Table 1-7 Description of Usages of Special Characters in a Regular Expression

Character	Symbol	Special Meaning
Period	.	Matches any single character.
Plus sign	+	Matches one or any sequence in a string.
Caret	^	Matches the start of a string.
Underline	_	Matches commas, brackets, start and end of a string, and spaces.
Square brackets	[ ]	Matches a single character within a range.

To search or filter the output of the **display** command, you must enter a vertical line (|). After the vertical line, specify the search or filter rules and content to be searched or filtered (characters or strings). Content to be searched and filtered is case-sensitive.

```
Device# display current-config | include interface
interface GigabitEthernet 0/0
interface GigabitEthernet 0/1
interface GigabitEthernet 0/2
interface GigabitEthernet 0/3
interface GigabitEthernet 0/4
interface GigabitEthernet 0/5
interface GigabitEthernet 0/6
interface GigabitEthernet 0/7
interface Mgmt 0
Device# display current-config | regexp include GigabitEthernet [0-9]/1
interface GigabitEthernet 0/1
```

1.1.10 Command Alias

You can specify any word as the alias of a command to simplify the input of the command string.

1. Default aliases in the system

In configuration mode or privileged EXEC mode, default aliases are available for some commands. You can run the **display aliases-cmd** command to display these default aliases.

```
Device(cfg)# display aliases-cmd
Exec mode alias-cmd:
  h                    help
  p                    ping
  s                    display
  u                    undebugging
  un                   undebugging
```

Note

Default aliases in the system cannot be deleted.

2. System help regarding command aliases

- The system provides help information for command aliases. An asterisk (*) is displayed in front of an alias in the following format:

  ```
  *command-alias=original-command
  ```

  For example, in privileged EXEC mode, the default command alias "s" represents the keyword **display**. If you enter "s?", help information of the keywords and aliases starting with "s" is displayed.

  ```
  Device# s?
  *s=display    display    start-chat    start-terminal-service
  ```

- If the command represented by an alias contains more than one word, the command is displayed in a pair of

quotation marks.

For example, in privileged EXEC mode, configure the alias "sv" to replace the **display version** command. If you enter "s?", help information of the keywords and aliases starting with "s" is displayed.

```
Device# s?
*s=display    *sv=”display version” display    start-chat
start-terminal-service
```

- You can use an alias to obtain help information of the command represented by the alias.

  For example, if you configure the alias "ia" in interface configuration mode to represent the **ip address** command, you can enter "ia?" in interface configuration mode to obtain help information of "**ip address**?", and replace the alias with the actual command.

```
Device(cfg-if)# ia ?
  A.B.C.D   IP address
  dhcp      IP Address via DHCP
Device(cfg-if)#ip address
```

Caution

If you enter a space in front of a command alias, the command represented by this alias is displayed.

1.1.11 Configuring an Alias for a Command Keyword

You can specify an alias to replace the command keyword on the device to make the command close to your usage habits.

1. Command Keyword Alias

After a keyword alias is specified, the original command is not available. When the original command is executed, an error message is displayed.

2. Alias in the Display Command Output

After a keyword alias is specified, the configuration in the **display current-config** command output is replaced by the alias.

1.1.12 Character Set Encoding

The function of character set encoding enables the device to use a unified character set encoding format by configuring character set encoding. After a command is entered in the CLI of a terminal, the command is automatically converted into a command in the unified character set encoding format before delivery.

Note

When hybrid formats exist in current running configurations, you must manually delete running configurations containing the encoding format different from the target format before modifying the character set encoding format.

1.1.13 Intelligent Switching of Configuration Modes

After intelligent switching of configuration modes is enabled, if a command to be executed does not exist in the current configuration mode, the system searches for this command in the upper-level configuration mode. If the

command exists, the system automatically switches to the upper-level configuration mode and executes the command. Intelligent switching to the upper-level configuration mode or to a further upper-level configuration mode next to the upper-level configuration mode is supported in some configuration modes.

For example, when you run a command in the G0/1 interface configuration mode for entering the G0/2 interface configuration mode, the system directly enters the G0/2 interface configuration mode:

```
Device> privilege
Device# config-system terminaltermianl
Device(cfg)# interface gigabitethernet0/1
Device(cfg-if-GigabitEthernet 0/1)#interface gigabitethernet 0/2
Device(cfg-if-GigabitEthernet 0/2)#
```

1.2 Restrictions and Guidelines

You can adopt UTF-8/GBK to perform configuration in Chinese. When you use the cursor to process words or delete **Backspace** in the CLI, the interface may display errors or garbled characters. You are advised to delete all current characters and then configure a unified character set encoding format. Upon word processing or Backspace deletion, you can run the **display current-config** command to check whether the configurations are correct. (To delete a Chinese character, you must press the **Backspace** key twice in the case of GBK, but must press the **Backspace** key three times in the case of UTF-8.)

1.3 Configuration Task Summary

CLI configuration includes the following tasks: All the configuration tasks below are optional. Select the configuration tasks as required.

- Operating the Python Script
- Configuring a Command Alias
- Configuring the Privilege Level of a Command
- Configuring the Character Set Encoding Format
- Disabling Intelligent Switching of Configuration Modes

1.4 Operating the Python Script

1. Overview

This section describes how to load and unload the Python script of CLI.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Operate the Python script. The configuration steps below are mutually exclusive. Select one of them for configuration.

- Load the Python script.

  cli-python insmod *python-filename*

- Unload the Python script.

cli-python rmmod *python-filename*

1.5 Configuring a Command Alias

1. Overview

You can specify any word as the alias of a command to simplify the input of the command string.

2. Restrictions and Guidelines

- The command replaced with an alias must start from the first character of the command line.
- The command replaced with an alias must be complete.
- An alias must be entered in full; otherwise, it cannot be identified.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an alias to replace the front part of a command.

alias-cmd *mode command-alias original-command*

Default aliases are available for some commands in global configuration mode or privileged EXEC mode by default.

1.6 Configuring the Privilege Level of a Command

1. Overview

This section describes how to configure the privilege level of a command.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the privilege level of a command.

privilege-cmd *mode* [**all-cmd**] { **cmd-level** *level* | **cmd-reset** } *command-string*

1.7 Configuring the Character Set Encoding Format

1. Overview

This section describes how to configure a unified character set encoding format for the device.

2. Restrictions and Guidelines

- When hybrid formats exist in current running configurations of the device, you must manually delete running

configurations containing the encoding format different from the target format before modifying the character set encoding format.

- Garbled characters may be displayed if the current encoding format configured for the device differs from that used by terminals.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the character set encoding format.

language character-set { **cmd-reset** | **gbk** | **utf-8** }

By default, hybrid formats are supported.

1.8 Disabling Intelligent Switching of Configuration Modes

1. Overview

Disabling intelligent switching of configuration modes on the device ensures that a command is run only when it is supported in the current configuration mode.

2. Restrictions and Guidelines

Intelligent switching of configuration modes is enabled by default.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Disable intelligent switching of configuration modes.
unset automatic-switching-mode

Intelligent switching of configuration modes is enabled by default.

1.9 Configuring an Alias for a Command Keyword

1. Overview

You can specify an alias to replace the command keyword on the device to make the command close to your usage habits.

2. Restrictions and Guidelines

- After the command alias is configured, the performance of viewing, saving, and restoring configurations is slightly deteriorated.
- After the command alias is configured, the command permission changed by the **privilege** command needs to be reconfigured using the new command.

- If the original command is displayed in the output of some commands, the original command is not replaced with the alias.
- Only the output of the **display current-config** command (running the **display current-config** { *interface* | *router-protocol* } command) is replaced with the alias. The output of other commands is not replaced.
- After a command is aliased, the parameter string entered in the command cannot be the same as the original command. For example, the command **username** *user* **password** *pass* is aliased as **user** *user* **pass** *pass*. In this case, the **user** *password* **pass** password is delivered, the command will be displayed as **user** *pass* **password** *password*.
- A command keyword may be a common parent keyword of multiple subcommand keywords. When an alias is configured, the parent keywords of all subcommands are aliased. For example, the **logging service** command is aliased as **syslog service**. Since the **logging** command is also the parent keyword of the **file** keyword in the **logging file** command, the **logging file** command is also aliased as **syslog file**.
- A command keyword may be registered by multiple function modules. To completely replace such keywords, you need to configure an alias without the subcommand keyword. For example, **ip** in the **ip dhcp pool** command is a common keyword of many function modules. When you use the **alias keyword** *config ip ipv4 dhcp dhcp-server* command to change the **ip dhcp pool alias** to **ipv4 dhcp-server pool**, the **ip dhcp** command is changed to **ipv4 dhcp-server**. Therefore, only the commands containing **ip dhcp** are changed.
- To change all **ip** commands to **ipv4** commands, run the **alias keyword** *config ip ipv4* command.
- The parameters cannot be aliased. In the alias command output, the content enclosed in <> brackets is a parameter. Some keywords follow parameters. After selecting the parameter value range, you can continue to select subsequent command keywords. For example, to alias the **username** *username-msg* **password** *password-msg* command to **user-test** *username-msg* **pass-test** *password-msg*, run the **alias keyword** *config username user-test <WORD> password pass-test* command, where <*WORD*> indicates the content of the *username-msg* parameter.
- Avoid the possible ambiguity between commands. If there is ambiguity, you need to redefine the aliases. The following example shows a case of ambiguous command. The device has the **username** *user* command. When the **display current-config** command is aliased as **username running-config**, the device has the **username** *user* and **username running-config** commands. When the **username running-config** command is entered, the device cannot identify which command to execute. Therefore, the command execution may be abnormal, failing the customer's expectation.
- One alias configuration corresponds to one **display current-config** content replacement. For the common command keyword mentioned above, running the **display current-config** command may still display the original command. In this case, the alias configuration of the corresponding command is required to display the command normally. For example, replace the **redistribute-route ospf** *number* **metric** command with r**edistribute-route ospf** *number* **metric-type**. When you run the command, the prompt is **metric-type**. However, the output of the **display current-config** command is not replaced by the alias. If you need to replace the configuration, you need to configure an alias that replaces the **redistribute-route bgp** *number* **metric** command with **redistribute-route bgp** *number* **metric-type**.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an alias for a command keyword.

alias keyword *mode original-command1 command-alias1 ...* [*original-command command-alias*]

No command keyword alias is configured by default.

1.10 Loading Command Keyword Aliases in Batches

1. Overview

You can load command keyword aliases in batches.

2. Restrictions and Guidelines

- The encoding format of the configuration file is UTF-8.
- The configuration file contains the keyword alias configuration in the required format.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Load command keyword aliases in batches.

load alias keyword { **flash**: **| tmp**: **|** *filename* }

1.11 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Table 1-8 Monitoring of the CLI

Command	Purpose
display aliases-cmd [*mode*]	Displays all command aliases or the command aliases in specific command mode.
display aliases keyword {**all** [*config*] | *mode_name*}	**D**isplays all keyword aliases or the keyword aliases in a specific command mode.

Contents

1 Configuring ZAM 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Restrictions and Guidelines 2
1.3 Configuring ZAM 2
1.3.1 Overview 2
1.3.2 Restrictions and Guidelines 2
1.3.3 Prerequisites 2
1.3.4 Procedure 3
1.4 Monitoring 4
1.5 Configuration Examples 4
1.5.1 Configuring ZAM 4
2 Configuring Auto-provision 8
2.1 Introduction 8
2.1.1 Overview 8
2.1.2 Principles 8
2.1.3 Protocols and Standards 8
2.2 Restrictions and Guidelines 8
2.3 Configuring Auto-provision 9
2.3.1 Overview 9
2.3.2 Procedure 9

2.4 Monitoring 9

2.5 Configuration Examples..... 10

2.5.1 Configuring the Auto-provision Function..... 10

1 Configuring ZAM

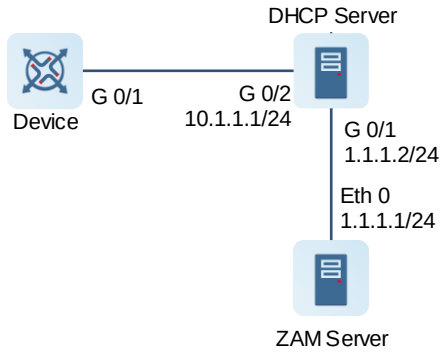
1.1 Introduction

1.1.1 Overview

The zero automatic manager (ZAM) function enables the device to automatically download the software version, upgrade the device version, and apply configuration files when the device has no configuration and the network administrator is not in contact with the field devices. ZAM simplifies the operation, maintenance, and deployment of massive devices, lowers the labor cost, reduces human errors, and improves the efficiency of network deployment.

1.1.2 Principles

Figure 1-1 ZAM Automatic Go-Online Solution



ZAM go-online and deployment include four stages as follows:

(1) Initialization

At this stage, a device that goes online without configurations connects to a network and completes initialization. The device automatically pre-deploys the ZAM environment. During the pre-deployment, the MGMT port or service port must be used as the ZAM interface. The device automatically performs configuration, without any need of extra configuration by the administrator.

(2) DHCP

Upon the pre-deployment above, the ZAM interface requests an IP address from the Dynamic Host Configuration Protocol (DHCP) server via the DHCP protocol. In a reply packet, the reply packet carries Option 66 (containing the IP address of the file server) and Option 67 (containing the name of the Python script to be downloaded). The requirements are as follows:

- The ZAM interface supports DHCP.
- Configure the ZAM interface with a DHCP-allocated IP address. DHCP requests support Option 66 and Option 67, and Option 66 and Option 67 have been configured in the address pool of the DHCP server.
- The device analyzes and deploys the IP address of the ZAM interface, and analyzes the content of Option 66 and Option 67.

(3) TFTP

Based on the file name of the Python script and the IP address of the ZAM server (on which the TFTP server function will be deployed) obtained during DHCP, the device downloads the corresponding configuration script from the ZAM server.

Upon downloading above, the device executes the configuration script and downloads the corresponding configuration file or version from the ZAM server.

(4) Configuration loading

After the Python script is executed, the device restarts automatically to update the device version and load the configuration file.

1.2 Restrictions and Guidelines

- An automatic retry is available after a ZAM failure. In this case, check whether the network functions properly and whether the DHCP server is correctly configured based on the ZAM failure prompt.
- For a product version with ZAM disabled by default, please manually enable this function.

1.3 Configuring ZAM

1.3.1 Overview

This section describes how to configure ZAM to enable a device with no configurations to implement automatic deployment.

1.3.2 Restrictions and Guidelines

To use the ZAM function, deploy the ZAMserver and DHCP server.

1.3.3 Prerequisites

- Configure the address pool, Option 66, and Option 67 for the DHCP server. The configuration example is as follows:

```
ip dhcp pool poap
 option 67 ascii zam.py
 option 66 ascii 100.1.1.2
```

- Configure the ZAM server. Store the Python script file in the root directory of the TFTP server, and create the folders listed in Table 1-1 in the root directory to store the files required during the ZAM configuration.

Table 1-1 Description of Folders to Be Created on the ZAM Server

Folder Name	Content	Function
POAP_CFG	● xxxx.cfg ● xxxx.params	● xxxx.cfg: Indicates the configuration file of the device. The configuration file of each device is named after the sn value of the device and uses .cfg as the suffix. ● xxxx.params: Indicates the parameter file of the device. It contains the path to the downloaded configuration file, path to the file used for version upgrade, and the name of the version file. ● The device automatically downloads the files.
POAP_IMAGE	Version binary file	● The folder stores the version file of the device. ● The device automatically downloads the file.
POAP_LOG	Log file	● The folder stores the log files uploaded during the ZAM configuration of the device. The log files of each device are named after the sn value of the device. ● The device automatically uploads the files.
POAP_STARTUP	Device sn value files	● The folder stores the sn value file of the device where ZAM needs to be configured. ● The device automatically uploads the file.
POAP_STATUS	-	● The folder stores the ZAM result log files, which are named after the sn value. ● The device automatically uploads the files.

1.3.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable ZAM.

zam

The ZAM function is enabled by default.

1.4 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-2 ZAM Monitoring

Command	Purpose
display zam	Displays the current configurations and status of ZAM.
debugging zam	Enables the ZAM framework event debugging.

1.5 Configuration Examples

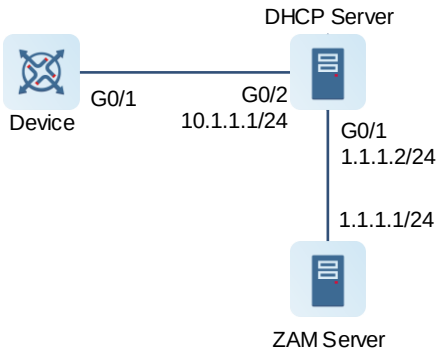
1.5.1 Configuring ZAM

1. Requirements

Manage and control onboarding and deployment of all devices in the network in a unified manner. Realize the unified management of onboarding and deployment of all devices.

2. Topology

Figure 1-2 Configuring ZAM



3. Notes

- Configure the DHCP server.
- Configure the ZAM server.
- Enable the ZAM function on the devices.

4. Procedure

(1) Configure the DHCP server.

Set the IP address of interface GigabitEthernet 0/1 of the DHCP server to 1.1.1.2/24.

```
Dhcp_Server> privilege
Dhcp_Server# config-system terminal
Dhcp_Server(cfg)#interface gigabitEthernet 0/1
Dhcp_Server(cfg-if-GigabitEthernet 0/1)#unset switch-port
Dhcp_Server(cfg-if-GigabitEthernet 0/1)# ip address 1.1.1.2 255.255.255.0
```

Set the IP address of interface GigabitEthernet 0/2 of the DHCP server to 10.1.1.1/24.

```
Dhcp_Server(cfg)#interface gigabitEthernet 0/2
Dhcp_Server(cfg-if-GigabitEthernet 0/2)#unset switch-port
Dhcp_Server(cfg-if-GigabitEthernet 0/2)# ip address 10.1.1.1 255.255.255.0
```

Enable the DHCP server function.

```
Dhcp_Server(cfg)# service dhcp
```

Configure the address pool of the DHCP server.

```
Dhcp_Server(cfg)# ip dhcp pool zam
Dhcp_Server(cfg-dhcp-pool)# network 10.1.1.1 255.255.255.0
Dhcp_Server(cfg-dhcp-pool)# default-router 10.1.1.1
```

Configure Option 66.

```
Dhcp_Server(cfg-dhcp-pool)# option 66 ascii 1.1.1.1
```

Configure Option 67.

```
Dhcp_Server(cfg-dhcp-pool)# option 67 ascii zam.py
```

(2) Configure the ZAM server. Store the intermediate file (Python script named **zam.py**), and the configuration files and version files to be delivered to the devices on the ZAM server. For details, see 1.3.3 Prerequisites.

(3) Enable the ZAM function on the devices.

Delete the configuration files from the devices.

```
Device# delete config.text
Do you want to delete [flash:/config.text]? [Y/N]:y
Delete success.
```

Run the **reboot** command to restart the devices to enable the ZAM function.

```
*May 13 15:47:36: %ZAM-6-EVENT: Get tftp server ip 1.1.1.1.
*May 13 15:47:36: %IPMIX-6-ADDRESS_ASSIGN: Interface VLAN 1 assigned DHCP address 10.1.1.2, mask 255.255.255.0.
*May 13 15:47:36: %ZAM-6-EVENT: Get boot file name zam.py.
*May 13 15:47:42: %ZAM-6-EVENT: Download script success, rename to zam.py.
*May 13 15:47:42: %ZAM-6-INFO: Begin to exec-prompt script zam.py.
Thu May 13 15:47:46 2021 INFO: script(version:1.0.12440) running...
Thu May 13 15:47:46 2021 DEBUG: server ip: 1.1.1.1, port type: SVI
Thu May 13 15:47:46 2021 DEBUG: current version: S5700H_RGOS12.5(master)
Thu May 13 15:47:46 2021 DEBUG: current patch: NA
```

```
Thu May 13 15:47:47 2021 DEBUG: execute upload command: copy flash:/G1NQ7UW700483.POAP
tftp://1.1.1.1/POAP_STARTUP/G1NQ7UW700483.POAP...
Thu May 13 15:47:47 2021 DEBUG: upload G1NQ7UW700483.POAP success
Thu May 13 15:47:47 2021 DEBUG: begin to download G1NQ7UW700483.params ...
Thu May 13 15:47:47 2021 DEBUG: execute download command: copy
tftp://1.1.1.1/POAP_CFG/G1NQ7UW700483.params flash:/poap.tmp.params...
Thu May 13 15:47:48 2021 DEBUG: download G1NQ7UW700483.params success
Thu May 13 15:47:48 2021 DEBUG: parse image_name=57H_5PL4_0329.bin
Thu May 13 15:47:48 2021 DEBUG: begin to download G1NQ7UW700483.cfg ...
Thu May 13 15:47:48 2021 DEBUG: execute download command: copy
tftp://1.1.1.1/POAP_CFG/G1NQ7UW700483.cfg flash:/poap.tmp.cfg...
Thu May 13 15:47:48 2021 DEBUG: download G1NQ7UW700483.cfg success
Thu May 13 15:47:48 2021 INFO: download G1NQ7UW700483.cfg success, save to flash:/poap.tmp.cfg
Thu May 13 15:47:48 2021 INFO: begin to download 57H_5PL4_0329.bin ...
Thu May 13 15:47:48 2021 DEBUG: execute download command: copy
tftp://1.1.1.1/POAP_IMAGE/57H_5PL4_0329.bin tmp:/poap.image.bin...
Thu May 13 15:54:20 2021 DEBUG: download 57H_5PL4_0329.bin success
Thu May 13 15:54:20 2021 INFO: download 57H_5PL4_0329.bin success, save to tmp:/poap.image.bin
Thu May 13 15:54:20 2021 INFO: upgrade system begin...
< The terminal is lock-session >
Upgrade start
!!!!100%
Upgrade skip, reason: The version in device is the same
< The terminal is unlock >
Thu May 13 15:54:43 2021 INFO: upgrade status: skip
Thu May 13 15:54:53 2021 DEBUG: rename file poap.tmp.cfg to config.text
Thu May 13 15:54:53 2021 INFO: script ending...
Thu May 13 15:54:54 2021 DEBUG: execute upload command: copy flash:/poap.tmp.log
tftp://1.1.1.1/POAP_LOG/G1NQ7UW700483.LOG...
Thu May 13 15:54:55 2021 DEBUG: upload G1NQ7UW700483.LOG success
Thu May 13 15:54:56 2021 DEBUG: execute upload command: copy flash:/G1NQ7UW700483.ok
tftp://1.1.1.1/POAP_STATUS/G1NQ7UW700483.ok...
Thu May 13 15:54:56 2021 DEBUG: upload G1NQ7UW700483.ok success
*May 13 15:55:00: %ZAM-6-EVENT: Script exec-prompt success, request device reset now.
*May 13 15:55:00: %DP-3-RESET_DEV: Reset device 1 due to zam reboot device.
*May 13 15:55:02: %ZAM-6-EVENT: Script exec-prompt success, request device reset now.
*May 13 15:55:02: %DP-3-MACHINE_RESTART: MACHINE_RESTART.
```

5. Verification

Run the **display dhcp lease** command to display the obtained IP address.

```
Device# display dhcp lease
Temp IP addr: 0.0.0.0    for slave on Interface: Mgmt 0
Temp    sub net mask: 0.0.0.0
     DHCP Lease server: 0.0.0.0, state: 3 Init-proc
     Retry count: 7      Client-ID: 01800588aba05e4D676D7430
Temp IP addr: 10.1.1.2    for slave on Interface: VLAN 1
```

```
Temp    sub net mask: 255.255.255.0
     DHCP Lease server: 10.1.1.1, state: 9 Inform-proc
     DHCP transaction id: 75cf492b
     Lease: 86400 secs,    Renewal: 43200 secs,    Rebind: 75600 secs
     Next timer fires after: 43054 secs
     Retry count: 0      Client-ID: 01800588aba05d564C414E31
```

Run the **display zam** command to check the execution status of the ZAM function.

```
Device# display zam
ZAM state                  :enable
ZAM status                 :Now is waiting for do shell exec-prompt
Server ip                 :1.1.1.1
Usb path                   :NULL
Interface name            :VLAN 1
Interface type           :SVI
Succ Interface name     :VLAN 1
Script URL                 :tftp://1.1.1.1/zam.py
```

6. Configuration Files

- DHCP server configuration file

```
hostname Dhcp_Server
!
ip dhcp pool zam
 option 66 ascii 1.1.1.1
 option 67 ascii zam.py
 network 10.1.1.0 255.255.255.0
 default-router 10.1.1.1
!
end
```

7. Common Errors

- The network connection between the device and the ZAM server is abnormal.
- The configuration file on the device is not deleted.

2 Configuring Auto-provision

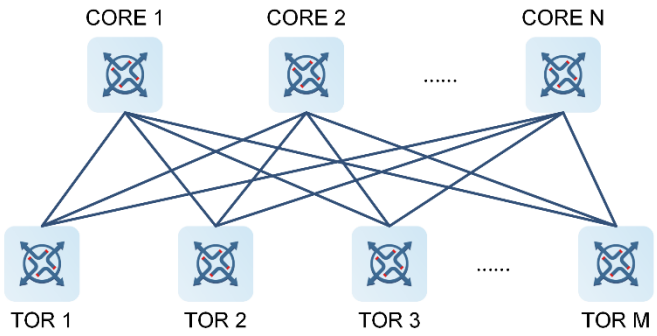
2.1 Introduction

2.1.1 Overview

In traditional data centers, the initial construction of network devices and the replacement of faulty devices heavily rely on manual participation, which poses very high requirements for participants' professional skills. Moreover, human errors are prone to occur, leading to network failures and other problems. In view of these, Ruijie proposes an efficient and fully automatic Auto-provision solution to achieve zero-configuration deployment of devices.

2.1.2 Principles

Figure 2-1 Auto-provision Topology



As shown in Figure 2-1, in access scenarios with a large number of top of rack (ToR) switches, the Auto-provision function is enabled by default and does not need to be configured.

- Each ToR switch obtains the port extension description from one core switch, and identifies its own IP address and configuration file name.
- ToR switches download configuration files from the Trivial File Transfer Protocol (TFTP) server and automatically load them to automatically configure the network, thereby achieving zero-configuration deployment.

2.1.3 Protocols and Standards

- RFC1350: TFTP Protocol
- IEEE 802.1AB 2005: Station and Media Access Control Connectivity Discovery
- ANSI/TIA-1057: Link Layer Discovery Protocol for Media Endpoint Devices

2.2 Restrictions and Guidelines

- ToR switches need to support the Auto-provision function.
- The Link Layer Discovery Protocol (LLDP) function of core switches need to support port extension descriptions.

- In the Auto-provision solution, to differentiate between configuration files of different vendors, the configuration files are named after IP addresses and vendors in the format of x.x.x.x_vendor.cfg. Use the IP address192.168.215.16 of a device as an example. The name of the configuration file is **192.168.215.16_ruijie.cfg**.
- In the Auto-provision solution, ToR switches exchange device information with core switches through LLDP. ToR switches need to obtain their own IP addresses, IP address of the configuration file server, and configuration file names. Therefore, they use extension port descriptions of core switches. Port descriptions are in the format of device hostname-port@third byte of the device IP address@fourth byte of the IP address of the configuration file server@first two bytes of the device IP address, for example, NMG-NMG01A334-B-S6220-16.Int-AE1@215@11@192.168.

2.3 Configuring Auto-provision

2.3.1 Overview

When ToR switches are connected to a network for the first time, they can automatically load configurations, achieving zero-configuration deployment.

2.3.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Privilege the Auto-provision function.

auto-provision

The Auto-provision function is enabled by default.

2.4 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

⚠ Note

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 2-1 Auto-provision Monitoring

Purpose	Command
Displays the current configurations and status of the Auto-provision function.	**display auto-provision**
Debugs Auto-provision.	**debugging auto-provision**

2.5 Configuration Examples

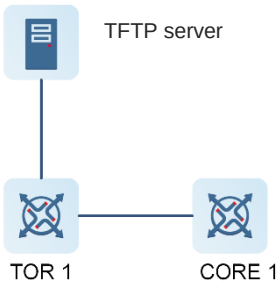
2.5.1 Configuring the Auto-provision Function

1. Requirements

When ToR switches are connected to a network for the first time, they can go online with zero configurations.

2. Topology

Figure 2-2 Auto-provision Topology



3. Notes

- If a device has no configuration file at startup, it automatically performs Auto-provision. If it has a configuration file, it does not perform Auto-provision.
- Before using the Auto-provision function, enable the SSH service on the TFTP server to enable TCP port 22 on the TFTP server.

4. Procedure

Privilege the Auto-provision function on the ToR switch.

```
TOR> privilege
TOR# config-system terminal
TOR(cfg)# auto-provision
```

5. Verification

Run the **display auto-provision** command to display the configuration result.

```
TOR# display auto-provision
Auto-provision manage info:
    Device mode   : S6220
    Admin state   : ENABLE
    Work status   : IDLE
    Process info : auto provision complete successfully.

Auto-provision port description info:
    Port name            : TenGigabitEthernet 2/0/1
    Port description : NMG-NMG01A334-B-S6220-241.Int-AE1@23@25@192.168
```

```
    Local ip              : 192.168.23.241
    Tftp server ip     : 192.168.23.25
    Syslog server ip : 192.168.23.2
Config file name : 192.168.23.241_ruijie.cfg
Device mode          : S6220
```

Contents

1 Configuring Basic Management 1

1.1 Overview 1

1.2 Basic Concepts 1

1.3 Protocol Specification 2

1.4 Configuration Task Summary 2

1.5 Configuring Basic System Parameters 3

1.5.1 Configuration Tasks 3

1.5.2 Configuring Connection to a Supervisor Module or Service Module 3

1.5.3 Configuring a System Name and Command Prompt 3

1.5.4 Configuring System Notifications 4

1.5.5 Configuring the System Clock-session 5

1.6 Enabling or Disabling a Specific Service 6

1.7 Configuring Passwords and Privilege Levels 6

1.8 Configuring Telnet Login 9

1.8.1 Configuration Tasks 9

1.8.2 Configuring Telnet Service 9

1.8.3 Configuring Telnet Access Security 11

1.9 Configuring Logging in to Other Devices through Telnet 13

1.10 Configuring Management 14

1.10.1 Configuration Tasks 14

1.10.2 Running Batch File Commands 14

1.10.3 Configuring Multiple-configuration Booting 15

1.10.4 Configuring Automatic Configuration File Backup to a Remote Server 15
1.10.5 Rolling Back System Configurations 16
1.10.6 Saving Configuration 17
1.11 Configuring a Restart Policy 17
1.11.1 Configuration Tasks 17
1.11.2 Configuring Immediate Restart 17
1.11.3 Configuring Scheduled Restart 18
1.12 Monitoring and Maintenance 18
1.13 Configuration Examples 20
1.13.1 Configuring Login Authentication and Telnet Service 20
1.13.2 Configuring Basic System Parameters 21

1 Configuring Basic Management

1.1 Overview

To reduce the inconvenience caused by on-site maintenance, IT devices are widely managed remotely. Even initial deployment can be carried out remotely. Remote management needs to consider the various operational needs of network administrators and the security of device management. A series of functions of basic management address security, maintenance and monitoring issues.

- Security: Basic management provides permissions, encryption authentication and access control to ensure device security.
- Maintenance: Basic management provides flexible configuration management methods, such as backup, rollback and batch execution, etc.
- Monitoring: Basic management provides various software and hardware monitoring functions to display information of memory, CPU, bus, session, etc..

1.2 Basic Concepts

Table 1-1 Basic Concepts

Concept	Description
TFTP	The Trivial File Transfer Protocol (TFTP) is a protocol used for simple file transfer between a client and a server in the Transmission Control Protocol (TCP)/Internet Protocol (IP) suite.
AAA	Authentication, Authorization and Accounting (AAA), including: ● Authentication: Verifies user identities and available network services. ● Authorization: Grants network services to users according to authentication results. ● Accounting: Records the network service consumption of users and to send the records to the billing system. After the AAA mode is enabled, some servers (or the local user database) are used to authenticate users' management permissions according to their usernames and password at terminal login and the configured AAA login authentication method list. For details about AAA, see *Configuring AAA*.
RADIUS	The Remote Authentication Dial in User Service (RADIUS) is the most widely used AAA protocol.
Telnet	Telnet is a terminal emulation protocol in the TCP/IP protocol suite which provides a connection to a remote host by creating a virtual terminal. It is a standard protocol at Layer 7 (application layer) of the Open System Interconnection (OSI) model and is used on the Internet for remote login. During remote login through telnet, users must enter the usernames and passwords for authentication. Telnet sets up a connection between the local personal computer (PC) and a remote host.

Concept	Description
System information	System information includes the system description, system power-on time, system hardware and software versions, control-layer software version, and boot-layer software version.
Hardware information	Hardware information includes the physical device information as well as information about pluggable modules on the device. ● The device information includes the device description and slot quantity. ● The slot information includes the slot ID, module description (which is empty if a slot does not have a module), number of physical ports on a module inserted into a slot, and maximum number of ports supported by a slot.
System configurations	System configurations include: ● Running configurations: Configurations running on all component modules of the system. ● Startup configurations: Configurations stored in the non-volatile random-access memory (NVRAM) of the system.

1.3 Protocol Specification

RFC 874: Telnet Protocol Specification

1.4 Configuration Task Summary

All the following configuration tasks are optional and may be selected as needed.

Basic functions:

- Configuring Basic System Parameters
  - Configuring Connection to a Supervisor Module or Service Module
  - Configuring a System Name and Command Prompt
  - Configuring System Notifications
  - Configuring the System Clock-session
- Enabling or Disabling a Specific Service

Permission and login control:

- Configuring Passwords and Privilege Levels
- Configuring Telnet Login
  - Configuring Telnet Service
  - Configuring Telnet Access Security

Configuration and maintenance:

- Configuring Logging in to Other Devices through Telnet
- Configuring Management
  - Running Batch File Commands
  - Configuring Multiple-configuration Booting

- Configuring Automatic Configuration File Backup to a Remote Server
- Rolling Back System Configurations
- Saving Configuration

- Configuring a Restart Policy
  - Configuring Immediate Restart
  - Configuring Scheduled Restart

1.5 Configuring Basic System Parameters

1.5.1 Configuration Tasks

Basic system parameter configuration includes the following tasks:

- Configuring Connection to a Supervisor Module or Service Module
- Configuring a System Name and Command Prompt
- Configuring System Notifications
- Configuring the System Clock-session

1.5.2 Configuring Connection to a Supervisor Module or Service Module

1. Overview

In a dual-node master/slave VSU environment (chassis-type devices) or multi-node VSU environment (box-type devices), the administrator can connect to a supervisor module or service module through this function.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Connect to a supervisor module or service module.

session { **master** | [**device** *device-number*] **slot** { **m**1 | **m**2 } }

1.5.3 Configuring a System Name and Command Prompt

1. Overview

To facilitate management, you can configure a system name for each device to identify the device. The default system name is **Ruijie**, and acts as the default command prompt. The command prompt changes with the system name. A system name longer than 32 characters is truncated to keep only the first 32 characters.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a system name.

hostname *hostname*

The default host name is **Ruijie**.

(4) Configure a command prompt.

prompt *prompt-string*

No CLI prompt is configured by default, and the system name is used as the command prompt.

1.5.4 Configuring System Notifications

1. Overview

System notifications are prompts displayed after user login and are classified into the following two types:

- Message of the day (MOTD): Sends urgent messages to users. MOTD information is displayed on the terminal after a user logs in to the device.
- Login banner: Provides some common login prompts and appears after MOTD information.

2. Restrictions and Guidelines

After entering a delimiter and pressing **Enter**, you can enter text, and then enter a delimiter and press **Enter** again to stop entering the text. Any characters following the ending delimiter are dropped. Text in the notification information must not contain the delimiter letter. The text length must not exceed

- 2047 bytes.
- 1023 bytes.
- 255 bytes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure MOTD information.

banner motd *c message c*

No MOTD information is configured by default.

(4) Configure login banner information.

banner line-login *c message c*

No login banner infoArmation is configured by default.

(5) Configure a prompt indicating the establishment of a reverse telnet connection.

banner incoming-prompt *c message c*

No prompt for the establishment of a reverse telnet connection is configured by default.

(6) Configure a prompt for the access to the privileged EXEC mode.

banner privilege-mode *c message c*

No prompt for the access to the privileged EXEC mode is configured by default.

(7) Configure a prompt for SLIP/PPP line connection.

banner slip-ppp-prompt *c message c*

No prompt for SLIP/PPP line connection is configured by default.

(8) Configure a prompt for user login authentication timeout.

banner authen-timeout-prompt *c message c*

No prompt for user login authentication timeout is configured by default.

(9) (Optional) Enter the line configuration mode.

line [**console** | **vty**] *first-line* [*last-line*]

(10) Configure a welcome prompt indicating that a user has entered the user EXEC mode of a line.

banner exec-prompt *c message c*

No welcome prompt indicating that a user has entered the user EXEC mode of a line is configured by default.

(11) (Optional) Configure a prompt indicating that the function of displaying EXEC prompt information is activated again for a specific line.

banner exec-banner-prompt *c message c*

The function of displaying EXEC prompt information is activated for all lines by default.

1.5.5 Configuring the System Clock-session

1. Overview

The system clock includes the date (year, month, and day), time (hour, minute, and second), and week information. This function is used to record event occurrence time, such as the system logging. When you use a device for the first time, set its system time to the current date and time manually.

2. Restrictions and Guidelines

- The device clock starts from the configured time and keeps running even when the device is powered off.
- If a device has no hardware clock, the manually configured time becomes invalid when the device is powered off.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure the system date and clock.

clock set *hh:mm*:*ss* [*MM* [*DD* [*YY*]]]

(3) (Optional) Update the hardware clock.

clock sync-to-calendar

The system is not configured to synchronize the hardware time with the software time by default.

(4) (Optional) Configure the hardware clock of the device

calendar set *hh:mm*:*ss* [*MM* [*DD* [*YY*]]]

The default hardware time of the device is 1970-01-01 00:00:00.

(5) (Optional) Configure the system to synchronize the software time with the hardware time.

clock sync-from-calendar

The system is not configured to synchronize the software time with the hardware time by default.

(6) (Optional) Configure the summer time.

clock fast-time *summer-time-zone* **start** *start-month* [*week* | **last**] *start-date hh*:*mm* **end** *end-month* [*week* | **last**] *end-date hh*:*mm* [**ahead** *hours-offset* [*minutes-offset*]

The summer time is not configured by default.

(7) (Optional) Configure the time zone.

clock timezone *timezone hours-offset* [*minutes-offset*]

The time zone is set as Universal Time Coordinated (UTC) by default.

1.6 Enabling or Disabling a Specific Service

1. Overview

When the system is running, you can dynamically adjust system services, including SSH server service, Telnet server service, SNMP agent service, and Web server service.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable a specific service.

enable-server service { **snmp**-**agent** | **ssh**-**server** | **telnet**-**server** }

The Simple Network Management Protocol (SNMP) agent service is enabled and the Secure Shell (SSH) server, Telnet server, and Web server services are disabled by default.

(4) (Optional) Save system configurations (**current-config**) to a specific position.

save [**auto-save interval** *interval-time* | **memory** [**auto-save interval** *interval-time*] | **terminal**]

1.7 Configuring Passwords and Privilege Levels

1. Overview

Privilege levels and passwords for different privilege levels can be configured to distinguish access needs of different users' devices.

16 privilege levels from 0 to 15 are defined for users in the command line interface (CLI) of network devices. Users of various privilege levels can run different commands. A smaller value indicates a lower privilege level. Level 0 is the lowest level and users of this level can run only a few commands, whereas level 15 is the highest level and users of this level can run all commands. Levels 0 and 1 are common user levels without the device configuration permission (users are not allowed to enter the global configuration mode). Levels 2 to 15 are privileged user levels with the device configuration permission.

Password protection is configured for each privilege level on the device so that users of different levels can use different command collections. An increase in privilege level requires the input of the correct password of the target privilege level, whereas a reduction in privilege level does not require password input.

Passwords fall into two types: Cipher and Safe-cipher.

- Cipher are simple encrypted passwords. You can set them for privilege levels 1 to 15.
- Safe-cipher are secure encrypted passwords. You can set them for privilege levels 1 to 15.

Passwords must be stored in encryption mode. Cipher use simple encryption, and Safe-cipher use secure encryption.

⚠ Caution

- If a privilege level is configured with both a Cipher and a Safe-cipher, the Cipher does not take effect.
- If no password is configured for a privileged user level, you do not need to input a password to enter this level. For security purposes, it is recommended that a password be configured for privileged user levels.

2. Restrictions and Guidelines

- When a Cipher or Safe-cipher is configured for the first time and the plaintext string is less than eight characters or contains only one type of characters, the system prompts you that it is a weak password.
- After logging in to a device, a user can run the **privilege**or **disable** command to raise or lower his/her privilege level to access commands at different privilege levels. An increase in privilege level requires the input of the correct password of the target privilege level by default.
- You can specify the **level** keyword in the password configuration command to configure a password for a specific privilege level. After you set a password for a specific privilege level, the password works for the users who need to access commands of this level.
- The **privilege** commands include **privilege** (for role switching), **privilege cipher**, and **privilege safe-cipher**. If no role is specified, these three commands are used to set the role configured by running the **privilege** command.
- **privilege cipher**
  - This command is used to configure Cipher only for level 15 and takes effect only when no Safe-cipher is configured. If you use this command to configure a cipher for a non-15 level, the system displays a warning and the cipher is automatically converted into a Safe-cipher.
  - If the cipher and Safe-cipher set for level 15 are the same, the system displays a warning.
  - If you specify an encryption type but enter a plaintext cipher during cipher configuration, you cannot enter the privileged EXEC mode again.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) (Optional) Display help information.

Help

(3) Change the privilege level.

- Raise the privilege level.

privilege [*privilege-level* | **role** *role-name*]

When the role-based access control (RBAC) function is enabled, this command can be used to switch the terminal role. If no role is specified, the system switches to role **network-admin** by default.

- Lower the privilege level.

  disable [*privilege-level*]

  The privilege level specified in this command must be lower than the current level.

  When the RBAC function is enabled, this command is unavailable.

(4) Enter the global configuration mode.

config-system terminal

(5) Configure a password.

privilege cipher [**level** *password-level* | **role** *role-name*] | [**0**] *password* | **7** *encrypted-password*]

Preamble spaces are allowed in front of the password but the spaces are ignored. Intermediate and trailing spaces are recognized.

(6) Configure a Safe-cipher.

privilege safe-cipher { [**level** *secret-level*] | [**role** *role-name*] } { [**0**] *password* | { **5** | **8** } *encrypted-secret* }

When the RBAC function is enabled and no role is specified, this command is used to set a password for role **network-admin** by default.

(7) (Optional) Configure a handling policy when the number of consecutive authentication failures to enter the privileged EXEC mode reaches the limit.

privilegeuser-block { **disable** | **failed-times** *times* **period** *period* | **reactive** *time* }

The allowed maximum number of consecutive authentication failures is 3, the period for resetting the consecutive authentication failure count is 10 minutes, and a user will be locked for 10 minutes after the number of consecutive authentication failures reaches the limit.

(8) (Optional) Configure the algorithm used for encryption.

privilege algorithm-type { **md5** | **sha256** }

The SHA-256 algorithm is used for encryption by default.

(9) (Optional) Configure the role for running the **enable** commands.

privilege default role *role-name*

The default role for running **privilege** commands is **network-admin**.

Only when the RBAC function is enabled, this command is available.

(10) (Optional) Enable level increase logging.

login privilege log

The level increase or role switching logging function is disabled by default.

(11) Configure command privilege levels.

privilege *mode* { **all** | **level** *level* | **reset** } *command-string*

(12) Enter the line configuration mode.

line [**console** | **vty**] *first-line* [*last-line*]

(13) Configure a password for line-based login.

cipher { [**0**] *password* | **7** *encrypted-password* }

(14) Verify the password for line-based login.

line-login

The verification function of simple login passwords is disabled for the console line and enabled for the virtual terminal lines by default.

(15) Exit the configuration mode and return to the upper-level mode or exit the command line interface (CLI) from the privileged EXEC mode.

exit

1.8 Configuring Telnet Login

1.8.1 Configuration Tasks

Telnet login includes the following tasks:

- Configuring Telnet Service
- Configuring the Number of Authentication Failures to Block an IP Address and the Period for Counting Consecutive Authentication Failures on the Telnet Server
- Configuring Telnet Access Security
- Configuring Logging in to Other Devices through Telnet

1.8.2 Configuring Telnet Service

1. Overview

As an application-layer protocol in the TCP/IP protocol suite, telnet provides the standard for remote login and virtual terminal communication on the Internet.

The Telnet client service allows a local or remote login user of the device to access other remote system resources on the Internet. As shown in the following figure, a user uses a PC to connect to device A by using the terminal emulation program or telnet program and then logs in to device B by running the **telnet** command to configure and manage device B.

Figure 1-1 Telnet Service



The telnet program can use an IPv4 or IPv6 address for communication. The Telnet server can accept telnet connection requests from IPv4 and IPv6 addresses. The Telnet client can send connection requests to hosts identified by IPv4 and IPv6 addresses.

When the telnet server service is enabled on the device, you can use the Telnet client to connect to the device and configure the following functions:

- Configure an access control list (ACL) for the Telnet server.
- Disable the IP address blocking function of the telnet server.
- Configuring the Number of Authentication Failures to Block an IP Address and the Period for Counting Consecutive Authentication Failures on the Telnet Server
- Configure the period for awakening blocked IP addresses on the Telnet server.
- Cleanup entries about blocking and authentication failures of all or specific IP addresses.

2. Restrictions and Guidelines

- When the number of authentication failures of telnet login meets the IP address blocking conditions in the authentication failure count period, the source IP address blocking is triggered. That is, the Telnet client of this source IP address is not allowed to log in to the device to prevent the device from being attacked. The Telnet client can log in to the device only after the IP address awakening period expires.
- In the enabled IP address blocking function, a user logs in to the device through Telnet. When the number of consecutive authentication failures reaches the configured count within the authentication failure count period, source IP address blocking is triggered. When such number does not reach the configured count or one authentication operation is successful within the authentication failure count period, the authentication failures are cleared.
- After the time for awaking a blocked source IP address comes, entries about the IP address blocking are cleared. The blocked IP address is awakened immediately and can log in to the device through the Telnet client.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ACL for the Telnet server.

ip telnet access-apply { *acl-number* | *acl-name* }

ipv6 telnet access-apply *acl-name*

(4) Disable the IP address blocking function of the Telnet server.

ip telnet ip-block disable

The IP address blocking function of the telnet server is enabled by default.

(5) Configure the number of authentication failures, beyond which an IP address is blocked, and the -period for counting consecutive authentication failures on the telnet server.

ip telnet ip-block failed-times *failed-times* **period** *failed-period-time*

The allowed maximum number of authentication failures is 6, and the -period for counting consecutive authentication failures is 5 minutes by default.

(6) Configure the period for awakening blocked IP addresses on the telnet server.

ip telnet ip-block reactive *reactive-period-time*

The default period for awakening blocked IP addresses is 5 minutes.

(7) Enter the line configuration mode.

line vty *line-number*

(8) (Optional) Configure the connection timeout time.

line-timeout *exec-timeout-minutes* [*exec-timeout-seconds*]

The default connection timeout time is 10 minutes.

If there is no input information during the specified time, the server interrupts the established connection.

(9) (Optional) Configure the session timeout time.

session-idle-time *session-timeout-time* [**output**]

The default session timeout time is 0 minute for remote terminals. That is, the sessions never time out. If there is no input information during the specified time, the device closes sessions established to a remote terminal on the current line and restores the terminal to the idle state.

(10) (Optional) Privilege EXEC prompt information display for a specific line.

exec-banner-prompt

EXEC prompt information display is enabled for all lines by default.

(11) (Optional) Privilege MOTD prompt information display for a specific line.

motd-banner

MOTD prompt information display is enabled for all lines by default.

1.8.3 Configuring Telnet Access Security

1. Overview

When AAA is disabled, you can configure a line password or local user authentication to control users to log in to and manage the device. When login authentication (via the **line-login** command) is configured for a line, only users who pass the line password verification are allowed to log in. When local user authentication (via the **line-login local** command) is configured for a line, the username and password entered by a user are checked against those stored in the local user database. If they are matched, the user can access the device with proper management permissions.

When AAA is enabled, some servers can be used to authenticate users' management permissions by their usernames and passwords at their login. Only authenticated users are allowed to log in. For example, a RADIUS server can authenticate usernames and passwords and control users' permissions to manage the device. Thus, instead of using locally stored password information for authentication, the device sends encrypted user information to the RADIUS server for verification. The server configures unified usernames, passwords, shared passwords, and access policies of users to manage and control user access and improve the security of user information.

⚠ Caution

After AAA is enabled, line password verification and local user authentication do not take effect.

2. Restrictions and Guidelines

- In the enabled AAA authentication mode, set line-based login for this authentication, and use the AAA authentication methods, including RADIUS authentication, local authentication, and no authentication.

- In the enabled AAA security service, to perform non-AAA authentication for a line run the **line-login access non-aaa** command. The configuration is valid for all terminals.
- The **user-account** command is used to create a local user database for authentication. The encryption type **7** needs to be specified only when encrypted passwords are copied and pasted. If the value **7** is specified as the encryption type, the entered ciphertext string must consist of an even number of characters. The login user cannot delete his/her account.
- To lock a session, enable locking on the terminal connected to a line in line configuration mode, and run the **lock-session** command in the EXEC mode of the terminal to lock the terminal. When a user enters any character on the locked terminal, the password prompt is displayed. The terminal will be automatically unlocked if the entered password is correct.

3. Procedure

(1) Enter the privileged EXEC mode.

(2) **privilege**Enter the global configuration mode.

config-system terminal

(3) Configure a local user account and optional authorization information.

user-account *username* [**algorithm-type** { **md5** | **sha256** }] [**login mode** { **console** | **ssh** | **telnet** | **ftp** }] [**online amount** *amount-number*] [**permission** *oper-mode path*] { [**privilege** *privilege-level*] | [**role** *text-string*] } [**reject remote-login**] [**vid** *vid-number*] [**web-auth**] [**pwd-modify**] [**nopassword** | **cipher** [**0** | **7**] *text-string* | **safe-cipher** [**0** | **5** | **8**] *text-string*]

No local user account or authorization information is configured by default.

(4) (Optional) Import user information from a file.

user-account import *file-name*

(5) (Optional) Export user information to a file.

user-account export *file-name*

(6) Enter the line configuration mode.

line [**console** | **vty**] *first-line* [*last-line*]

(7) Configure the encryption type for line-based login.

privilege algorithm-type { **md5** | **sha256** }

The default encryption type for line-based login is SHA256.

(8) Configure an MD5/SHA-256 irreversible encrypted password for line-based login.

privilege safe-cipher { [**0**] *password* | { **5** | **8** } *encrypted-secret* }

No encrypted password is configured for line-based login by default.

(9) Configure a password for line-based login.

privilege cipher { [**0**] *password* | **7** *encrypted-password }*

No password is configured for line-based login by default.

(10) Configure local authentication for line-based login.

line-login local

When AAA is enabled, no local user authentication is configured for lines by default,

(11) (Optional) Configure AAA authentication for line-based login.

line-login authentication { *list-name* | **default** }

When AAA is enabled, the **default** authentication method is used by default.

Note

This command is not supported by only the S6520-64CQ.

(12) (Optional) Configure non-AAA authentication for line-based login when AAA is enabled.

line-login access non-aaa

When AAA is enabled, non-AAA authentication is disabled by default.

(13) (Optional) Privilege locking on a terminal connected to a line.

lock-session-enable

The function of locking terminals connected to the current line is disabled by default.

(14) Return to the privileged EXEC mode.

end

1.9 Configuring Logging in to Other Devices through Telnet

1. Overview

As an application-layer protocol in the TCP/IP protocol suite, telnet provides the standard for remote login and virtual terminal communication on the Internet. The telnet client service allows a local or remote login user of the device to access other remote system resources on the Internet.

2. Restrictions and Guidelines

If you have run the **telnet** command to initiate a telnet client session, you can press **Ctrl+Shift+6+X** (press **Ctrl+Shift+6**, release the buttons, and then press **X**) to temporarily exit the session. To restore this session, run the <**1**–**99**> command. To display information about established sessions, run the **display login-sessions** command.

- Before using a Telnet device, you need to run the **active service telnet-server** command to enable the Telnet server service on the device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Run the **telnet** command to log in to the telnet server.

telnet [**oob**] { *hostname* | *ipv4-address* | *ipv6-address* } [*port-number*] [**/source** { **ip** *ipv4-address*| **ipv6** *ipv6-address*| **interface** *interface-type interface-name* }] [**/vrf** *vrf-name*] [**via mgmt** *mgmt-number*]

(3) Run the **do telnet** command to log in to the telnet server.

do telnet [**oob**] { *hostname* | *ipv4-address* | *ipv6-address* } [*port-number*] [/**source** { **interface** *interface-type interface-number* | **ip** *ipv4-address* | **ipv6** *ipv6-address* | **interface** *interface-type interface-name* }] [**/vrf** *vrf-name*] [**via mgmt** *mgmt-number*]

(4) (Optional) Restore the established telnet client session.

1–99

(5) (Optional) Disconnect a suspended telnet client session.

remove *session-id*

(6) (Optional) Specify the IP address of an interface as the source IP address of a Telnet connection. .

ip telnet source *interface-type interface-number*

The Telnet source interface is the reachable outbound interface of the Telnet server by default.

(7) (Optional) Return to the privileged EXEC mode.

end

(8) (Optional) Lock the current terminal connected to a line.

lock-session

1.10 Configuring Management

1.10.1 Configuration Tasks

Management Configuration includes the following tasks. All the following configuration tasks are optional and may be selected as needed.

- Running Batch File Commands
- Configuring Multiple-configuration Booting
- Configuring Automatic Configuration File Backup to a Remote Server
- Rolling Back System Configurations
- Saving Configuration

1.10.2 Running Batch File Commands

1. Overview

To management system functions, it may take a long time to enter many commands on the CLI. This process is prone to errors and omissions. You can put the commands in a batch file according to configuration steps, and execute the file to complete related configurations.

2. Restrictions and Guidelines

- You can specify the name and content of the batch file on your PC and transfer the file to the flash memory of the device through TFTP. The content of the batch file simulates user input. Therefore, you must edit the content according to the configuration sequence of the CLI commands. For some interactive commands, you must write the responses in the batch file to ensure that the commands are normally run.
- The batch file must not exceed 128 KB in size; otherwise, it will fail to be executed. You can divide a large batch file into multiple files smaller than 128 KB in size each.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Run the **execute** command to execute the batch file.

execute { [**flash:**] *file-name* }

1.10.3 Configuring Multiple-configuration Booting

1. Overview

Multiple-configuration booting allows users to modify the saving paths and names of startup configuration files of the device. This function saves configurations to an extended flash memory or an extended Universal Serial Bus (USB) flash drive of the device only.

2. Restrictions and Guidelines

- The startup configuration file name can be a file name or an existing path. For example, if the startup configuration file name is set to **flash:/Hostname/Hostname.text** or **usb0:/Hostname/Hostname.text**, **flash:/Hostname.** The file folder **flash:/Hostname** and **usb0:/Hostname** must exist, or the **save** command will fail to save the configurations. In the master-slave mode, the paths must exist on all devices.
- To save configurations to an extended USB flash drive, the device must support at least one USB port. If the device supports two or more USB ports, this function saves startup configurations to USB 0 only.
- To save the startup configuration file to a USB flash drive, the device must provide a USB port with a USB flash drive inserted. Otherwise, the **save** command will fail to save the configurations. In the master-slave mode, all devices must have a USB flash drive inserted.

3. Procedure

(1) Enter the privileged EXEC mode.

(2) **privilege**Enter the global configuration mode.

config-system terminal

(3) Modify the saving path and name of startup configuration file.

boot config { **flash:***file-name* | **usb0:***file-name* }

1.10.4 Configuring Automatic Configuration File Backup to a Remote Server

1. Overview

By configuring the specific information and interval, you can automatically back up the configuration file of the device to the remote server.

2. Restrictions and Guidelines

- If no configuration file exists during command execution, an error is displayed.
- If the configuration file is deleted after the configuration command takes effect, the system stops backing up the configuration file to the remote server after the preset time expires.

3. Procedure

(1) Enter the privileged EXEC mode.

(2) **privilege**Enter the global configuration mode.

config-system terminal

(3) Configure automatic backup of the configuration file backup to a remote server.

auto-backup configuration [**oob**] { **ftp** *server* [**port** *port-number*] **username** *username* **cipher** { [0] *password* | 7 *encrypted-password* } | **tftp** *server* } **interval** *interval-time* [**path** *folder*]] [**via mgmt** *mgmt_name*] [**vrf** *vrf_name*]

The automatic configuration file backup function is not configured by default.

1.10.5 Rolling Back System Configurations

1. Overview

Rollback configuration allows you to make a snapshot for the current configurations, that is, a copy or checkpoint of the current configurations, and apply the checkpoint configurations to the device without restarting the device. This function applies to the following scenarios:

- When current system configurations contain too many errors to locate or roll back one by one, you must roll back the current configurations to a previous correct state.
- When the device application environment changes and the device needs to run the configurations in a configuration file, you can roll back the current configurations to the specified configuration file state without restarting the device.

During rollback, the system compares and handles the differences between the current configurations and checkpoint configurations.

- For the same commands in both configurations, the system does not process them.
- For the commands only in the current configurations, the system cancels them.
- For the commands only in the checkpoint configurations, the system runs them.
- For the different commands between both the current configurations and checkpoint configurations, the system cancels them, and then runs related commands in the checkpoint configurations.

2. Restrictions and Guidelines

- The checkpoint quantity varies with the flash memory size of the device. For devices with a small flash memory, four checkpoints are supported. The default checkpoint quantity is 10.
- Only one user can create checkpoints and configure rollback on a device at a time.
- It is recommended that you check the consistency of serial port baud rates between the current system configurations and checkpoint configurations before you perform rollback. If they are inconsistent, you are advised to change the serial port baud rate to that of the checkpoint configurations. Otherwise, a rate change will occur during rollback, causing a failure to display the rollback process information.
- During configuration rollback, do not hot-swap any supervisor module, line card, or service board and ensure that the device topology environment is the same as the environment of checkpoint creation. For example, if the device topology is a standalone environment during checkpoint creation but a virtual switching unit (VSU) environment during rollback, configuration rollback may fail.
- If an “Increased configuration:” message is displayed after rollback, the content after the message are configurations increasing from the checkpoint configurations. The message is displayed because some commands cannot be reversed or fail to be reversed. For details, see the command reference of specific functions, and manually reserve these commands.
- If a "Decreased configuration:" message is displayed after rollback, configurations decrease from the checkpoint configurations. This is because some commands fail to be executed during rollback. For details, see the command reference of specific functions, and manually run these commands.

3. Procedure

(1) Enter the privileged EXEC mode.

(2) **privilege**Create a checkpoint.

checkpoint [*checkpoint -name*] [**description** *description*]

No checkpoint is configured by default.

(3) Roll back configurations.

rollback current-config checkpoint *checkpoint-info -name* [**display-differences** | **ignore-results**]

1.10.6 Saving Configuration

1. Overview

The device supports manual saving and automatic saving. The running configuration is not saved by default, and the device will use the startup configuration file after restarting. Therefore, you are advised to use this function in time to save the running configuration to the startup configuration file

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Save system configurations (**current-config**) to a specific position.

save [**auto-save interval** *interval-time* | **memory** [**auto-save interval** *interval-time*] | **terminal**]

1.11 Configuring a Restart Policy

1.11.1 Configuration Tasks

The restart policy configuration includes the following tasks:

- Configuring Immediate Restart
- Configuring Scheduled Restart

1.11.2 Configuring Immediate Restart

1. Overview

Immediate restart applies when the device needs to be restarted immediately.

2. Restrictions and Guidelines

- A restart may interrupt services. Exercise caution.
- If the device to be restarted is being upgraded, it does not perform the restart.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure immediate device restart.

reboot

1.11.3 Configuring Scheduled Restart

1. Overview

A restart policy enables the device to restart as scheduled. The following two scheduled restart functions are supported:

- Configure the system to restart after an interval. The interval is in the format of *mmm* or *hhh*:*mm*, in minutes. You can select either of the formats. You can specify an interval name to reflect the restart purpose.
- Configure the system to restart at a future time point.

2. Restrictions and Guidelines

- A restart may interrupt services. Exercise caution.
- If the device to be restarted is being upgraded, the system does not perform the restart.
- The restart time must be later than the current system time but cannot be more than 31 days later than the current system time. After you configure a restart schedule, do not change the system clock (for example, change the system time to a time after the restart time). Otherwise, the configuration may fail.
- To restart the system at a future time point, the system must support the clock function and the input time value must be a future time point. A new restart schedule overwrites the existing one. If the system is restarted before a restart schedule takes effect, the schedule will be lost.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure scheduled device restart.

- Configure a scheduled restart schedule.

  reboot at *hh*:*mm*:*ss* [*MM* [*DD* [*YY*]]]

- Configure a countdown restart schedule.

  reboot in { [*hh* :] *mm* }

  No restart function is configured by default.

(3) (Optional) Cancel the restart schedule.

reboot cancel

1.12 Monitoring and Maintenance

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-2 Monitoring and Maintenance of Basic Management

Command	Function
cleanup checkpoint-info database	Cleanup checkpoints and related data.

Command	Function
cleanup enable user-block { **all** | **no-user** | *user* }	Unlock users who are locked because the number of consecutive authentication failures to enter the privileged EXEC mode exceeds the limit.
cleanup telnet ip-block { **all** | [*ipv4-address* | *ipv6-address*] }	Cleanup entries about blocked IP addresses and authentication failures.
cpu high-watermark set [[**up** *up-value*] [**down** *down-value*]]	Configure the maximum threshold for the total CPU usage of all control cores and enable CPU usage monitoring.
memory history cleanup [**all** | **half** | **one-forth**]	Cleanup historical memory usage records.
memory low-waterline set *memory-threshold*	Enable the monitoring of memory usage threshold.
display boot config	Display the saving path and name of startup configuration file.
display calendar	Display the hardware time of the system.
display checkpoint-info { **different-from** {*checkpoint-name* | **current-config**} } [**all**] | **summary** }	Display all information about a checkpoint or a summary of all checkpoints.
display clock	Display the current system time.
display cpu [**core**]	Display CPU usage information of system tasks on control cores and non-virtual cores.
display debugging	Display information about enabled debugging functions.
display enable user-block	Display users who are blocked from privileged EXEC mode authentication on the device.
display hostname	Display the host name of the device.
display language character-set	Display the current character set encoding format of the device.
display line { **console** *line-number* **| vty** *line-number* **|** *line-number* }	Display the line configurations.
display memory [**history-stats** | **low-waterline** | *process-id* | *process-name* | **slot** [*uslot* [**sorted total**] | *all-info* [**sorted total**]] | **sorted total**]	Display memory information.
display memory vsd *vsd-id*	Display Virtual Storage Director (VSD) information.
display pci-bus	Display information about devices mounted on the Peripheral Component Interconnect (PCI) bus.

Command	Function
display processes-info cpu detailed { *process-id* | *process-name* }	Display details about a specific task.
display processes-info cpu [**history-stats** [**table**] | [**5sec** | **1min** | **5min** | **15min**] [**nonzero**] [**record**]]	Display system tasks by VSD.
display processes-info vsd *vsd-id* **cpu**	Display details about a specific task.
display reboot	Display system restart settings.
display current-config [**interface** *interface-type interface-number* | **router** { **bgp** | **isis** | **ospf** }]	Display the running configurations of the device system or configurations of an interface.
display service	Display the service status (enabled/disabled).
display login-sessions	Display information about established Telnet client instances.
display bootup-config	Display the device configurations stored in the NVRAM.
display sysmon grpc info	Display information about the gRPC function registered in the system monitoring process.
display telnet ip-block { **all** | **list** }	Displays information about blocked IP addresses and authentication failures.
display this-config	Displays effective system configurations in current mode.
display usb-bus	Display information about devices mounted on the USB bus.
display version	Displays system information.

1.13 Configuration Examples

1.13.1 Configuring Login Authentication and Telnet Service

1. Requirements

You can use Device A as a jump server through the workstation to manage Device B remotely.

- Establish a telnet session to a remote device.
- Complete login identity authentication.

2. Topology

Figure 1-2 Configuring the Telnet Service



3. Notes

- Establish a Telnet session to the remote device whose IPv4 address is 192.168.65.119.
- Establish a Telnet session to the remote device whose IPv6 address is 2AAA:BBBB::CCCC.

4. Procedure

(1) Configure user and authorization information on Device B.

```
Hostname> privilege
Hostname# config-system terminal
Hostname(cfg)# line vty 0
Hostname(cfg-line)# cipher Guestuser
Hostname(cfg-line)# line-login
```

(2) Establish a telnet session to a remote device on Device A.

Run the **telnet** command in privileged EXEC mode, or run the **do telnet** command in user EXEC mode, privileged EXEC mode, or interface configuration mode.

```
Hostname# telnet 192.168.65.119
Trying 192.168.65.119 ... Open
User Access Verification
Password: Guestuser
Hostname# telnet 2AAA:BBBB::CCCC
Trying 2AAA:BBBB::CCCC ... Open
User Access Verification
Password:
Hostname(cfg)# do telnet 2AAA:BBBB::CCCC
Trying 2AAA:BBBB::CCCC ... Open
User Access Verification
Password: Guestuser
```

5. Verification

- Run the **ping** command to display the configurations. If the remote device can be pinged, the telnet service is configured.
- Verify the login identity. If the login is successful, login authentication is configured.

1.13.2 Configuring Basic System Parameters

1. Notes

- Configure the system time.

- Configure MOTD information.
- Configure login banner information.
- Set the serial port baud rate to 57,600 bps.

2. Procedure

(1) Configure the system time.

Set the system time to June 20, 2003, 10:10:12.

```
Hostname> privilege
Hostname# clock set 10:10:12 6 20 2003
```

(2) Configure MOTD information.

Set the MOTD content to "Notice: system will shutdown on July 6th." with the pound key (#) as the delimiter.

```
Hostname# config-system terminal
Hostname(cfg)# banner motd #
Enter TEXT message.  End with the character '#'.
Notice: system will shutdown on July 6th.#
Hostname(cfg)#
```

(3) Configure login banner information.

Set the login banner content to "Access for authorized users only. Please enter your password." with the pound key (#) as the delimiter.

```
Hostname(cfg)# banner line-login #
Enter TEXT message.  End with the character '#'
Access for authorized users only. Please enter your password.
```

(4) Set the serial port baud rate to 57,600 bps.

```
Hostname# config-system terminal
Hostname(cfg)# line console 0
Hostname(cfg-line)# speed 57600
```

3. Verification

- Verify the system time.

  Run the **display clock** command in privileged EXEC mode to display the system time.

```
Hostname# display clock
clock: 2003-6-20 10:10:54
```

- Verify MOTD information.

  Connect to the local device through the console, telnet, or SSH, and check whether the MOTD information is displayed before the CLI appears.

```
Hostname# telnet 192.168.65.236
Notice:  system will shutdown on July 6th.
Access for authorized users only. Please enter your password.
 User Access Verification
Password:
```

- Verify login banner information.

Connect to the local device through the console, telnet, or SSH, and check whether the login banner information is displayed before the CLI appears.

```
Hostname# telnet 192.168.65.236
Notice:    system will shutdown on July 6th
Access for authorized users only. Please enter your cipher
  User Access Verification
Password:
```

- Verify that the serial port baud rate is set to 57,600 bps.

  Run the **display line** command to display the configurations.

```
Hostname# display line console 0
CON         Type       speed      Overruns
* 0         CON        57600      0
Line 0, Location: "", Type: "vt100"
Length: 25 lines, Width: 80 columns
Special Chars: Escape    Disconnect    Activation
                       ^^x         none              ^M
Timeouts:           Idle EXEC       Idle Session
                       never              never
History is enabled, history-stats size is 10
Total input: 22 bytes
Total output:    115 bytes
Data overflow:    0 bytes
stop rx interrupt:    0 times
Modem: READY
```

Contents

1 Configuring RBAC 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Concepts 1
1.2 Configuration Task Summary 2
1.3 Configuring a Feature Group 3
1.3.1 Overview 3
1.3.2 Restrictions and Guidelines 3
1.3.3 Procedure 3
1.4 Configuring Role Permissions 3
1.4.1 Overview 3
1.4.2 Configuration Tasks 4
1.4.3 Enabling the RBAC Function 4
1.4.4 Configuring Roles 4
1.4.5 Configuring Rule Permissions for a Role 5
1.4.6 Configuring Description of a Role 6
1.4.7 Prohibiting a Role from Operating All Interface Resources 7
1.4.8 Allowing a Role to Operate a Specific Interface Resource 7
1.4.9 Prohibiting a Role from Operating All VLAN Resources 7
1.4.10 Allowing a Role to Operate a Specific VLAN Resource 8
1.4.11 Prohibiting a Role from Operating All VRF Resources 8
1.4.12 Allowing a Role to Operate a Specific VRF Resource 9

1.5 Monitoring 9

1.6 Configuration Examples..... 10

1.6.1 Configuring Role Permissions 10

1 Configuring RBAC

1.1 Introduction

1.1.1 Overview

Role-based access control (RBAC) associates roles with permissions. Users are assigned with appropriate roles with -permissions. The authorization structure of user-role-permission is formed to simplify permission management. Roles are defined to complete various tasks, and a device administrator can predefine all the roles and their permissions. To change a user's permission, use RBAC to change only the permission of his role. This process reduces the authorization workload and device management overhead.

1.1.2 Basic Concepts

- Feature

  Features vary with CLI commands. CLI commands with the same features constitute one feature.

- Feature group

  A feature group is composed of features. Different features are classified as needed to form one feature group. In short, a feature group is composed of several types of CLI commands.

- User

  By default, a user does not have any permissions to operate a device. Only after you specify a proper role to a user, the user has the permission of this role. There are local users and remote AAA server users. Local users are used for local authentication while remote AAA server users are used for remote AAA authentication.

- Role

  There are predefined system roles and user-defined roles. The system predefines 18 roles, including **network-admin**, **network-operator**, and **priv-***n* (0–15). Each role is granted with specific operation permissions. Table 1-1 lists these roles and their permissions.

Table 1-1 Roles and Permissions

Role	Default Permission
network-admin	Administrator role, with all operation permissions.
network-operator	Operator role, with the following permissions: ● Permission control: ○ CLI commands: allowed to run the **ping**, **display**, **ssh**, **telnet**, **traceroute**, **ssh-session**, and **terminal** commands and the command to change the current local user password. The **display** command displays all information. ● Resource control: ○ Interfaces: allowed to operate all interfaces. ○ Virtual local area networks (VLANs): allowed to operate all VLANs. ○ VPN routing and forwarding (VRF): allowed to operate all VRF instances.

priv-0	Level-0 role, assigned with the following permissions: ● Permission control: ○ CLI commands: allowed to run the **ping**, **ssh**, **telnet**, **traceroute**, **ssh-session**, and **privilege** commands. ● Resource control: ○ Interfaces: allowed to operate all interfaces. ○ VLANs: allowed to operate all VLANs. ○ VRF: allowed to operate all VRF instances.
priv-n (1-13)	Level-1 to level-13 roles, without any permission
priv-14	Level-14 role, with the following permissions: ● Permission control: ○ CLI commands: allowed to run CLI commands other than the **more**, **upgrade**, and **debugging** commands that are executable only by an administrator. ● Resource control: ○ Interfaces: allowed to operate all interfaces. ○ VLANs: allowed to operate all VLANs. ○ VRF: allowed to operate all VRF instances.
priv-15	Level-15 role, with the same all operation permissions as the role **network-admin**.

- Permission

  Permissions fall into three modes - read, write, and execute - or into two types - rule permissions and resource permissions. Rule permissions include those of command-based rules, those of feature-based rules, and those of feature group-based rules. Resource permissions include those of interface resources, those of VLAN resources, and those of VRF resources.

  Rule permissions configured for user roles are divided into the following categories:

  - Prohibit a role from running or allow a role to execute a specific command line.
  - Prohibit a role from running or allows a role to run one or several types of commands for specified or all features.
  - Prohibit a role from running or allows a role to run one or several types of commands for all features in a feature group.

  Resources permissions configured for user roles are divided into the following categories:

  - Prohibit a role from operating or allow a role to operate all or some interfaces.
  - Prohibit a role from operating or allow a role to operate all or some VLANs.
  - Prohibit a role from operating or allow a role to operate all or some VRF instances.

1.2 Configuration Task Summary

RBAC configuration includes the following tasks:

(1) (Optional) Configuring a Feature Group

(2) Configuring Role Permissions

a Enabling the RBAC Function

b Configuring Roles

c Configuring Rule Permissions for a Role

d (Optional) Configuring Description of a Role

e (Optional) Prohibiting a Role from Operating All Interface Resources

f (Optional) Allowing a Role to Operate a Specific Interface Resource

g (Optional) Prohibiting a Role from Operating All VLAN Resources

h (Optional) Allowing a Role to Operate a Specific VLAN Resource

i (Optional) Prohibiting a Role from Operating All VRF Resources

j (Optional) Allowing a Role to Operate a Specific VRF Resource

1.3 Configuring a Feature Group

1.3.1 Overview

This section describes how to create a feature group and add features to the feature group.

1.3.2 Restrictions and Guidelines

- Feature groups predefined in the system cannot be deleted or modified.
- Up to 64 feature groups can be customized.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a feature group and enter the feature group configuration mode.

role feature-group name *group-name*

The system predefines feature groups **L2** and **L3** by default. Feature group **L2** contains all commands for functions related to Layer 2 protocols. Feature group **L3** contains all commands for functions related to Layer 3 protocols.

(4) Add features to the feature group.

feature *feature-name*

By default, a system predefined feature group contains default features while a user-defined feature group contains no feature.

1.4 Configuring Role Permissions

1.4.1 Overview

This section describes how to create a user role and configure its operation permissions. After a user is authenticated to get a proper role, he has operation permissions.

1.4.2 Configuration Tasks

User role permission configuration includes the following tasks:

(1) Enabling the RBAC Function

(2) Configuring Roles

(3) Configuring Rule Permissions for a Role

(4) (Optional) Configuring Description of a Role

(5) (Optional) Prohibiting a Role from Operating All Interface Resources

(6) (Optional) Allowing a Role to Operate a Specific Interface Resource

(7) (Optional) Prohibiting a Role from Operating All VLAN Resources

(8) (Optional) Allowing a Role to Operate a Specific VLAN Resource

(9) (Optional) Prohibiting a Role from Operating All VRF Resources

(10) (Optional) Allowing a Role to Operate a Specific VRF Resource

1.4.3 Enabling the RBAC Function

1. Overview

This section describes how to enable the RBAC function.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the RBAC function.

role enable

The RBAC function is disabled by default.

1.4.4 Configuring Roles

1. Overview

This section describes how to customize a role.

2. Restrictions and Guidelines

- System predefined roles cannot be deleted by running the **no** command. The default permission of only the **priv**-*n* (0–13) role can be restored by running the **default** command.
- Permissions can be added to the **priv**-*n* (0–13) role only.
- Users can customize up to 64 roles and configure permissions for the roles. .

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a role and enter the role configuration mode.

role name *role-name*

By default, the system predefines 18 roles, including **network-admin**, **network-operator**, and **priv**-*n* (0–15). Each role is granted with specific operation permissions.

1.4.5 Configuring Rule Permissions for a Role

1. Overview

This section describes how to configure the rule permissions for a role.

2. Restrictions and Guidelines

- By default, system predefined roles have predefined rule permissions while user-defined roles have no rule permission.
- During rule configuration, if the specified rule number does not exist, create a rule; otherwise, modify the rule corresponding to the specified rule number. The modified rule supports newly authenticated users only.
- A user role is allowed to create multiple rules, and permissions executable by this role is a union set of these rules. If permissions defined by these rules conflict with each other, rules with larger serial numbers prevail. For example, if command A is prohibited by rule 1, and command B is prohibited by rule 2, but command A is allowed by rule 3, rule 2 and rule 3 finally take effect. Specifically, command A is allowed and command B is prohibited.
- Predefined rules for predefined roles in the system cannot be deleted or modified. If there is a conflict between system predefined rules and user-defined rules, user-defined rules prevail.
- Up to 256 rules can be configured for each role. A maximum of 1024 rules can be configured for all roles on the device.
- To configure command-based rules, follow the rules below:
  - Division of segments. To describe a multi-level mode command, divide the command character string into multiple segments by a semicolon (;). Each segment represents one or a series of commands. The command in the latter segment is used to execute the mode of a command in the preceding segment. A segment must contain at least one printable character.
  - Use of semicolons. To describe a multi-level mode command, separate the command segments with a semicolon. For example, the character string **config ; logging on** is used to grant a permission over the **logging on** command in configuration mode. The semicolon in the last command segment indicates that the permission is granted over the current mode command. For example, the character string **config ; interface *** is used to grant a permission over only the command to enter the interface configuration mode. The absence of a semicolon in the last command segment indicates that permissions are granted over the current command mode and all commands in this mode. For example, the character string **config ; interface *** is used to grant permissions over all commands in interface mode.
  - Use of asterisks. Each command segment contains at least one asterisk (*). An asterisk resides either in the middle or at both ends of a command segment. Each asterisk serves to fuzzily match a command. For example, the character string **config ; *** is used to grant permissions over all the commands in

configuration mode. The character string **config ; logging * flush** is used to grant a permission over a command starting with **logging** and ending with **flush** in configuration mode. The character string **config ; logging *** is used to grant permissions over all commands starting with **logging on** in configuration mode. When an asterisk resides in the middle of a command segment and the asterisk is used to match the command, the command is matched up to only the first asterisk in the middle, and the subsequent command segments are all considered matched. An execution command must be fully matched.

- Matching of keyword prefixes. A prefix matching algorithm is used for the matching between the command keyword and the command character string. That is, if the first several consecutive characters or all characters of a keyword in the command line match the keyword defined in a rule, the command line matches this rule. Therefore, a command character string may include a partial or complete command keyword. For example, if the rule **rule 1 deny command display ssh** is effective, the **display ssh** command and the **display ssh-session** command are disabled.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the role configuration mode.

role name *role-name*

(4) Configure rule permissions for a role.

rule *rule-number* { **permit** | **deny** } { **command** *command-string* | { **read** | **write** | **execute** }* { **feature** [*feature-name*] | **feature-group** *feature-group-name* } }

By default, predefined roles have predefined rule permissions while user-defined roles have no rule permission.

1.4.6 Configuring Description of a Role

1. Overview

This section describes how to configure the description for a role.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the role configuration mode.

role name *role-name*

(4) Configure the description for a role.

description *description*

By default, a predefined role is provided with a default description with the user-defined role is provided with no description.

1.4.7 Prohibiting a Role from Operating All Interface Resources

1. Overview

This section describes how to prohibit a role from -creating, deleting or applying all interface resources.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the role configuration mode.

role name *role-name*

(4) Prohibit a role from operating all interface resources and enter the role interface configuration mode.

interface policy deny

By default, a role has the permission to operate all interface resources.

1.4.8 Allowing a Role to Operate a Specific Interface Resource

1. Overview

This section describes how to allow a role to create, delete or apply a specific interface resource.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the role configuration mode.

role name *role-name*

(4) Prohibit a role from operating all interface resources and enter the role interface configuration mode.

interface policy deny

(5) Enable the role to operate a specific interface resource.

permit interface *interface-type interface-number-list*

By default, a role is prohibited from operating all interface resources.

1.4.9 Prohibiting a Role from Operating All VLAN Resources

1. Overview

This section describes how to prohibit a role from creating, deleting or applying all VLAN resources.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the role configuration mode.

role name *role-name*

(4) Prohibit a role from operating all VLAN resources and enter the role VLAN configuration mode.

vlan policy deny

By default, a role has the permission to operate all VLAN resources.

1.4.10 Allowing a Role to Operate a Specific VLAN Resource

1. Overview

This section describes how to allow a role to create, delete, and apply a specific VLAN resource.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the role configuration mode.

role name *role-name*

(4) Prohibit a role from operating all VLAN resources and enter the role VLAN configuration mode.

vlan policy deny

(5) Allow a role to operate a specific VLAN resource.

permit vlan *vlan-list*

By default, a role is prohibited from operating all VLAN resources.

1.4.11 Prohibiting a Role from Operating All VRF Resources

1. Overview

This section describes how to prohibit a role from creating, deleting or applying all VRF resources.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the role configuration mode.

role name *role-name*

(4) Prohibit a role from operating all VRF resources and enter the role VRF configuration mode.

vrf policy deny

By default, a role has the permission to operate all VRF resources.

1.4.12 Allowing a Role to Operate a Specific VRF Resource

1. Overview

This section describes how to allow a role to create, delete or apply a specific VRF resource.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the role configuration mode.

role name *role-name*

(4) Prohibit a role from operating all VRF resources and enter the role VRF configuration mode.

vrf policy deny

(5) Enable a role to operate a specific VRF resource.

permit vrf *vrf-name*

By default, a role is prohibited from operating all VRF resources.

1.5 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** command to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Run the **debugging** command to output various debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-2 RBAC Monitoring

Command	Purpose
display role [**name** *role-name*]	Displays the information about a specific role or all roles.
display role feature [**detail** | **name** *feature-name* [**detail**]]	Displays the basic information or details about a specific feature or all features.
display role feature-group [**detail** | **name** *group-name* [**detail**]]	Displays the basic information or details about a specific feature group or all feature groups.
debugging rbac	Debugs the RBAC module.

1.6 Configuration Examples

1.6.1 Configuring Role Permissions

1. Requirements

(1) Create the role **test**, set the description to "test role", and configure permissions to:

- Check all device information.
- Execute all commands of features snmpd and syslogd.
- Execute commands to create and delete interfaces, VLANs and VRF instances.
- Prohibit the role from operating all interface resources, but allow the role to operate interface VLAN 1.
- Prohibit the role from operating all VLAN resources, but allow the role to operate VLAN 1.
- Prohibit the role from operating all VRF resources, but allow the role to operate the VRF instance **test**.

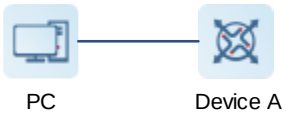
(2) Create a user with username **user**, password **user123**, and role **test**.

(3) Configure username and password-based authentication for telnet login.

(4) Let a user log in to a device from a PC in telnet mode, and perform authentication with username **user** and password **user123**. Upon login, the user is assigned with role **test** and has permissions of role **test**.

2. Topology

Figure 1-1 Configuring Role Permissions



3. Notes

- Enable the RBAC function.
- Create a role and configure its description.
- Configure role permissions to:

- Operate all **display** commands.
- Operate all read, write, and execution commands of features snmpd and syslogd.
- Execute interface, VLAN and VRF commands, as well as all commands in corresponding modes.
- Prohibit the role from operating all interface resources, but allow it to operate interface VLAN 1.
- Prohibit the role from operating all VLAN resources, but allow it to operate VLAN 1.
- Prohibit the role from operating all VRF resources, but allow it to operate VRF instance **test**.
- Create a user with username **user** and password **user123**, and assign the user with the role **test**.
- Configure username and password-based authentication for Telnet login.

4. Procedure

(1) Enable the RBAC function.

Enable the RBAC function for device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# role enable
```

5. Verification

View the role information.

```
DeviceA# display role name test
Role: test
  Description: test role
  Interface policy: deny
  Permit interfaces:
  VLAN1
  VLAN policy: deny
  Permit VLANs: 1
  Vrf policy: deny
  Permit vrfs: test
  ------------------------------------------------------------------
  Rule      Perm     Type   Scope              Entity
  ------------------------------------------------------------------
  1         permit           command            display *
  2         permit RWX       feature            snmpd
  3         permit RWX       feature            syslogd
  4         permit           command            config;interface *
  5         permit           command            config;vlan *
  6         permit           command            config;vrf definition *
  R:Read W:Write X:Execute
```

Run all read, write, and execution commands of features snmpd and syslogd.

```
DeviceA# display syslog
Syslog syslog: enabled
  Console syslog: level debugging, 46 messages logged
  Monitor syslog: level debugging, 19 messages logged
```

```
  Buffer syslog: level debugging, 46 messages logged
  Standard format:false
  Timestamp debugging messages: log-datetime
  Timestamp log messages: log-datetime
  Sequence-number log messages: disable
  Sysname log messages: disable
  Count log messages: disable
  Trap syslog: level informational, 46 message lines logged,0 fail
Log Buffer (Total 1048576 Bytes): have written 4462
*Oct 16 07:23:17: %CLI-6-STARTUP: Cli server process startup.
*Oct 16 07:23:17: %LOCALEAP-6-PKIMANAGE: Self-Signed PKI is activated
DeviceA# display snmp
Chassis: 60FF60
0 SNMP packets type
        0 Bad SNMP version errors
        0 Unknown community name
        0 Illegal operation for community name supplied
        0 Encoding errors
        0 Number of requested variables
        0 Number of altered variables
        0 Get-request PDUs
        0 Get-next PDUs
        0 Set-request PDUs
        0 Drop PDUs
        0 UDP parse errors
0 SNMP packets output
        0 Too big errors (Maximum packet size 1472)
        0 No such name errors
        0 Bad values errors
        0 General errors
        0 Response PDUs
        0 Trap PDUs
SNMP global trap: disabled
SNMP syslog: disabled
SNMP agent: enabled
SNMP v1: enabled
```

Run the interface access command. Then, a prompt for no permission appears.

```
DeviceA(cfg)# snmp-agent community test rw
DeviceA(cfg)# syslog server 1.1.1.1
DeviceA(cfg)# interface vlan 1
DeviceA(cfg-if-VLAN 1)# description test
DeviceA(cfg-if-VLAN 1)# exit
DeviceA(cfg)# interface gigabitethernet 0/1
% User doesn't have sufficient privilege-cmd to execute this command.
DeviceA(cfg)# vlan 1
DeviceA(cfg-vlan)# name test
```

```
DeviceA(cfg-vlan)# exit
DeviceA(cfg)# vlan 2
% User doesn't have sufficient privilege-cmd to execute this command.
DeviceA(cfg)# vrf definition test
DeviceA(cfg-vrf)# descriptive-name test
DeviceA(cfg-vrf)# exit
DeviceA(cfg)# vrf definition test1
% User doesn't have sufficient privilege-cmd to execute this command.
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
username user password user123
username user role test
!
role enable
!
description test role
 rule 1 permit command display *
 rule 2 permit read write execute feature snmpd
 rule 3 permit read write execute feature syslogd
 rule 4 permit command config;interface *
 rule 5 permit command config;vlan *
 rule 6 permit command config;vrf definition *
 interface policy deny
  permit interface VLAN 1
 vlan policy deny
  permit vlan 1
 vrf policy deny
  permit vrf test
!
line vty 0 4
 line-login local
```

Contents

1 Configuring Lines 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.2 Configuration Task Summary 1

1.3 Configuring the Number of VTY Terminals 1

1.3.1 Overview 1

1.3.2 Restrictions and Guidelines 1

1.3.3 Procedure 2

1.4 Configuring Line Attributes 2

1.4.1 Overview 2

1.4.2 Restrictions and Guidelines 2

1.4.3 Procedure 3

1.5 Configuring Terminal Attributes 6

1.5.1 Overview 6

1.5.2 Procedure 6

1.6 Monitoring 8

1.7 Configuration Examples 9

1.7.1 Configuring Line Attributes 9

1 Configuring Lines

1.1 Introduction

1.1.1 Overview

There are various types of terminal lines on network devices. You can group and manage terminal lines by types. Configurations of these terminal lines are called line configurations. On network devices, terminal lines are classified into multiple types such as CTY and virtual type terminal (VTY).

1.1.2 Principles

1. Basic Concepts

- CTY

  The CTY line is a line connected to the console port. Most network devices have one console port. You can access the local system through the console port.

- VTY

  A VTY line is a virtual terminal line not connected to any hardware. It is used for telnet or Secure Shell (SSH) connection.

1.2 Configuration Task Summary

Line configuration includes the following tasks.

The following tasks are optional. Select them based on your requirements.

- (Optional)Configuring the Number of VTY Terminals
- (Optional)Configuring Line Attributes
- (Optional)Configuring Terminal Attributes

1.3 Configuring the Number of VTY Terminals

1.3.1 Overview

This section describes how to enter the line configuration mode to configure other functions.

1.3.2 Restrictions and Guidelines

- The **login-line vty** command allows you to enter the VTY line configuration mode and specify the number of VTY connections.
- To reduce the number of available VTY connections, run the **unset login-line vty** *line-number* command.
- To configure the allowed maximum number of VTY connections, run the **login-line maximum-vty** command. If the allowed maximum number of VTY connections is set to **0**, all remote connections fail. If the allowed maximum number of VTY connections is set to a value less than the number of online remote connections, the configuration fails and a prompt appears.

Note

- Remote connections include telnet, SSH, and session connections.
- The allowed maximum number of VTY connections and the number of available VTY connections are separately managed. A remote connection is established successfully only when both conditions are met.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the maximum number of VTY connections.

line maximum-vty *max-number*

The maximum number of VTY connections is 36 by default.

(4) Enter the line configuration mode.

login-line { **console** | **vty | maximum-vty** } *first-line* [*last-line*]

- Increase or decrease VTY connections.

  line vty *line-number*

  There are five VTY connections, numbered from 0 to 4 by default. You can increase the number of VTY connections to 64, with new ones numbered from 5 to 63.

  VTY connections numbered from 0 to 4 are default connections of the system, which cannot be deleted. Only new connections can be deleted.

1.4 Configuring Line Attributes

1.4.1 Overview

This section describes how to configure line attributes in line configuration mode.

1.4.2 Restrictions and Guidelines

- When the absolute timeout time set for a line expires, the device immediately disconnects the line even if a user is operating the line terminal. Before the line is disconnected, the system displays the remaining time and a prompt, indicating that the terminal will exit.
- After a user acts as a dumb terminal to connect to the device through an asynchronous serial port, the user remotely logs in to the specified host through telnet specified in the **autocommand** command or obtains the specified application-based terminal service by running the **autocommand** command.
- The asynchronous hardware of the device generates seven data bits with parity check in flow communication mode. If parity is generated, specify seven data bits per character. If no parity is generated, specify eight data bits per character. Only early devices support five or six data bits, which are seldom used.
- By running the **flow-control** command, you can specify the flow control mode to keep the Tx rate of one end the same as the Rx rate of the peer end. Since terminals cannot receive data while sending data, flow control can prevent data loss. When high-speed data processing devices communicate with low-speed data

processing devices (for example, a printer communicates with a network port), you also need to enable flow control to prevent data loss. The device provides the following two flow control modes:

- Software flow control, also called soft flow control, which uses control keys for control. The default stop and start characters for software flow control are **Ctrl+S** (XOFF, ASCII value 19) and **Ctrl+Q** (XON, ASCII value 17), respectively. You can run the **stop-character** and **start-character** commands to configure them.
- Hardware flow control, also called hard flow control, which uses hardware for control.

- In telnet connection scenarios, you can run the **terminal-type** command to configure the type of terminals simulated on the terminal connected to a line as required. Upon telnet connection, one end negotiates with the other end about the terminal type based on its terminal type configuration (telnet negotiation ID: 0x18). For details, see RFC 854. You can run the **terminal-type vt100** command to restore the default terminal type.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of VTY connections.

line maximum-vty *max-number*

The maximum number of VTY connections is 36 by default.

(4) Enter the line configuration mode.

line { **console** | **vty** } *first-line* [*last-line*]

(5) (Optional) Configure an access control list (ACL) to control terminal login.

(IPv4 network)

access-apply { *acl-name* | *acl-number* } { **in** | **out** }

(IPv6 network)

ipv6 access-apply { *acl-name* | *acl-number* } { **in** | **out** }

No ACL is configured for controlling terminal login by default.

(6) (Optional) Enable accounting for a line. All the following configurations are optional and may be selected as needed.

- Enable command accounting for a line.

  aaa-config accounting cmd [*accounting-commands-level*] { *list-name* | **default** }

  The command accounting function is disabled by default.

- Enable user access accounting for a line.

  aaa-config accounting execute { *list-name* | **default** }

  The user access accounting function is disabled by default.

- Enable command authorization for a line.

  aaa-config authorization cmd *authorization-commands-level* { *list-name* | **default** }

The command authorization function is disabled by default.

(7) (Optional) Enable EXEC authorization for a line.

aaa-config authorization execute { **default** | *list-name* }

The EXEC authorization function is disabled by default.

(8) Configure the access to the CLI through a line.

exce

Accessing the CLI through lines is enabled by default.

(9) Configure line attributes. The following configurations are optional. Select them based on your requirements.

- Configure the absolute timeout time for a line.

  absolute-timeout *absolute-timeout-time*

  No absolute timeout time is configured for a line by default.

- Configure the character for activating a null terminal session.

  activation-character *ascii-value*

  The default character for activating a terminal session is the carriage return character (ASCII value 13).

- Enable automatic command execution.

  autocommand *autocommand-command*

  Automatic command execution for a line is disabled by default.

- Configure the hotkey for disconnecting terminal connections.

  session-terminate-character *ascii-value*

  The default hotkey for disconnecting terminal connections is **Ctrl+D** (ASCII value 4).

- Configure the character for exiting a line.

  exit-character *escape-value*

  The default character for exiting a line is **Ctrl+Shift+6** (ASCII value 30).

- Configure the access to the CLI through a line.

  exce

  Accessing the CLI through lines is enabled by default.

- Enable historical command recording for a specified line terminal and configure the allowed maximum number of recorded historical commands for the current terminal.

  command-history [**size** *size*]

  The historical command recording function is enabled by default, and the default number of recorded historical commands is **10**.

- Configure the maximum number of lines that are displayed on a single screen on a specified line terminal.

  length *screen-length*

  The maximum number of lines that are displayed on a single screen is **24** by default.

- Configure location description for a specific line.

  location *location*

No location description is configured for a specific line by default.

- Enable logging display on terminals.

  Monitor

  Logging display on terminals is disabled by default.

- Configure a privilege level for line-based login.

  privilege-cmd level *privilege-level*

  The default privilege level for line-based login is **1**.

- Configure the prompt for refusing line-based login.

  deny-message [*c message c*]

  No prompt is configured for refusing line-based login by default.

- Configure the baud rate for the current terminal.

  terminal speed *baudrate*

  The default baud rate is **9600 bps**. For S6920-4C, the default baud rate of terminals is **115,200 bps**.

- Configure the maximum number of columns that are displayed in a single line on the terminal connected to a specified line, that is, the line width.

  width *screen-width*

  The line width is **79** by default.

- Configure a role for a line.

  role *role-name*

  A role is configured for each line by default. The default role for serial ports, auxiliary ports, and extended serial ports is **network-admin**. The default role for other lines is **network-operator**.

⚠ Caution

Each line is configured with 1 to 64 roles. The last role cannot be deleted. If it is deleted, a failure prompt is displayed.

(10) Configure asynchronous line attributes. The following configurations are optional. Select them based on your requirements.

- Configure the number of data bits per character for asynchronous lines in flow communication mode.

  databits *bit*

  The default number of data bits per character for asynchronous lines in flow communication mode is **8**.

- Configure the CLI character encoding format for asynchronous lines.

  exec-character-bits { **7** | **8** }

  The system selects a full 8-bit ASCII character set as the CLI character set by default.

  To enter Chinese characters or display Chinese characters, images, or other international characters in the CLI, run the **exec-character-bits 8** command.

- Configure the flow control mode for asynchronous lines.

  flowcontrol { **hardware** | **none** | **software** }

No flow control is configured for asynchronous lines by default.

- Configure the parity bit for asynchronous lines.

  parity { **even** | **none** | **odd** }

  When using certain hardware (such as an asynchronous serial port and console port) for communication, you usually need to configure a parity bit.

- Configure the start character for software flow control for asynchronous lines.

  start-character *ascii-value*

  The default start character for software flow control for asynchronous lines is **Ctrl+Q** (ASCII value 17).

  After software flow control is enabled for an asynchronous line, the start character indicates the start of data transmission.

- Configure the stop character for software flow control for asynchronous lines.

  stop-character *ascii-value*

  The default stop character for software flow control for asynchronous lines is **Ctrl+S** (ASCII value 19).

  After software flow control is enabled for an asynchronous line, the stop character indicates the end of data transmission.

- Configure the number of stop bits in each byte transmitted through asynchronous lines.

  stopbits { **1** | **2** }

  The default number of stop bits in each byte transmitted through asynchronous lines is **2**.

  You should configure the stop bits for communication between an asynchronous line and the connected network device (such as a conventional numb terminal and modem).

- Configure the type of terminals simulated by an asynchronous line terminal.

  terminal-type *terminal-type-string*

  The default type of terminals simulated by a line is **vt100**.

1.5 Configuring Terminal Attributes

1.5.1 Overview

You can configure temporary terminal attributes in the CLI. These configurations are invalid after the connection is disconnected.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure terminal attributes. The following configurations are optional. Select them based on your requirements.

- Configure the number of data bits per character for asynchronous lines in flow communication mode.

  terminal databits *bit*

  The default number of data bits per character for the current terminal in flow communication mode is **8**.

- Configure the character for exiting the current terminal.

terminal exit-character *escape-value*

The default character for exiting the current terminal is **Ctrl+Shift+6** (ASCII value 30).

- Configure the CLI character encoding format for the current terminal.

  terminal exec-character-bits { **7** | **8** }

  To enter Chinese characters or display Chinese characters, images, or other international characters in the CLI, run the **terminal exec-character-bits 8** command.

- Configure the flow control mode for the current terminal.

  terminal flowcontrol { **hardware** | **none** | **software** }

  No flow control is configured for asynchronous lines by default.

- Enable historical command recording for the current terminal and configure the allowed maximum number of recorded historical commands.

  terminal command-history [**size** *size*]

  The historical command recording function is enabled by default, and the default number of recorded historical commands is **10**.

- Configure the maximum number of lines that are displayed on a single screen on the current terminal.

  terminal length *screen-length*

  The maximum number of lines that are displayed on a single screen is **24** by default.

- Configure location description for the current terminal.

  terminal location *location*

  No location description is configured for the current terminal by default.

- Configure the parity bit for the asynchronous line corresponding to the current terminal.

  terminal parity { **even** | **none** | **odd** }

  No parity bit is configured for the asynchronous line corresponding to the current terminal by default.

  When using certain hardware (such as an asynchronous serial port and console port) for communication, you usually need to configure a parity bit.

- Configure the baud rate for the current terminal.

  terminal speed *baudrate*

  The default baud rate is **9600 bps**. For S6920-4C, the default baud rate of terminals is **115,200 bps**.

- Configure the start character for software flow control for the current terminal.

  terminal start-character *ascii-value*

  The default start character for software flow control for the current terminal is **Ctrl+Q** (ASCII value 17).

- Configure the stop character for software flow control for the current terminal.

  terminal stop-character *ascii-value*

  The default stop character for software flow control for the current terminal is **Ctrl+S** (ASCII value 19).

- Configure the number of stop bits in each byte transmitted through the current terminal.

  terminal stopbits { **1** | **2** }

  The default number of stop bits in each byte transmitted through the current terminal is **2**.

- Configure the type of terminals simulated by the current terminal.

  terminal terminal-type *terminal-type-string*

  The default terminal type is **vt100**.

- Configure the maximum number of columns that are displayed in a single line on the current terminal, that is, the line width.

  terminal width *screen-width*

  The maximum number of columns that are displayed in a single line is **79** by default.

(3) Enter the global configuration mode.

config-system terminal

(4) Enter the line configuration mode.

login-line { **console** | **vty** } *first-line* [*last-line*]

(5) (Optional) Configure the login authentication timeout time for the current terminal.

line-login login timeout *response-timeout-time*

The default authentication timeout time for line-based login is **30** seconds.

(6) (Optional) Configure communication protocols supported by the current terminal.

protocol type { **total** | **ssh** | **telnet** | **none** }

All communication protocols are supported by default. That is, both SSH and telnet are supported.

(7) (Optional) Configure the prompt for terminal logout.

vacant-message [*c message c*]

No prompt is configured for terminal logout by default.

(8) (Optional) Configure the maximum number of columns that are displayed in a single line on the current terminal, that is, the line width.

width *screen-width*

The maximum number of columns that are displayed in a single line is **79** by default.

1.6 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup login-line { *line-number* | **console** *console-line-number* | **vty** *vty-line-number* }	Clears the connection status of a line.

Command	Purpose
cleanup command-history all-users	Clears historical records.
display line-info { *line-number* | **console** *console-line-number* | **vty** *vty-line-number* }	Displays the line configurations.
display command-history	Displays historical command records of a line.
display command-history all-users	Displays all historical command records.
display privilege-cmd	Display the privilege level of a line.
display user [**total**]	Displays login user information of a line.

1.7 Configuration Examples

1.7.1 Configuring Line Attributes

1. Requirements

When users use VTY connections to remotely log in to the device through telnet, the network administrator can configure VTY line attributes based on the use and security requirements of the device. Configure an ACL to control terminal login, set the privilege level of VTY line users to **14**, and require users of VTY line-based login to enter a password. Set the absolute timeout time of VTY connections to 30 minutes.

2. Topology

Figure 1-1 Configuring Line Attributes



3. Notes

- Set the maximum number of VTY line users to **6**.
- Configure an ACL to allow users of some IP addresses to access the device.
- Configure attributes of VTY line user.
- Configure the privilege level, login verification method, and verification password of VTY line users.

4. Procedure

Configure an ACL named **acl1** to control terminal login.

```
Device> privilege
Device# config-system terminal
Device(cfg)# ip access-control-list standard acl1
Device(cfg-stdacl)# permit 192.168.1.0 0.0.0.255
```

```
Device(cfg-stdacl)# exit
Device(cfg)# line v 0 6
Device(cfg-line)# access-apply acl1 in
```

Set the maximum number of VTY line users to **6**.

```
Device(cfg)# line v 0 6
```

Set the baud rate to 115200 bps.

```
Device(cfg-line)# speed 115200
```

Set the number of data bits to **8**.

```
Device(cfg-line)# databits 8
```

Set the parity bit to **even**.

```
Device(cfg-line)# parity even
```

Set the number of stop bits to **1**.

```
Device(cfg-line)# stopbits 1
```

Set the absolute timeout time of VTY connections to 30 minutes.

```
Device(cfg-line)# absolute-timeout 30
```

Configure software flow control.

```
Device(cfg-line)# flowcontrol software
```

Set the maximum number of lines that are displayed on a single screen to **100**.

```
Device(cfg-line)# length 100
```

Set the maximum number of columns that are displayed in a line to **256**.

```
Device(cfg-line)# width 256
```

Enable historical command recording and set the number of recorded historical commands to **200**.

```
Device(cfg-line)# command-history size 200
```

Set the privilege level of line-based login to **14**.

```
Device(cfg-line)# privilege-cmd level 14
```

Configure the user verification method.

```
Device(cfg-line)# aaa-config authentication local
Device(cfg-line)# exit
```

Configure a username.

```
Device(cfg)# username Device privilege-cmd 15 password test_1234
```

5. Verification

Users of the IP address 192.168.1.100 can log in to the device through tenet. Users beyond the IP addresses in the 192.168.1.0/24 network segment cannot log in to the device. Run the **display users** command to display users who log in to the device.

```
Device# display users
Line                User            Host(s)                Idle          Location
---------------- ------------ -------------------- ---------- ------------------
*   0 con 0         ---             idle                    00:00:00     ---
```

```
  1 vty 0          Device          idle                        00:00:04    192.168.1.100
```

Run the **display login-line vty** command to display the line status of the console.

```
Device# display line-info vty 1

VTY        Type      speed     Overruns
  1        VTY       115200    0
Line 2, Location: "", Type: "vt100"
Length: 100 lines, Width: 256 columns
Special Chars: Escape    Disconnect    Activation
                  ^^x         ^D              ^M
Timeouts:         Idle EXEC       Idle Session
                  00:10:00        never
History is enabled, command-history size is 200.
Total input: 0 bytes
Total output: 0 bytes
Data overflow: 0 bytes
stop rx interrupt: 0 times
```

6. Configuration Files

Device configuration file

```
hostname Device
!
username Device privilege-cmd 15 password test_1234
!
line console 0
 exec-timeout 0 0
 speed 115200
line console 0
 exec-timeout 0 0
 speed 115200
line vty 0 6
 absolute-timeout 30
 access-apply acl1 in
 privilege-cmd level 14
 authentication local
 password test_123
 flowcontrol software
 parity even
 stopbits 1
 width 256
 length 100
 monitor
 command-history size 200
 speed 115200
line vty 7 35
```

```
 access-apply acl1 in
 authentication
!
end
```

Contents

1 Configuring File System Management....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Basic Concepts....1
1.1.3 Way to Manage Files....2
1.2 Configuration Task Summary....2
1.3 Configuring a Directory....3
1.3.1 Displaying the Working Directory....3
1.3.2 Changing the Working Directory....3
1.3.3 Displaying Directory Information....3
1.3.4 Creating a Directory....3
1.3.5 Deleting a Directory....4
1.4 Configuring a File....4
1.4.1 Displaying File Content....4
1.4.2 Displaying File Information....4
1.4.3 Copying a File....4
1.4.4 Renaming a File....5
1.4.5 Deleting a File....5
1.4.6 Ejecting a USB Device....5
1.4.7 Erasing a File System....5
1.4.8 Configuring Prompt Level....6
1.5 Monitoring....6

1.6 Configuration Examples.....6

1.6.1 Configuring Basic Features of File System Management.....6

1 Configuring File System Management

1.1 Introduction

1.1.1 Overview

Files required for running a device, including configuration files and system software, are saved in the storage media of the device. File system management refers to the management of directories and files in storage media, including creation, deletion, modification and viewing of files.

1.1.2 Basic Concepts

The storage media supported by devices include fixed media (Flash) and pluggable media (USB flash drive). Each storage medium is called a file system.

1. Storage media and file system name

- The file system of Flash-type storage media is named as "flash:".
- The file system of USB drive-type storage media is named as "usb:".

2. Default file system

The file system used by default after a user logs in to the device when multiple storage media are available for the device.

3. Name of folder and file

The name of a folder or file can contain numbers, letters or special characters except asterisk (*), vertical bar symbol (|), backslash (\), slash (/), question mark (?), angle brackets (<>), quotation mark ("), and colon (:).

4. Root directory

The root directory is the default directory once the user logs in to the device. It is expressed by a slash (/). For example, "flash:/" indicates the root directory of Flash.

5. Working directory

The working directory is also called the current working directory. The default working directory is the root directory of Flash.

6. Path

The location of a file or folder.

Note

filesystem: specifies the uniform resource locator (URL) of a file system, followed by a colon (:). File systems include **flash:**, **usb:,** and **tmp:**. In addition, *directory* indicates a file name with the path, or specifies a path name. If the name starts with a slash (/), the path is an absolute path; otherwise, the path is a relative path.

1.1.3 Way to Manage Files

To manage files, log in to the system directly or use File Transfer Protocol (FTP) or Trivial File Transfer Protocol (TFTP).

Table 1-1 Way to Manage Files

Way to Manage Files	Application Scenario
Direct login to the system	A device fails to access information, and you need to repair the device or manage the files and directories in the device.
FTP	FTP is applicable to file transfer scenarios which do not require high security of network, and widely used for version upgrade.
TFTP	In laboratory LAN with good network conditions, use TFTP to load and upgrade versions online.

1.2 Configuration Task Summary

Configuration of file system management includes the following tasks: All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Directory
  - Displaying the Working Directory
  - Changing the Working Directory
  - Displaying Directory Information
  - Creating a Directory
- Configuring a File
  - Displaying File Content
  - Displaying File Information
  - Copying a File
  - copy source-url destination-url [*vrf-name*]

    copy *source-url dstination-url*
  - Renaming a File
  - Deleting a File
  - Ejecting a USB Device
  - Erasing a File System
  - Configuring Prompt Level

1.3 Configuring a Directory

1.3.1 Displaying the Working Directory

1. Overview

This feature displays the complete path of the current working directory.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Display the complete path of the current working directory.

pwd

1.3.2 Changing the Working Directory

1. Overview

This feature changes the current working directory.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Change the current working directory.

cd [*filesystem*:] [*directory*]

1.3.3 Displaying Directory Information

1. Overview

This feature displays the list of files and sub-directories under a directory.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Display the files under a directory.

dir [*filesystem*:] [*file-url*]

1.3.4 Creating a Directory

1. Overview

This feature creates a directory.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Create a directory.

mkdir [*filesystem*:] *directory*

1.3.5 Deleting a Directory

1. Overview

This feature deletes a directory.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Delete an empty directory.

rmdir [*filesystem*:] *directory*

1.4 Configuring a File

1.4.1 Displaying File Content

1. Overview

This feature displays the content of a file.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Display the content of a file.

more [**/ascii** | **/binary**] [*filesystem*:] *file-url*

1.4.2 Displaying File Information

1. Overview

This feature displays the information of a file.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Display the information of a file.

file [*filesystem*:] *file-url*

1.4.3 Copying a File

1. Overview

This feature copies a file.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Copy a file.

copy *source-url destination-url* [*vrf-name*]

copy *source-url dstination-url*

1.4.4 Renaming a File

1. Overview

This feature renames a file.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Rename a file or folder.

rename *source-url destination-url*

1.4.5 Deleting a File

1. Overview

This feature deletes a file.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Delete a file.

delete { [*filesystem*:] *file-url* | **bootup-config** }

delete [*filesystem*:] *file-url*

1.4.6 Ejecting a USB Device

1. Overview

This feature ejects a USB device.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Eject a USB device.

eject usb0

1.4.7 Erasing a File System

1. Overview

This feature erases a file system of a device.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Erase the file system of a USB device.

erase usb0

1.4.8 Configuring Prompt Level

1. Overview

This feature configures the prompt level for executing a file or folder.

- When the prompt level is set to **frequency**, the system asks you to confirm all the files.
- When the prompt level is set to **slient**, the system seldom gives a prompt.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure the prompt level for operating a file.

file prompt [**frequency** | **slient**]

1.5 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-2 File System Management Monitoring

Command	Purpose
display disk [**usb** | **flash**]	Displays the USB flash drive or Flash information.
display file system-info	Displays information about file systems.
display mount	Displays information about the file system mounted on the device.

1.6 Configuration Examples

1.6.1 Configuring Basic Features of File System Management

1. Requirements

Operate the files in the device after logging in to the device through the Console port or with a Telnet connection.

2. Notes

- Display the files and sub-directories under the current directory.
- Create the **test** directory.
- Copy the **config.text** file to the **test** directory and name it **test.text**.

3. Procedure

(1) Display the files and sub-directories under the current directory.

```
Device# dir flash:/
Directory of flash:/
Number    Properties        Size                    Time                          Name
------    ----------    ----------    ------------------------    --------------------
1           -rw-          9.9k            Fri Jan    3 02:49:39 2020    Hostname.text
2           -rw-          113.3M           Sat Jun 13 14:15:17 2020
3           -rw-          5.3k            Wed Jun 10 14:12:45 2020    cfgmpls
4           drwx           4.0k            Mon Jan    6 21:12:20 2020    startup
5           -rw-          2.6k            Sat Jun 13 16:13:12 2020    cfgpol
6           -rwx           1.6k            Thu May 21 13:39:30 2020    rsa1_private.bin
7           drwx           4.0k            Thu Nov 14 23:34:40 2019    vsd
8           -rw-          3.9k            Thu Jan    1 10:53:27 1970    virtual_switch.text
9           drwx           4.0k            Mon Jan    6 21:13:23 2020    uft
10            drwx           4.0k             Fri Nov 15 00:31:04 2019    licns
11            -rw-           7.5k             Wed Jun 30 10:33:50 2021    config.text
…
33 files, 22 directories
1,939,972,096 bytes data total (383,688,704 bytes free)
3,959,422,976 bytes flash total (383,688,704 bytes free)
```

(2) Create the **test** directory. Copy the **config.text** file to the **test** directory and name it **test.text**.

```
Device# mkdir test
```

(3) Copy the **config.text** file to the **test** directory and name it **test.text**.

```
Device# copy config.text flash:/test/test.text
Copying, press Ctrl+C to quit
!
Accessing flash:/config.text finished, 7652 bytes prepared
Flushing data to flash:/test/test.text...
Flush data done
Copy success.
```

4. Verification

Enter the **test** directory.

```
Device# cd test
```

Display the current working path.

```
Device# pwd
flash:/test
```

Display the files under the **test** directory.

```
Device# dir
Directory of flash:/test
Number    Properties        Size                    Time                          Name
------    ----------    ----------    ------------------------    --------------------
   1           -rw-          7.5k             Tue Jan    7 19:56:44 2020    test.text
1 file, 0 directories
1,939,972,096 bytes data total (383,676,416 bytes free)
```

```
3,959,422,976 bytes flash total (383,676,416 bytes free)
```

Contents

1 Configuring USB....1

1.1 Introduction....1

1.1.1 Overview....1

1.2 Configuration Task Summary....1

1.3 Using a USB Device....1

1.4 Monitoring....2

1.5 Configuration Examples....2

1.5.1 Configuring USB Basic Features....2

1 Configuring USB

1.1 Introduction

1.1.1 Overview

Universal Serial Bus (USB) is an external bus standard. In this document, it refers to peripheral devices in line with the USB standard, such as USB flash drives. USB devices are hot-swappable. They serve to copy files (such as configuration files and log files) from a communication device or copy external data (such as system upgrade files) to an internal storage device.

USB devices are applied in scenarios such as file system management and upgrade. The detailed scenarios are specified in configuration guides for different features. This document describes only the identification, viewing, use, and removal of USB devices.

1.2 Configuration Task Summary

USB configuration includes the following tasks:

(1) Using a USB

1.3 Using a USB Device

1. Overview

When you insert a USB device into a USB port, the system automatically finds the USB device. The driver module of the system, upon identifying the USB device, initializes the device and loads the file system in it. Then, the system reads and writes the USB device. Run commands (**cd**, **copy, delete** or **dir**) of the file system to operate the USB device.

2. Restrictions and Guidelines

- The system only uses products that support standard Small Computer System Interface (SCSI) commands (generally USB flash drives). Other products, including USB flash drives attached to USB network interface cards (NICs) and those with virtual USB drives, are not supported.
- A USB device supports file allocation tables (FATs) only. Other file systems can be used on a USB device only after they are converted to FATs through a PC.

3. Prerequisites

Insert a USB device.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Run commands of the file system to operate the USB device. Configure one of the following tasks.

- Enter the USB device partition.

cd usb: [*directory*]

The path name, if not specified, is the name of the root path of the USB device partition by default.

- Copy files between file systems.

  copy *source-url dstination-url*

  When the file to be copied exists on the target URL, the target file system determines the action, for example, reporting an error, overwriting the file, or asking users to make the choice.

- Delete a file in the USB device.

  delete usb: [*file-url*]

- Check whether the file is in the USB device.

  dir usb: [*file-url*]

  The path name, if not specified, is the name of the root path of the USB device partition by default.

Note

- When multiple partitions exist in the USB device, only the first FAT partition is accessible.
- The concept of parent directory is not available for the USB device path. After accessing the USB device by running **cd usb*:**, return to the flash file system by running **cd flash:**.

1.4 Monitoring

Run the **display** command to check the USB device.

Table 1-1 Monitoring

Command	Purpose
display usb	Displays the information about the inserted USB device.

1.5 Configuration Examples

1.5.1 Configuring USB Basic Features

1. Requirements

Operate the files in the device as follows after logging in to the device through the Console port or with a Telnet connection.

- Insert the USB flash drive into the port.
- Copy the **config.txt** file in the drive to **Flash**.
- Run the **eject usb0** command to remove the USB device.

2. Notes

- After you insert a USB flash drive into the port, the system displays a prompt of successful loading of the drive.
- Before removing the USB device, run the corresponding command to eject it. This is to avoid an error

resulted from the system using the device.

3. Procedure

(1) After you insert a USB flash drive into the port, the system finds the USB device and loads the drive. After that, the system prints the following prompt information:

```
*Jan    1 00:09:42: %USB-5-USB_DISK_FOUND: USB Disk <Mass Storage> has been inserted to USB port 0!
*Jan    1 00:09:42: %USB-5-USB_DISK_PARTITION_MOUNT: Mount usb0 (type:FAT32),size : 15789711360B(15789.7MB)
```

Note

Mass Storage specifies the name of the identified device. "**usb0**" specifies the first USB device. **Size** specifies the size of the partition. As shown above, the USB flash drive has a space of 15789.7 MB.

(2) Copy the **config.txt** file in the drive to Flash.

```
Hostname# dir usb0:/
Directory of usb0:/
    1 -rwx              4    Tue Jan    1 00:00:00 1980    fac_test
    2 -rwx              1    Mon Sep 30 13:15:48 2013    config.txt
2 files, 0 directories
15,789,711,360 bytes total (15,789,686,784 bytes free)
Hostname#
Hostname# copy usb0:/config.txt flash:/
Copying: !
Accessing usb0:/config.txt finished, 1 bytes prepared
Flushing data to flash:/config.txt...
Flush data done
```

(3) Remove the USB device.

```
Hostnam# eject usb 0
OK, now you can pull out the device 0.
```

4. Verification

Insert the USB flash drive to check the USB device.

```
Hostname> privilege
Hostname# display usb
Device: Mass Storage
ID: 0
URL prefix: usb0
Disk Partitions:
usb0(type:vfat)
Size:15789711360B(15789.7MB)
Available size:15789686784B(15789.6MB)
```

Run the **dir** command to check whether the **config.txt** file in the flash is copied successfully.

```
Hostname# dir flash:/
```

```
Directory of flash:/
   20 ----              1  Thu Jan  1 00:04:51 2020  config.txt
1 files, 0 directories
5,095,424 bytes total (4,960,256 bytes free)
```

Run the **display usb** command to check whether the USB device is removed. The device with ID 0 is not displayed when the **display usb** command is run.

```
Hostname# display usb
  No partition found.
```

5. Common Errors

- A device that uses non-standard SCSI is inserted into the USB port.
- The system cannot identify the USB device as the device is not a FAT.

Contents

1 Configuring Syslog 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Classification of System Logs 1
1.1.3 Levels of System Logs 1
1.1.4 Output Direction of System Logs 2
1.1.5 RFC3164 Log Format 2
1.1.6 RFC5424 Log Format 6
1.1.7 System Log Filtering 8
1.1.8 Log Reporting 9
1.1.9 Configuring System Log Monitoring 9
1.1.10 Protocols and Standards 10
1.2 Configuration Task Summary 10
1.3 Configuring Basic Syslog Features 10
1.3.1 Overview 10
1.3.2 Procedure 10
1.4 Configuring the System Log Format 11
1.4.1 Overview 11
1.4.2 Configuration Tasks 11
1.4.3 Configuring the RFC3164 Log Format 11
1.4.4 Configuring the RFC5424 Log Format 12
1.4.5 Configuring a Space Next to <pri> in the Log Format 12

1.5 Configuring Log Reporting 13
1.5.1 Configuration Tasks 13
1.5.2 Restrictions and Guidelines 13
1.5.3 Configuring Level-based Log Reporting 13
1.5.4 Configuring Delayed Log Reporting 14
1.5.5 Configuring Periodical Log Reporting 15
1.6 Configuring System Log Monitoring 16
1.6.1 Overview 16
1.6.2 Restrictions and Guidelines 16
1.6.3 Procedure 17
1.7 Configuring the Output Direction of System Logs 17
1.7.1 Configuration Tasks 17
1.7.2 Configuring the Output of System Logs to the Console 17
1.7.3 Configuring the Output of System Logs to the Monitor Terminal 18
1.7.4 Configuring the Function of Writing System Logs into the Memory Buffer 19
1.7.5 Configuring the Transmission of System Logs to the Log Server 19
1.7.6 Configuring the Function of Writing System Logs into Log Files 20
1.8 Configuring System Log Filtering 22
1.8.1 Overview 22
1.8.2 Restrictions and Guidelines 22
1.8.3 Procedure 22
1.9 Configuring System Log Redirection 23
1.9.1 Overview 23
1.9.2 Restrictions and Guidelines 23

1.9.3 Procedure....23
1.10 Configuring Performance Syslog Function....24
1.10.1 Overview....24
1.10.2 Restrictions and Guidelines....24
1.10.3 Procedure....24
1.11 Configuring Synchronization of User Input and Log Output....24
1.11.1 Overview....24
1.11.2 Restrictions and Guidelines....24
1.11.3 Procedure....24
1.12 Monitoring....25
1.13 Configuration Examples....25
1.13.1 Configuring the RFC3164 Log Format....25
1.13.2 Configuring the RFC5424 Log Format....29

1 Configuring Syslog

1.1 Introduction

1.1.1 Overview

If link status change or events such as receiving of exception packets and processing exception occur during the device running, a log packet in a fixed format is generated automatically. A log packet can be added with a timestamp and a sequence number, classified by the log priority, and output to the console, monitor terminal, log server, or other media. Logs are used by administrators to monitor the running status of the device, analyze the network conditions, and locate problems.

1.1.2 Classification of System Logs

System logs fall into two types:

- Common logs
- Debugging logs

Note

To generate debugging logs, you need to enable a debugging command. Such logs are used to locate problems, and may be ignored by users.

1.1.3 Levels of System Logs

Eight severity levels are defined for system logs in a descending order, including emergency, alert, critical, error, warning, notification, informational, and debugging, which correspond to eight numerical values from 0 to 7 respectively. A smaller value indicates a higher level. Table 1-1 describes the log levels.

Note

Only logs with a level equal to or higher than the specified level are output. For example, if the level of logs is set to informational (level 6), logs of level 6 or higher are output.

Table 1-1 Description of Log Levels

Keyword	Level	Description
emergencies	0	Indicates that the system cannot run normally.
alerts	1	Indicates that corrective measures must be taken immediately.
critical	2	Indicates a critical circumstance.
errors	3	Indicates an error message.
warnings	4	Indicates a warning.

Keyword	Level	Description
notifications	5	Indicates a common but important message that requires attention.
informational	6	Indicates an informational message.
debugging	7	Indicates debugging information.

1.1.4 Output Direction of System Logs

System logs can be output to the console, monitor, server, buffer, and file. The default level and type of logs vary with the output direction. You can customize filtering rules for different output directions. Table 1-2 describes output directions of system logs.

Table 1-2 Description of System Log Output Directions

Name of Output Direction	Default Output Direction	Default Output Level	Description
console	Console	debugging (Level 7)	Outputs logs and debugging information.
monitor	Monitor terminal	debugging (Level 7)	Outputs logs and debugging information to facilitate remote maintenance.
server	Log server	informational (Level 6)	Outputs logs and debugging information.
buffer	Log buffer	debugging (Level 7)	Outputs logs and debugging information. The log buffer is used to store system logs during the device running.
file	Log file	informational (Level 6)	Outputs logs and debugging information, and periodically writes logs in the log buffer into files.

1.1.5 RFC3164 Log Format

Formats of system logs may vary with the system log output direction.

1. Non-server Direction

When logs are output to the console, monitor terminal, log buffer, or log file, the system logs are in the following format:

```
seq no: *timestamp: sysname %module-level-mnemonic: content
```

The log format is described below:

Sequence number: *timestamp: system name %module name-severity-mnemonic: log text

For example, if you exit the configuration mode, the following log is displayed on the console:

```
001233: *May 22 09:44:36: Ruijie %SYS-5-CONFIG_I: Configured from console by console
```

2. Server Direction

When logs are output to the log server, the system logs are in the following format:

```
<priority>seq no: *timestamp: sysname %module-level-mnemonic: content
```

The log format is described below:

<Priority> sequence number: *timestamp: system name %module name-severity-mnemonic: log text

For example, if you exit the configuration mode, the following log is displayed on the log server:

```
<189>001233: *May 22 09:44:36: Ruijie %SYS-5-CONFIG_I: Configured from console by console
```

The following details each field in the log format:

- priority

  The priority is calculated using the following formula: Facility × 8 + Level. "Level" indicates the log level and "Facility" indicates the numerical code of the facility. The default facility value is local7 (23). Table 1-3 lists the value range of the facility.

Note

This field is valid only when logs are output to the log server.

Table 1-3 Description of the Log Priority Parameter

Numerical Code	Facility Keyword	Facility Description
0	kern	Kernel messages
1	user	User-level messages
2	mail	Mail system
3	daemon	System daemons
4	auth1	Security/Authorization messages
5	syslog	Messages generated internally by syslogd
6	lpr	Line printer subsystem
7	news	Network news subsystem
8	uucp	UUCP subsystem
9	clock1	Clock daemon
10	auth2	Security/Authorization messages
11	ftp	FTP daemon
12	ntp	NTP subsystem
13	logaudit	Log audit

Numerical Code	Facility Keyword	Facility Description
14	logalert	Log alert
15	clock2	Clock daemon
16	local0	Local use 0 (local0)
17	local1	Local use 1 (local1)
18	local2	Local use 2 (local2)
19	local3	Local use 3 (local3)
20	local4	Local use 4 (local4)
21	local5	Local use 5 (local5)
22	local6	Local use 6 (local6)
23	local7	Local use 7 (local7)

- seq no (sequence number)

  The sequence number of a system log is a 6-digit integer, and increases with system logs. By default, the sequence number is not displayed. You can run a command to display or hide this field.

- timestamp

  The timestamp records the generation time of a system log to help you check and locate system events. Ruijie devices support two formats of system log timestamps: datetime and uptime.

  - Datetime format

  The complete datetime format is as follows:

  ```
  Mmm dd yyyy hh:mm:ss.msec
  ```

  Table 1-4 describes each parameter of the timestamp in the datetime format.

  - uptime format:

  The complete uptime format is as follows:

  ```
  dd:hh:mm:ss
  ```

  Table 1-5 describes each parameter of the timestamp in the uptime format.

Table 1-4 Description of Parameters in the Timestamp in the Datetime Format

Timestamp Parameter	Parameter Name	Description
Mmm	Month	Indicates the abbreviated month. The 12 months in a year are abbreviated as Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, and Dec.
dd	Day	Indicates the day in the current month.
yyyy	Year	Indicates the year, not displayed by default.

Timestamp Parameter	Parameter Name	Description
hh	Hour	Indicates the hour.
mm	Minute	Indicates the minute.
ss	Second	Indicates the second.
msec	Millisecond	Indicates the millisecond.

Note

By default, the timestamp in the datetime format in system logs does not contain the year and milliseconds. You can run a command to display or hide the year and millisecond in the timestamp in the datetime format.

Table 1-5 Description of Parameters in the Timestamp in the Uptime Format

Timestamp Parameter	Parameter Name	Description
dd	Day	Indicates the day in the current month.
hh	Hour	Indicates the hour.
mm	Minute	Indicates the minute.
ss	Second	Indicates the second.

Note

If the device does not have a real time clock (RTC), which is hardware used to record the absolute time of the system, the device uses its startup time (in the uptime format) as the system log timestamp by default. If the device has an RTC, the device uses its absolute time (in the datetime format) as the system log timestamp by default.

- sysname (system name)

  This field records the name of the device that generates a log so that the log server can identify the host of such device. By default, the field is not displayed. You can run a command to display or hide this field.

- module (module name)

  This field indicates the name of the module that generates a log. The value is an upper-case string of 2 to 20 characters, which can contain upper-case letters, digits, and underscores. The **module** field is mandatory for common logs by default, and optional for debugging logs.

- level (log level)

  Eight system log levels from 0 to 7 are defined. The level of system logs generated by each module is determined during development and cannot be changed.

- mnemonic

  This field indicates a summary of the generation of a log. The value is an upper-case string of 4 to 32 characters, which can contain upper-case letters, digits, and underscores. The **mnemonic** field is mandatory for common logs by default, and optional for debugging logs.

- content (log text)

  This field indicates the content of a system log.

1.1.6 RFC5424 Log Format

All system logs are in the following format regardless of their output directions:

```
<priority>version timestamp sysname MODULE LEVEL MNEMONIC [structured-data] description
```

The log format is described below:

<Priority>version timestamp system name module name level mnemonic structured parameter area log content

For example, if you exit the configuration mode, the following log is displayed on the console:

```
<133>1 2013-07-24T12:19:33.130290Z Ruijie SYS 5 CONFIG - Configured from console by console
```

The following details each field:

- priority

  The priority is calculated using the following formula: Facility × 8 + Level. "Level" indicates the log level and "Facility" indicates the facility value. The facility value can be set during log configuration. When the RFC5424 log function is enabled, the default facility value is local0 (16).

- version

  According to RFC5424, the version is always **1**.

- timestamp

  The timestamp records the generation time of a system log to help you check and locate system events. The devices use the following uniform timestamp format when the RFC5424 syslog function is enabled:

  ```
  YYYY-MM-DDTHH:MM:SS.SECFRACZ
  ```

  Table 1-6 describes each parameter.

Table 1-6 Description of Timestamp Parameters

Timestamp Parameter	Parameter Name	Description
YYYY	Year	Indicates the year.
MM	Month	Indicates the month in the current year.
DD	Day	Indicates the day in the current month.
T	Separator	A date must end with "T".
HH	Hour	Indicates the hour.
MM	Minute	Indicates the minute.
SS	Second	Indicates the second.

Timestamp Parameter	Parameter Name	Description
SECFRAC	Millisecond	Indicates the millisecond (1–6 digits).
Z	End mark	Time must end with "Z".

- sysname (system name)

  This field indicates the name of the device that generates a log so that the log server can identify the host of such device.

- MODULE (module name)

  This field indicates the name of the module that generates a log. The value is an upper-case string of 2 to 20 characters, which can contain upper-case letters, digits, and underscores. The **module** field is mandatory for common logs by default, and optional for debugging logs.

- LEVEL (log level)

  Eight system log levels from 0 to 7 are defined. The level of system logs generated by each module is determined during development and cannot be changed.

- MNEMONIC

  This field indicates a summary of the generation of a log. The value is an upper-case string of 4 to 32 characters, which can contain upper-case letters, digits, and underscores. The **mnemonic** field is mandatory for common logs by default, and optional for debugging logs.

- structured-data (structured parameter area)

  This field is introduced to RFC5424, to describe log parameters in a way that helps device parsing. Each log can contain 0 or multiple parameters. If no parameter exists, the placeholder (-) must be used. Each parameter is in the following format:

  ```
  [SD_ID@enterpriseID PARAM-NAME=PARAM-VALUE]
  ```

  Table 1-7 describes each parameter.

Table 1-7 Description of Structured Parameters

Structured Parameter	Parameter Name	Description
SD_ID	Name of parameter information	The name of parameter information is capitalized, and must be unique in a log.
@	Separator	"@enterpriseID" is necessary for the customized parameters, but not for parameters defined in RFC5424.
enterpriseID	Enterprise ID	The enterprise ID is maintained by the Internet Assigned Numbers Authority (IANA). The enterprise ID of the devices is 4881. You can query the enterprise ID at the official website of IANA.

Structured Parameter	Parameter Name	Description
PARAM-NAME	Parameter name	The parameter name is capitalized, and must be unique in the structured parameter area of a log.
PARAM-VALUE	Parameter value	The parameter value must be enclosed in double quotation marks. Values of the IP address and MAC address must be capitalized, and values of other parameters are capitalized as required.

- description (log text)

  This field indicates the content of a system log.

1.1.7 System Log Filtering

By default, the logs generated by the system are output in all directions. The device provides the log filtering function, which allows you to filter logs by log output direction, log keyword, and matching rule. When you do not care about some logs or you care about some logs only, use the log filtering function to filter the logs.

- Filtering direction

  Four log filtering directions are defined:

  - **buffer**: Filters the logs sent to the log buffer, that is, the logs displayed by the **display syslog** command.
  - **file**: Filters the logs written into syslog files.
  - **server**: Filters the logs sent to the log server.
  - **terminal**: Filters the logs sent to the console and monitor terminal (including telnet and SSH).

Note

The four filtering directions can be used either collectively (logs in various directions are filtered) or separately (logs in one direction are filtered).

- Filtering Mode

  Two filtering modes are available:

  - **contains-only**: Indicates that only the logs that contain keywords specified in the filtering rules are output. When you care about some logs only, you can apply the contains-only mode on the device to output only the logs that match the filtering rules on the terminal. Thus, you can check whether any event occurs.
  - **filter-only**: Indicates that the logs that contain keywords specified in the filtering rules are filtered and are not output. Too many logs from a module may result in spamming on the terminal CLI. If you do not care about this type of logs, you can apply the filter-only mode and configure filtering rules to filter such logs.

Note

The two filtering modes are mutually exclusive, that is, you can configure only one filtering mode at a time.

- Filtering Rule

  Two log filtering rules are available:

  - **exact-match**: If exact-match is selected, you must select all the three filtering options (log module, log level, and log mnemonic). To filter a specific log, use the exact-match filtering rule.
  - **single-match**: If single-match is selected, you only need to select one of the three filtering options (log module, log level, and log mnemonic). To filter a specified type of log, use the single-match filtering rule.

Note

If the same module name, log level, or mnemonic is configured in both the single-match and exact-match rules, the single-match rule prevails.

1.1.8 Log Reporting

The log reporting functions falls into level-based reporting, delayed reporting, and periodical reporting, which are described as follows:

- Level-based reporting

  You can perform the level-based policy function to send logs of modules and severity levels to different destinations.

  For example, you can configure a command to send OSPF module logs of level 4 or lower to the log server, and OSPF module logs of level 5 or higher to local log files.

- Delayed reporting

  Delayed log reporting means that instead of being directly sent to the log server, system logs generated in the system are buffered in the log files in the device and then periodically sent to the log server. Delayed log reporting can be configured to reduce the packet transmission and interaction frequency between the device and the log server. In this way, the performance pressure on the device and the log server and the burden of the intermediate network is reduced.

- Periodical reporting

  Logs about device performance statistics are periodically sent. All timers for periodically sending logs are managed by the syslog module. When a timer expires, the syslog module calls the log processing function registered by each module to display and output the performance statistic logs to the remote syslog server in real time. The server analyzes these logs to evaluate the device performance.

Caution

- To configure log reporting, enable the RFC5424 log format function; otherwise, you cannot configure level-based reporting, delayed reporting, and periodical reporting.
- When the RFC5424 log format is enabled, logs can be output in all directions, and delayed reporting is enabled by default. At the same time, the periodical reporting function is disabled.

1.1.9 Configuring System Log Monitoring

System log monitoring means that the system monitors external connections to the device and records logs.

- After user login/logout syslog is enabled, the system records the user's connections to the device. The recorded information includes the login username and source address.

- After user operation syslog is enabled, the system records device configuration changes. The recorded information includes the operation username, source address, and operation content.

1.1.10 Protocols and Standards

- RFC 3164: The BSD syslog Protocol
- RFC 5424: The_Syslog_Protocol

1.2 Configuration Task Summary

Syslog configuration includes the following tasks:

(1) Configuring Basic Syslog Features

(2) (Optional) Configuring the System Log Format. "Configuring the RFC3164 Log FormatConfiguring the RFC3164 Log Format " and "Configuring the RFC5424 Log Format" are mutually exclusive.

- Configuring the RFC3164 Log Format
- Configuring the RFC5424 Log Format
- (Optional) Configuring a Space Next to <pri> in the Log Format.

(3) Configuring Log Reporting Configure at least one of the functions.

- Configuring Level-based Log Reporting
- Configuring Delayed Log Reporting
- (Optional) Configuring Periodical Log Reporting

(4) (Optional) Configuring System Log Monitoring

(5) (Optional) Configuring the Output Direction of System Logs. Configure at least one of the output directions.

- Configuring the Output of System Logs to the Console
- Configuring the Output of System Logs to the Monitor Terminal
- Configuring the Function of Writing System Logs into the Memory Buffer
- Configuring the Transmission of System Logs to the Log Server
- Configuring the Function of Writing System Logs into Log Files

(6) (Optional) Configuring System Log Filtering

(7) (Optional) Configuring System Log Redirection

(8) (Optional) Configuring Performance Syslog Function

(9) (Optional) Configuring Synchronization of User Input and Log Output

1.3 Configuring Basic Syslog Features

1.3.1 Overview

This section describes how to enable the syslog function so that the system processes logs and users view the logs generated by the device.

1.3.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable syslog.

syslog on

By default, syslog is enabled.

1.4 Configuring the System Log Format

1.4.1 Overview

By configuring system log display, you can adjust the display format of system logs.

1.4.2 Configuration Tasks

System log format configuration includes the following tasks. "Configuring the RFC3164 Log Format" and "Configuring the RFC5424 Log Format " are mutually exclusive. Configure either of the tasks.

- Configuring the RFC3164 Log Format
- Configuring the RFC5424 Log Format
- Configuring a Space Next to <pri> in the Log Format

1.4.3 Configuring the RFC3164 Log Format

1. Restrictions and Guidelines

- If the device does not have an RTC, which is hardware used to record the absolute time of the system, the system uses the device startup time (in the **uptime** format) as the log timestamp. In this case, the configured device time is invalid. If the device has an RTC, the system uses the device time (in the **datetime** format) as the log timestamp.
- Add a system name to logs to identify the device that generates a system log after the system log is sent to the server.
- Add a sequence number to logs. The sequence number starts from 1. After the sequence number is added, the system can monitor the loss and the generation sequence of logs. The log sequence number is a 6-digit integer. Each time a log is generated, the sequence number increases by one. When the sequence number increases from 1 to 1,000,000, or reaches 2^32, it starts from 000000 again.
- After the RFC3164 format is enabled, the **syslog delay-send**, **syslog policy**, and **syslog statistic** commands that are applicable to the RFC5424 format will become invalid and be hidden.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the timestamp format for system logs.

service syslog-timestamp [*message-type* [**uptime** | **datetime** [**msec**] [**year**]]]

The system logs use the datetime timestamp format, and the timestamp does not contain the year and milliseconds by default.

(4) (Optional) Add the system name to system logs.

service sysname

By default, system logs do not contain the system name.

(5) (Optional) Add the sequence number to system logs.

service syslog-seq-numbers

By default, system logs do not contain the sequence number.

(6) Configure the system log format. The configurations below are mutually exclusive. Configure one of the log formats.

- Set the system log format to the standard log format.

  service standard-syslog

- Set the system log format to the private log format.

  service private-syslog

1.4.4 Configuring the RFC5424 Log Format

1. Restrictions and Guidelines

- After the RFC5424 log format is enabled, a uniform timestamp format is adopted, and **uptime** and **datetime** are not differentiated.
- In the RFC5424 log format, the timestamp may or may not contain the time zone. Currently, only the timestamp without the time zone is supported.
- After the RFC5424 format is enabled, the **service sequence-numbers**, **service sysname**, **service timestamps**, **service private-syslog**, and **service standard-syslog** commands that are applicable to the RFC3164 format will become invalid and be hidden.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Set the system log format to the RFC5424 log format.

service log-format rfc5424

1.4.5 Configuring a Space Next to <pri> in the Log Format

1. Restrictions and Guidelines

- if there is no special requirement, configure a space on the device that needs a space next to <pri> in the log format.
- The default format of logs sent to a server are as follows:

```
<pri>*timestamp: %module-level-mnemonic: content
```

The order is as follows:

```
<pri>*timestamp: %module-level-mnemonic: content
```

If the function is enabled, the format of logs sent to a server are as follows:

```
<pri> *timestamp: %module-level-mnemonic: content
```

Compared with the default log format, a space is added next to <pri>.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a space next to <pri> in the log format.

syslog priv-format private

1.5 Configuring Log Reporting

1.5.1 Configuration Tasks

The log reporting configuration includes the following tasks. Configure at least one of the tasks.

- Configuring Level-based Log Reporting
- Configuring Delayed Log Reporting
- (Optional) Configuring Periodical Log Reporting

1.5.2 Restrictions and Guidelines

The log reporting function can be configured and takes effect only when the RFC5424 format is enabled.

1.5.3 Configuring Level-based Log Reporting

1. Overview

You can configure a level-based log reporting policy to output logs of different modules and severity levels to different destinations.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Set the system log format to the RFC5424 log format.

service log-format rfc5424

(4) Configure a level-based log reporting policy.

syslog policy module *module-name* [**not-lesser-than**] *policy-level* **direction** { **all** | **buffer** | **console** | **file** | **monitor** | **server** }

No level-based log reporting policy is configured by default.

1.5.4 Configuring Delayed Log Reporting

1. Overview

Delayed log reporting means that, instead of being directly sent to the log server, system logs generated in the system are buffered in the log files in the device and then periodically sent to the log server. Delayed log reporting can be configured for system logs to reduce the message transmission and interaction frequency between the device and the log server. In this way, the performance pressure on the device and the syslog server and the burden of the intermediate network is reduced.

2. Restrictions and Guidelines

- Generally, you are not advised to output logs reported in a delayed manner to the console or remote terminal. Otherwise, many logs reported in a delayed manner are displayed, increasing the burden on the device.
- The configured file name cannot contain any dot (.) because the system automatically adds the index and the file name extension (**.txt**) to the file name when generating a locally buffered file. The index increases each time a new file is generated. In addition, the file name cannot contain characters prohibited by the file system of your PC, such as \, /, :, *, ", <, >, and |. For example, the configured file name is **log_server**, the current index is 5, the file size is 1000 bytes, and the source IP address of the device sending the log file is 10.2.3.5. The name of the log file sent to the remote server is **log_server_1000_10.2.3.5_5.txt** while the name of the log file stored on the device is **log_server_5.txt**. If the source IP address is an IPv6 address, the colon (:) in the IPv6 address must be replaced by the hyphen (-) because the colon (:) is prohibited by the file system. For example, the file name is **log_server**, the current file index is 6, the file size is 1000 bytes, and the source IPv6 address of the device sending the log file is 2001::1. The name of the log file sent to the remote server is **log_server_1000_2001-1_6.txt** while the name of the log file stored on the device is **log_server_6.txt**.
- If few logs are generated on the device, you are advised to set the interval of delayed log reporting to a large value so that more logs are sent to the remote server at a time.
- By default, the log file sent to the remote server is named **File size_device IP address_index.txt**. If the name of the file for delayed log reporting is modified, the log file sent to the remote server is named **Configured file name prefix_file size_device IP address_index.txt**. The file stored on the local flash space of the device is named **Configured file name prefix_index.txt**. The default file name prefix is **syslog_ftp_server**, the interval for delayed log reporting is 3600s (1 hour), and the log file size is 128 KB.
- The maximum interval for delayed log reporting is 65535s (18 hours). If you set the interval for delayed log reporting to a larger value, the size of logs generated in this period may exceed the file size (128 KB). To prevent loss of logs, the logs are written into a new log file, and the index increases by 1. When the timer expires, all log files buffered in this period are sent to the log server at a time.
- The flash space for buffering local log files on the device is limited. Therefore, up to eight log files are buffered on the device. If more than eight log files are buffered on the device, all the log files generated earlier are sent to the log server at a time.
- You can send logs to the log server through File Transfer Protocol (FTP) or Trivial File Transfer Protocol (TFTP). The addresses of up to five log servers are configured for one device. Either FTP or TFTP is

specified for each server.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Set the system log format to the RFC5424 log format.

service log-format rfc5424

(4) Configure delayed log reporting. The following configurations are optional. Configure at least one of them as actually needed.

- Configure delayed log reporting to the console and remote terminal.

  syslog delay-send terminal

  Delayed log reporting to the console and remote terminal is disabled by default.

- Configure the name of the file for delayed log reporting.

  syslog delay-send file flash: *delay-send-filename*

  The default format of the log file name is file size_device IP address_index.txt.

- Configure the interval for delayed log reporting.

  syslog delay-send interval *delay-send-interval*

  The interval for delayed log reporting is 3600s (1 hour) by default.

- Configure the server address and reporting mode.

  syslog delay-send server [**oob**] { *ipv4-address* | **ipv6** *ipv6-address* } [**vrf** *vrf-name*] [**via** *mgmt-name*] **mode** { **ftp user** *username* **password** [**0** | **7**] *password* | **tftp** }

  Delayed log reporting is disabled by default.

1.5.5 Configuring Periodical Log Reporting

1. Overview

This section describes how to configure periodical log reporting so that the server can collect all the logs on the device at the same time point.

2. Restrictions and Guidelines

- The interval of periodical log reporting and the function of outputting logs to the console and remote terminal take effect only after periodical log reporting is enabled.
- You are advised to disable the function of outputting periodically reported logs to the console and remote terminal. Otherwise, when the reporting timer expires, many performance statistics logs are displayed, increasing the burden on the device.
- To ensure that the server collects all performance statistics logs from the device at the same time point, the timers of all statistical objects are restarted when you modify the interval of one statistic object.
- The default interval of periodical log reporting is 15 minutes. To enable the server to collect all performance statistics logs from the device at the same time point, you need to set the log reporting interval of different

statistic objects to be a multiple of another interval. Currently, the interval can be set to 0, 15, 30, 60, or 120. Here, **0** indicates that periodical log reporting is disabled.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Set the system log format to the RFC5424 log format.

service log-format rfc5424

(4) Configure periodical log reporting.

syslog statistic enable

Periodical log reporting is disabled by default.

(5) (Optional) Configure periodical log reporting to the console and remote terminal.

syslog statistic terminal

Periodical log reporting to the console and remote terminal is disabled by default.

(6) (Optional) Configure the interval for periodical log reporting.

syslog statistic mnemonic *mnemonic* **interval** *logging-statistic-interval*

The default interval of periodical log reporting is 15 min.

1.6 Configuring System Log Monitoring

1.6.1 Overview

System log monitoring enables the system to monitor the external connections to the device and record logs.

1.6.2 Restrictions and Guidelines

- If both the **syslog userinfo** command and the **syslog userinfo command-log** command are configured on the device, only the results shown by the **syslog userinfo command-log** command are displayed when you run the **display current-config** command.
- The **syslog userinfo** command is run in global configuration mode to enable login/logout syslog. After this function is configured, the device displays logs when users access the devices through telnet, Secure Shell (SSH), or Hypertext Transfer Protocol (HTTP) so that the administrator monitors the device connection status.
- The **syslog userinfo command-log** command is run in global configuration mode to enable user operation syslog. After this function is configured, the system displays related logs to notify the administrator of configuration changes.
- User operations are logged when commands are configured and run. By default, the device does not generate operation logs when a user modifies the device configuration. If the 5424 log format is configured, that is, the **service log-format rfc5424** command is configured, you need to configure the **syslog delay-send terminal** command so that operation logs are output to the terminal (because delayed log reporting is registered for operation logs).

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure system log monitoring. The following configurations are all optional. Configure at least one of them as actually needed.

- Enable user login/logout syslog.

  syslog userinfo

  User login/logout syslog is disabled by default.

- Enable user operation syslog.

  syslog userinfo command-log

  User operation syslog is disabled by default.

1.7 Configuring the Output Direction of System Logs

1.7.1 Configuration Tasks

System log output direction configuration includes the following tasks. Configure at least one of the tasks.

- (Optional) Configuring the Output of System Logs to the Console
- (Optional) Configuring the Output of System Logs to the Monitor Terminal
- (Optional) Configuring the Function of Writing System Logs into the Memory Buffer
- (Optional) Configuring the Transmission of System Logs to the Log Server
- (Optional) Configuring the Function of Writing System Logs into Log Files

1.7.2 Configuring the Output of System Logs to the Console

1. Overview

This section describes how to configure the output of system logs to the console so that the device can output logs generated by the system to the console. Then, administrator can monitor the running status of the system.

2. Restrictions and Guidelines

- If too many system logs are generated, you can limit the syslog rate to reduce logs output to the console.
- By default, system syslog is enabled. You are advised not to disable it. If too many system logs are displayed, you can configure the level of logs to be displayed on different devices to reduce the logs displayed.
- The **syslog count** command is run to enable the log statistics function in global configuration mode. After this function is enabled, the system records the number of times logs are generated by each module and the generation time of the last log.
- The default level of logs that are displayed on the console is debugging (Level 7). You can run the **display syslog config** command in privileged EXEC mode to display the level of logs that are displayed on the

console.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable log statistics collection.

syslog count

Log statistics collection is disabled by default.

(4) (Optional) Configure the level of logs that are output to the console.

syslog console [*severity-level*]

The default level of logs that are displayed on the console is 7 (debugging information).

(5) (Optional) Configure syslog rate limiting.

syslog rate-limit [**console**] [**all**] *number* [**except** [*severity-level*]]

Syslog rate limiting is disabled by default.

1.7.3 Configuring the Output of System Logs to the Monitor Terminal

1. Overview

This section describes how to configure the output of system logs to the monitor terminal so that the device can output system logs to the remote monitor terminal. Then, the administrator can monitor the running status of the system.

2. Restrictions and Guidelines

- If too many system logs are generated, you can limit the syslog rate to reduce logs output to the monitor terminal.
- By default, the current monitor terminal is not allowed to output logs after you remotely connect to the device. You need to manually run the **terminal monitor** command to allow the current monitor terminal to display logs. The **terminal monitor** command is valid only for the current connection. When the terminal reconnects to the device, the default settings of this command are restored**.**

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enable log display on the current monitor terminal.

terminal monitor

Log display in the window of the monitor terminal is disable by default.

(3) Enter the global configuration mode.

config-system terminal

(4) (Optional) Configure the level of logs to be output to the monitor terminal.

syslog monitor [*severity-level*]

The default level of logs that are displayed in the window of the monitor terminal is 7 (debugging information).

1.7.4 Configuring the Function of Writing System Logs into the Memory Buffer

1. Overview

This section describes how to configure the function of writing system logs into the memory buffer so that the device can write generated system logs into the memory buffer. Then, the administrator can view recent system logs by running the **display syslog** command.

2. Restrictions and Guidelines

- If the buffer is full, earlier logs are overwritten when system logs are written into the memory buffer.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the parameters for the memory buffer, into which logs are to be written.

syslog buffered [*buffer-size*] [*severity-level*]

The buffer size is 1 mega-byte and the log severity level is 7 by default.

1.7.5 Configuring the Transmission of System Logs to the Log Server

1. Overview

This section describes how to configure the device to transmit generated system logs to the log server so that the administrator can monitor device logs on the server.

2. Restrictions and Guidelines

- The log timestamp or sequence number function must be enabled. Otherwise, the logs are not sent to the log server.
- If the device has an MGMT interface and is connected to the log server through the MGMT interface, you must add the **oob** option (indicating that system logs are sent to the log server through the MGMT interface) when configuring the **syslog server** command. You can use the **via** parameter only when **oob** is specified in the command. In this case, **/vrf** is unavailable.
- The **syslog server** command is used to specify the address of the log server that receives logs. You can specify multiple log servers, and logs are sent simultaneously to all these log servers.
- Up to five log servers are configured for a Ruijie product.
- When a domain name is entered to configure a log server, the **syslog hostname** command is disabled.
- To track and manage logs, you can use the **syslog source interface** command to set the source IP address of all log packets to the IP address of an interface. Thus, the administrator can identify the device that sends the logs based on the unique address. If this source interface is not configured or the IP address is not configured for this source interface, the source IP address of log packets is the IP address of the interface

that sends the log packets.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the transmission of logs to a specified log server.

syslog server [**oob**] { *hostname* | *ipv4-address* | **ipv6** *ipv6-address* } [**via** *mgmt-name*] [**udp-port** *port-number*] [**vrf** *vrf-name*] [**facility** facility-type] [**level** inform-level]

syslog { *ipv4-address* | **ipv6** *ipv6-address* } [**vrf** *vrf-name*] [**udp-port** *port-number*] [**facility** facility-type] [**level** inform-level]

(4) (Optional) Configure the level of logs that are sent to the log server.

syslog trap [*severity-level*]

The default level of logs that are sent to the log server is 6 (informational messages).

(5) (Optional) Configure the facility value for logs to be sent to the log server.

syslog facility *facility-type*

When the RFC5424 log format is enabled, the default facility value is 16 (Local0, Local use); when the RFC5424 log format is disabled, the default facility value is 23 (Local7, Local use).

(6) (Optional) Configure the source interface for logs to be sent to the log server.

syslog source interface *interface-type interface-number*

No log source address is configured by default, and the source IP address of the log packets sent to the server is the IP address of the interface that sends the log packets.

(7) (Optional) Configure the source address for logs to be sent to the log server.

- Configure an IPv4 source address for the logs to be sent to the log server.

  syslog source ip *ipv4-address*

- Configure an IPv6 source address for the logs to be sent to the log server.

  syslog source ipv6 *ipv6-address*

No log source address is configured by default, and the source IP address of the log packets sent to the server is the IP address of the interface that sends the log packets.

1.7.6 Configuring the Function of Writing System Logs into Log Files

1. Overview

This section describes how to configure the function of writing system logs into log files so that the device can save generated system logs to the log files for viewing. Logs are saved in a log file buffer before being saved to the log file. The system writes the content in the log file buffer to the log file at the specified frequency, and you can also save the logs to a log file manually. After the logs are saved, the content in the log file buffer will be cleared. When logs need to be saved, the device automatically generates log files.

2. Restrictions and Guidelines

- System logs are not immediately written into log files. They are first buffered in the memory buffer, and then written into log files either periodically (at an interval of 1 hour by default) or when the buffer is full.
- When no log file is generated and the remaining space of the storage medium is insufficient, the device does not save the newly generated logs to a log file. Therefore, you must regularly clean up the storage space of the storage medium to ensure the log file function.
- The **syslog file** command is used to create a log file with a specified file name on a specified file storage device. The file size increases with logs, but cannot exceed the configured value of **max-file-size**. If the value of **max-file-size** is not specified, the default size of a log file is 128 KB.
- After the **syslog file** command is configured, the system saves logs to log files. A log file name does not contain any file type extension. The log file name extension is always **txt**, which cannot be changed.
- After this function is configured, logs will be written into log files at an interval of 1 hour. If you have run the **syslog file flash:syslog** command, 16 log files are created, such as **syslog.txt**, **syslog_1.txt**, **syslog_2.txt**, …, **syslog_14.txt**, and **syslog_15.txt**. Logs are overwritten into the 16 log files in sequence and cyclically. For example, the system writes logs into **syslog_1.txt** after **syslog.txt** is fully occupied. When **syslog_15.txt** is fully occupied, logs are written into **syslog.txt** again.
- If no extended flash space is available, the **syslog file flash** command is automatically hidden and is not configured. If no FLASH2 is available, the **syslog file flash2** command is automatically hidden and is not configured. If FLASH2 is available and the **syslog file flash** command is configured, logs are recorded in FLASH2.
- The system will not delete the generated log files after the number of log files is modified. Therefore, to save the space of the extended flash space, you need to manually delete the log files generated in the system (before deletion, you can transfer the log files to an external server through TFTP). For example, 16 log files are created by default after the function of writing logs into log files is enabled. If the device has generated 16 log files and if you want to change the number of log files to 2, new logs are overwritten into the log files with the index of 0 and 1 by turns. The existing log files with the index of 2 to 16 are retained. You can manually delete these log files as needed.
- After the time-based log storage is enabled, the system writes logs of the same level that are generated in the same day into the same log file. The log file is named **yyyy-mm-dd_filename_level.txt**, where **yyyy-mm-dd** indicates the absolute time of the day when the logs are generated, **filename** indicates the log file name configured by the **syslog file flash** command, and **level** indicates the log level.
- After you specify the storage time for logs of a level, the system will delete the logs once the storage time expires. For better network management, the storage time ranges from 7 days to 365 days.
- If the time-based log storage is not enabled, logs are stored based on the file size to support old configuration commands.
- After the **syslog flash flush** command is configured, the logs in the buffer are immediately written into a log file.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the parameters for the log file, into which logs are written.

syslog file { **sata0:***filename* | **flash:***filename* | **flash2:***filename* | **usb0**:*filename* | **usb1**:*filename* | **sd0**:*filename* } [*max-file-size*] [*inform-level*]

(4) Configure the number of log files.

syslog file numbers *file-numbers*

The default number of log files is 16.

(5) Configure the interval for writing logs into log files.

syslog flash interval *log-write-flash-interval*

By default, logs are written into flash files at an interval of 3600s.

(6) Configure the storage time of logs written into log files.

syslog life-time level *inform-level life-time-days*

No storage time is configured by default. The storage time depends on the configured log file size.

(7) Configure the function of immediately writing logs in the buffer into log files.

syslog flash flush

1.8 Configuring System Log Filtering

1.8.1 Overview

By default, all the logs generated by the system are displayed on the console or other terminals. By configuring log filtering, the network administrator can filter the generated system logs, select only the required logs to be displayed, or have the logs displayed on a specified terminal.

1.8.2 Restrictions and Guidelines

- Two filtering modes are available: **contains-only** and **filter-only**, which are mutually exclusive. You can configure only one filtering mode at a time.
- Log filtering rules fall into **exact-match** and **single-match**. The **single-match** rule prevails over the **exact-match** rule. If the same module, mnemonic, or information level is configured in both the **single-match** and **exact-match** rules, the logs complying with the **single-match** rule are filtered first.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the log filtering direction. The following configurations are optional. Select at least one of the configurations as actually needed.

- Configure the log filtering direction.

  syslog filter direction { **all** | **buffer** | **file** | **server** | **terminal** }

  Logs sent to all the directions are filtered by default, namely, **all** is set.

- Configure the log filtering mode.

  syslog filter type { **contains**-**only** | **filter**-**only** }

  The filtering type is set to **filter-only** by default.

(4) Configure the log filtering rule.

syslog filter rule { **exact-match module** *module-name* **mnemonic** *mnemonic-name* **level** *inform-level* | **single-match** { **level** *inform-level* | **mnemonic** *mnemonic-name* | **module** *module-name* } }

No log filtering rule is configured by default.

1.9 Configuring System Log Redirection

1.9.1 Overview

In a virtual switching unit (VSU) environment, you can configure the system log redirection function to redirect the logs of the slave device to the active device so that the network administrator can manage the logs uniformly.

1.9.2 Restrictions and Guidelines

- The system log redirection function is used in a VSU active-standby environment only, and does not apply to the standalone environment.
- You are advised to run the **syslog rd rate-limit** command to limit the rate of the logs redirected to the active device. This is to limit the number of the logs that are redirected from the slave device to the active device per second. This prevents considerable logs on the slave device from burdening the system.
- In a VSU environment, logs on the slave device can be displayed in its console window, redirected to the active device for display in the console or virtual type terminal (VTY) window of the active device, or stored in the memory buffer of the active device, extended flash space or syslog server.
- In a VSU environment composed of box-type devices, after the log redirection function is enabled, logs on the slave or standby device will be redirected to the active device for output, and the role flag string "(*device ID)" will be added to the beginning of each log, to indicate that the log is a redirected log. Assume that four devices form a VSU. The ID of the active device is 1, the ID of the slave device is 2, and the IDs of two standby devices are 3 and 4. No role flag string is added to logs generated by the active device. The role flag string (*2) is added to logs redirected from the slave device to the active device. The role flag strings (*3) and (*4) are added respectively to logs redirected from the two standby devices to the active device.
- In a VSU environment composed of chassis-type devices, after the log redirection function is enabled, logs generated by the slave or standby Supervisor Engine will be redirected to the master Supervisor Module for output, and a role flag string "(*device ID/Supervisor Engine name)" will be added to each log to indicate that a log is a redirected log. If four Supervisor Engines are used in a VSU, the role flag strings are as follows: (*1/M1), (*1/M2), (*2/M1), and (*2/M2).

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the log redirection function.

syslog rd on

(4) (Optional) Configure the rate limit on redirected logs.

syslog rd rate-limit *number* [**except** [*severity-level*]]

The log redirection function limits the maximum number of logs to be redirected per second to 200 by default.

1.10 Configuring Performance Syslog Function

1.10.1 Overview

After the performance syslog function is enabled, the logs that are output through the performance syslog interface will be transmitted through the performance syslog channel.

1.10.2 Restrictions and Guidelines

The performance syslog function needs to be configured only when massive logs are displayed on the server within a short period. Only several functional services require this function.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the performance syslog function.

syslog performance switch

The performance syslog function is disabled by default.

1.11 Configuring Synchronization of User Input and Log Output

1.11.1 Overview

When the synchronization of user input and log output is enabled, even if logs are displayed during user input, the user input information is displayed after display, thereby ensuring the input integrity and continuity.

1.11.2 Restrictions and Guidelines

The **syslog synchronous** command is configured in line configuration mode, and on each line that enables this function.

1.11.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the specified line configuration mode.

line { **console** | **vty** } *first-line* [*last-line*]

(4) Synchronize user input and log output.

syslog synchronous

Synchronization of user input and log output is disabled by default.

1.12 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-8 Syslog Monitoring

Command	Purpose
cleanup syslog	Clears logs in the memory buffer.
display syslog	Displays log statistics and logs in the memory buffer based on the timestamp from earliest to latest.
display syslog reverse [**timestamp** *YY MM DD hh:mm:ss*]	Displays log statistics and logs in the memory buffer based on the timestamp from latest to earliest.
display syslog last *number*	Displays log statistics, and the specified number of logs in the memory buffer based on the timestamps from latest to earliest.
display syslog config	Displays system log configurations and statistics.
display syslog count	Displays log statistics of each module in the system.

1.13 Configuration Examples

1.13.1 Configuring the RFC3164 Log Format

1. Requirements

The network administrator can check system logs to learn about the operation status of the device, better understand and manage the device, or locate problems.

2. Topology

Figure 1-1 Configuring the RFC3164 Log Format



3. Notes

- Configure the L3 network reachable between the device and the server.
- Enable syslog.
- Configure the log format.
  - Set the system log format to RFC3164 format.
  - Configure the log display format.
  - Set the timestamp format to datetime and add the millisecond and year to the timestamp.
  - Add the system name to the system log format.
  - Add the sequence number to the system log format.
- Configure log monitoring.
  - Configure log statistics collection.
  - Set the rate of outputting system logs to the console to 50 logs per second.
  - Configure the display of logs on the monitor terminal.
- Synchronize user input and log output.
- Enable performance syslog.
- Configure log filtering.
  - Set the filtering directions of logs to **terminal** and **server**.
  - Set the log filtering mode to **filter-only**.
  - Set the log filtering rule to **single-match** to filter the logs with a module name containing "SYS".
- Configure the output direction of system logs.
  - Configure the function of writing system logs into a log file named **syslog**.
  - Configure the transmission of system logs to the log server with the IPv4 address of 192.168.1.20.
  - Configure the size of the system logs to be written into the memory buffer to 128 KB (131072 bytes).

4. Procedure

(1) Configure the management IP address for the device.

```
Device> privilege
Device# config-system terminal
Device(cfg)# interface gigabitethernet 0/1
Device(cfg-if-gigabitethernet 0/1)# ip address 192.168.1.20 255.255.255.0
Device(cfg-if-gigabitethernet 0/1)# exit
```

(2) Enable syslog.

```
Device(cfg)# syslog on
```

(3) Configure the system log format.

Set the system log format to RFC5424 log format.

```
Device(cfg)# no service log-format rfc5424
```

Set the log display format to standard log format.

```
Device(cfg)# service standard-syslog
```

Set the timestamp format to datetime and add the millisecond and year to the timestamp.

```
Device(cfg)# service timestamps log datetime year msec
Device(cfg)# service timestamps debugging datetime year msec
```

Add the system name to the system log format.

```
Device(cfg)# service sysname
```

Add the sequence number to the system log format.

```
Device(cfg)# service sequence-numbers
```

(4) Configure log monitoring.

```
Device(cfg)# syslog userinfo
Device(cfg)# syslog userinfo command-log
```

Configure log statistics collection.

```
Device(cfg)# syslog count
```

Configure the output of system logs to the console.

```
Device(cfg)# syslog console informational
```

Set the rate of outputting system logs to the console to 50 logs per second.

```
Device(cfg)# syslog rate-limit console 50
```

Configure the output of system logs to the monitor terminal.

```
Device(cfg)# syslog monitor informational
Device(cfg)# line vty 0 4
Device(cfg-line)# monitor
```

(5) Synchronize user input and log output.

```
Device(cfg-line)#syslog synchronous
Device(cfg-line)#exit
```

(6) Enable performance syslog.

```
Device(cfg)# syslog performance switch
```

(7) Configure log filtering.

Set the filtering directions of logs to **terminal** and **server**.

```
Device(cfg)# syslog filter direction server
Device(cfg)# syslog filter direction terminal
```

Set the log filtering mode to **filter-only**.

```
Device(cfg)# syslog filter type filter-only
```

Set the log filtering rule to **single-match** to filter the logs with a module name containing "SYS".

```
Device(cfg)# syslog filter rule single-match module SYS
```

(8) Configure the output direction of system logs.

Configure the function of writing system logs into a log file named **syslog**.

```
Device(cfg)# syslog file flash:syslog debugging
Device(cfg)# syslog flash interval 600
```

Configure the transmission of system logs to the log server with the IPv4 address of 192.168.1.20.

```
Device(cfg)# syslog server 192.168.1.20
```

Configure the size of the system logs to be written into the memory buffer to 128 KB (131072 bytes).

```
Device(cfg)# syslog buffered 131072 informational
```

5. Verification

Run the **ping** command to check whether the L3 route between the server and the log server is reachable.

```
Device# ping 192.168.1.20
Sending 5, 100-byte ICMP Echoes to 192.168.1.20, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/3 ms
```

Run the **display syslog config** command to check the configuration.

```
Device# display syslog config
Syslog syslog: enabled
  Console syslog: level informational, 82 messages logged
  Monitor syslog: level informational, 0 messages logged
  Buffer syslog: level debugging, 83 messages logged
  File syslog: level debugging, 83 messages logged
  File name:syslog.txt, size 128 Kbytes, the 1 file is currently being written
  Standard format:true
  Timestamp debugging messages: datetime
  Timestamp log messages: datetime
  Sequence-number log messages: enable
  Sysname log messages: enable
  Count log messages: enable
  Trap syslog: level debugging, 82 message lines logged,8 fail
    syslog to   192.168.1.20
```

6. Configuration Files

Device configuration file

```
hostname Device
!
service standard-syslog
service sysname
service sequence-numbers
service timestamps debugging datetime msec year
```

```
service timestamps log datetime msec year
syslog filter direction server
syslog filter direction terminal
syslog filter rule single-match module SYS
syslog rate-limit console 50
syslog count
syslog userinfo command-log
syslog buffered 131072 informational
syslog file flash:syslog debugging
syslog flash interval 600
syslog console informational
syslog monitor informational
syslog server 192.168.1.20
syslog performance switch
!
interface gigabitethernet 0/1
 ip address 192.168.1.10 255.255.255.0
!
end
```

1.13.2 Configuring the RFC5424 Log Format

1. Requirements

The network administrator can check system logs to learn about the operation status of the device, better understand and manage the device, or locate problems.

2. Topology

Figure 1-2 Configuring the RFC5424 Log Format



3. Notes

- Configure the L3 network reachable between the device and the server.
- Enable syslog
- Set the system log format to RFC5424 format.
- Configure log reporting.
  - Configure level-based log reporting.
  - Configure delayed log reporting.
  - Configure periodical log reporting.
- Configure log monitoring.
  - Configure the log statistics collection.

- Set the rate of outputting system logs to the console to 50 logs per second.
- Configure the display of logs on the monitor terminal.

- Synchronize user input and log output.
- Enable performance syslog.
- Configure log filtering.
  - Set the filtering directions of logs to **terminal** and **server**.
  - Set the log filtering mode to **filter-only**.
  - Set the log filtering rule to **single-match** to filter the logs with a module name containing "SYS".
- Configure the output direction of system logs.
  - Configure the function of writing system logs into a log file named **syslog**.
  - Configure the transmission of system logs to the log server with the IPv4 address of 192.168.1.20.
  - Configure the size of the system logs to be written into the memory buffer to 128 KB (131072 bytes).

4. Procedure

(1) Configure the management IP address for the device.

```
Device> privilege
Device# config-system terminal
Device(cfg)#interface gigabitethernet 0/1
Device(cfg-if-gigabitethernet 0/1)# ip address 192.168.1.10 255.255.255.0
Device(cfg-if-gigabitethernet 0/1)# exit
```

(2) Enable syslog

```
Device(cfg)# syslog on
```

(3) Set the system log format to RFC5424 format.

```
Device(cfg)# service log-format rfc5424
```

(4) Configure log reporting.

Configure delayed log reporting.

```
Device(cfg)# syslog delay-send terminal
Device(cfg)# syslog delay-send interval 7200
Device(cfg)# syslog delay-send file flash:syslog_Ruijie
Device(cfg)# syslog delay-send server 192.168.1.20 mode ftp user admin password admin
```

Configure level-based log reporting.

```
Device(cfg)# syslog policy module SYS not-lesser-than 5 direction console
Device(cfg)# syslog policy module SYS 3 direction buffer
```

Configure periodical log reporting.

```
Device(cfg)# syslog statistic enable
Device(cfg)# syslog statistic terminal
Device(cfg)# syslog statistic mnemonic TUNNEL_STAT interval 30
```

(5) Configure log monitoring.

```
Device(cfg)# syslog userinfo
```

```
Device(cfg)# syslog userinfo command-log
```

Configure log statistics collection.

```
Device(cfg)# syslog count
```

Configure the output of system logs to the console.

```
Device(cfg)# syslog console informational
```

Set the rate of outputting system logs to the console to 50 logs per second.

```
Device(cfg)# syslog rate-limit console 50
```

Configure the output of system logs to the monitor terminal.

```
Device(cfg)# syslog monitor informational
Device(cfg)# line vty 0 4
Device(cfg-line)# monitor
```

(6) Synchronize user input and log output

```
Device(cfg-line)# syslog synchronous
Device(cfg-line)# exit
```

(7) Enable performance syslog.

```
Device(cfg)# syslog performance switch
```

(8) Configure system log filtering.

Set the filtering directions of logs to **terminal** and **server**.

```
Device(cfg)# syslog filter direction server
Device(cfg)# syslog filter direction terminal
```

Set the log filtering mode to **filter-only**.

```
Device(cfg)# syslog filter type filter-only
```

Set the log filtering rule to **single-match** to filter the logs with a module name containing "SYS".

```
Device(cfg)# syslog filter rule single-match module SYS
```

(9) Configure the output direction of system logs.

Configure the function of writing system logs into a log file named **syslog**.

```
Device(cfg)# syslog file flash:syslog debugging
Device(cfg)# syslog flash interval 600
```

Configure the transmission of system logs to the log server with the IPv4 address of 192.168.1.20.

```
Device(cfg)# syslog server 192.168.1.20
```

Configure the size of the system logs to be written into the memory buffer to 128 KB (131072 bytes).

```
Device(cfg)# syslog buffered 131072 informational
```

5. Verification

(1) Run the **ping** command to check whether the L3 route between the server and the log server is reachable.

```
Device# ping 192.168.1.20
Sending 5, 100-byte ICMP Echoes to 192.168.1.20, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
```

```
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/3 ms.
```

(2) Run the **display syslog config** command to check the configuration result.

```
Device# display syslog config
Syslog syslog: enabled
  Console syslog: level informational, 84 messages logged
  Monitor syslog: level informational, 0 messages logged
  Buffer syslog: level debugging, 92 messages logged
  File syslog: level debugging, 105 messages logged
  File name:syslog.txt, size 128 Kbytes, the 1 file is currently being written
  Statistic log messages: enable
  Statistic log messages to terminal: enable
  Delay-send log messages to terminal: enable
  Delay-send file name:syslog_Ruijie, Current write index:0, Current send index:0, Cycle:7200 seconds
  Count log messages: enable
  Trap syslog: level debugging, 84 message lines logged,10 fail
    syslog to   192.168.1.20
  Delay-send syslog: 0 message lines logged
    syslog to   192.168.1.20   by ftp
```

6. Configuration Files

Device configuration file

```
hostname Device
!
service log-format rfc5424
syslog filter direction server
syslog filter direction terminal
syslog filter rule single-match module SYS
syslog rate-limit console 50
syslog count
syslog userinfo command-log
syslog buffered 131072 informational
syslog file flash:syslog debugging
syslog flash interval 600
syslog console informational
syslog monitor informational
syslog facility local7
syslog server 192.168.1.20
syslog performance switch
syslog policy module SYS not-lesser-than 5 direction console
syslog policy module SYS 3 direction buffer
syslog statistic enable
syslog statistic terminal
syslog statistic mnemonic TUNNEL_STAT interval 30
syslog delay-send terminal
syslog delay-send interval 7200
```

```
syslog delay-send file flash:syslog_Ruijie
syslog delay-send server 192.168.1.20 mode ftp user admin password admin
!
interface gigabitethernet 0/1
  ip address 192.168.1.10 255.255.255.0
!
end
```

Contents

1 Configuring Software Upgrade 1

1.1 Introduction 1

1.2 Restrictions and Guidelines 1

1.3 Principles of System Version Upgrade 1

1.4 Principles of Patch Version Upgrade 4

1.4.1 Basic Concepts 4

1.4.2 Patch Package Management 4

1.4.3 Patch Auto-Sync 4

1.5 Configuration Task Summary 4

1.6 Configuring System Version Upgrade 5

1.6.1 Configuration Tasks 5

1.6.2 Configuring Subsystem Upgrade/Degrade 5

1.6.3 Configuring Version Auto-sync 7

1.7 Configuring Patch Version Upgrade 8

1.7.1 Overview 8

1.7.2 Configuration Restrictions and Guidance 8

1.7.3 Configuration Tasks 8

1.7.4 Installing a Patch Package 8

1.7.5 Activating a Patch Package by One Click 9

1.7.6 Uninstalling a Patch Package 10

1.8 Monitoring 11

1.9 Configuration Examples 12

1.9.1 Configuring Subsystem Upgrade....12
1.9.2 Configuring Auto-sync Upgrade....14
1.9.3 Configuring Patch-based Upgrade....15
1.9.4 Uninstalling a Patch Package....18
1.9.5 Installing a Patch Package by One Click....20

1 Configuring Software Upgrade

1.1 Introduction

Adopting a modular structure, Ruijie General Operating System (RGOS) supports overall system upgrade and subsystem upgrade, as well as the upgrade through patches. The package management module of RGOS is used to install, query, and maintain components of the device. By upgrading the software of the device, users can install software that is more stable or contains more features in the system.

1.2 Restrictions and Guidelines

If the password encryption type is 7 or 8, when the device version is downgraded to a version that does not support the AES128/SHA256 encryption algorithm, the password may not be identified. Therefore, before the device is downgraded, you must reconfigure the password as a plaintext password or a type 7 password ciphertext generated on a lower version device.

1.3 Principles of System Version Upgrade

1. Basic Concepts

- Subsystem

  A subsystem exists on a device in the form of images. The RGOS subsystems include:

  - Uboot: After being powered on, the device loads and runs the Uboot subsystem. This subsystem initializes the device, and loads and runs system images.
  - Rboot: This subsystem is a set of dedicated programs used to install or upgrade the main program.
  - Basic Input Output System (BIOS): This subsystem is dedicated for devices running the X86 architecture.
  - Main program: This subsystem is a collection of Operating Systems (OSs) and applications in the system.
  - CPLD: complex programmable logic device (CPLD) on the device.
  - MCU: microcontroller unit (MCU) on the device, which is dedicated for S6510-48VS8CQ and S6250-48XS8CQ.
  - Optical module: Optical interface on the device.

- Main package

  The main package is often used to upgrade/degrade subsystems of box-type devices. The main package is a combination package of the Uboot, Rboot, and main program subsystems. The main package is used for overall system upgrade/degradation.

- Rack package

  A rack package is used to upgrade/degrade subsystems of rack-type devices. It contains the Main packages of the supervisor modules, switch fabric modules, and line cards. The rack package allows all the modules on a rack-type device to be upgraded once for all.

- BIOS package

The BIOS package is used to upgrade the BIOS of the device. It is used on the devices running the X86 architecture only.

- CPLD package

  The CPLD package is used to upgrade the CPLD of the device.

- MCU package

  The MCU package is used to upgrade the MCU of the device. This package is used only for S6510-48VS8CQ and S6250-48XS8CQ.

- Optical module package

  The optical module package is used to upgrade the optical module of the device.

⚠ Caution

An "installation package" in this document refers to installation files of all the subsystems of the device.

2. Upgrading/Degrading and Managing Subsystems

Subsystem upgrade/degradation aims to update the software functions by replacing the subsystem components in the device with the ones in an installation package. Redundancy design is adopted for subsystems, so subsystems of the device are often not directly replaced with the subsystems in the package during upgrade/degradation. Instead, subsystems are added to the device and then activated.

- Upgrade/Degradation

  Subsystems exist on the device in different forms. Therefore, they are upgraded or degraded differently.

  - Uboot: Generally, this subsystem exists in the NOR flash memory of the device in the form of images. It is upgraded/degraded by writing images into the NOR flash memory of the device.
  - Rboot: Generally, this subsystem exists in the NOR flash memory of the device in the form of images. It is upgraded/degraded by writing images into the NOR flash memory of the device.
  - BIOS: Generally, this subsystem exists in the flash memory of the device in the form of images. It is upgraded/degraded by writing images into the flash memory of the device.
  - Main program: Generally, this subsystem exists in the NAND flash memory of the device in the form of images. It is upgraded/degraded by writing images into the NAND flash memory of the device.
  - CPLD: Generally, this subsystem exists in the CPLD in the form of images. It is upgraded/degraded by using the CPLD firmware to upgrade/degrade the CPLD.
  - MCU: Generally, this subsystem exists in the MCU in the form of images. It is upgraded/degraded by using the MCU firmware to upgrade/degrade the MCU.

- Management

  During upgrade of the subsystem component of the main program, the upgrade/degradation module always records the subsystem component in use, the redundant subsystem components, and management information of various versions in the configuration file. The module queries the available subsystem components and then loads subsystem components as needed.

  The redundancy design is adopted for each subsystem component. During the upgrade/degradation:

  - Uboot: This subsystem contains a master boot subsystem and a slave boot subsystem. Only the master boot subsystem is upgraded, and the slave boot subsystem is always redundant.

- Rboot: This subsystem contains one program. More redundancies are allowed if there is enough space.
- BIOS: This subsystem contains a master BIOS and a slave BIOS. The two BIOSs can be upgraded.
- Main program: One redundancy is allowed if there is enough space.
- CPLD: This subsystem contains one program. More redundancies are allowed if there is enough space.
- MCU: This subsystem contains one program. More redundancies are allowed if there is enough space.
- Optical module: Generally, this subsystem exists in the optical interface in the form of images. It is upgraded/degraded by using the optical module firmware.

- Upgrade

  Put upgrade files in the local device, and then run the **upgrade** command.

3. Auto-Sync Upgrade

Auto-sync upgrade aims to coordinate multiple subsystems (boards, chassis, and virtual switching unit (VSU) system) within one system. During software upgrade, the system automatically sends an upgrade package to all matched members, or upgrades the software version of new members to the specified version according to the specified upgrade policy to ensure that the software version of the entire system meets expectations.

- Select one from the following auto-sync upgrade policies:
  - Check the software version of a new member, and upgrade the member if its version is earlier than the version of the system installation package. Otherwise, no upgrade is performed.
  - Do not auto-sync new members.
  - Upgrade new members to the version of the system upgrade package regardless of the current version of the new members.
- Auto-sync upgrade check is performed in the following scenarios:
  - During upgrade, if no upgrade target is specified, the upgrade function module sends the upgrade package to all matched members (including modules, chassis, and VSU system) for auto-sync upgrade.
  - When the device is started, the version of every member is checked, and auto-sync upgrade is performed according to the version check result and auto-sync upgrade policy.
  - When a new member is added, the version of the member is checked and auto-sync upgrade is performed according to the version check result and auto-sync upgrade policy.
- Management

  To perform auto-sync upgrade, configure an auto-sync upgrade policy and install an upgrade package.
- Configuring an auto-sync upgrade policy

  To perform the expected upgrade when a new member is added to the system, check the auto-sync upgrade configuration, such as preparation of the installation package.

  If a module is not checked because the system is not configured with an auto-sync upgrade policy, you can upgrade the specific members.
- Restriction

  The upgrade package used for auto-sync upgrade can be a main package or rack package.

1.4 Principles of Patch Version Upgrade

1.4.1 Basic Concepts

- Patch package

  A patch package contains several patches. You can upgrade the package to fix various problems for the main program. If a patch package contains hot patches, they take effect before you restart the device restart. If a patch package contains cold patches, they take effect only after the device is restarted.

1.4.2 Patch Package Management

Upgrade of cold patches is to expand the component files in a package into the device. The upgrade takes effect only after the device is restarted.

Such process is similar to the upgrade of hot patches. But in the latter upgrade, only the files to be revised are replaced and then the new files take effect automatically.

Patch management is to use the database to record the patch information. The query, installation, and uninstallation correspond to the insertion, query, and deletion of the database.

Patch management includes three states: installed, activated, and confirmed:

- A patch in the installed state only indicates that this patch exists on the device, but it has not taken effect.
- Only a patch in the activated state is effective, but it will automatically roll back to the previous state after the device is restarted.
- A confirmed patch is also an effective patch, and the patch will not roll back even if the device is restarted.

1.4.3 Patch Auto-Sync

Patch auto-sync aims to coordinate multiple subsystems (boards and chassis) within a system. During software upgrade, the system automatically sends the patch installed on the current supervisor module to all the members meeting component installation rules. When a new member is added, the system automatically synchronizes the patch installed on the supervisor module to the new member. Thus, no resolved patch problem occurs even if synchronization is omitted.

Patch auto-sync rules:

- The subsystems comply with all the normal patch installation rules.
- The versions of all the subsystems must come from the same rack package before patch auto-sync is performed.
- The patch installed on the master supervisor module is used as the benchmark. If the same patch is installed on a newly connected member but the patch status is higher than that of the current supervisor module, the patch on the new member is not be processed. The status order of patches is: add, active, and running.

1.5 Configuration Task Summary

Software upgrade configuration includes the following tasks:

- Configuring System Version Upgrade

  a Configuring Subsystem Upgrade/Degrade

- (Optional) Configuring Version Auto-syncConfiguring Patch Version Upgrade

a Installing a Patch Package

b (Optional) Activating a Patch Package by One Click

c Download and activate the patch package permanently by one click.

patch autoload *url*

d Uninstalling a Patch Package

1.6 Configuring System Version Upgrade

1.6.1 Configuration Tasks

System version upgrade configuration includes the following tasks:

(1) Configuring Subsystem Upgrade/Degrade

(2) (Optional) Configuring Version Auto-sync

1.6.2 Configuring Subsystem Upgrade/Degrade

Warning

Boot upgrade is risky. If the device is powered off during upgrade, the device may be damaged and thus does not work. If the upgrade fails, do not power off the device, but attempt to upgrade the boot again.

1. Overview

Subsystem installation packages include the following types:

- Main package for device upgrade: After the upgrade using the main package is completed, all system software in the device is updated, and the overall software is enhanced.
- Rack package for upgrade: After the upgrade using the rack package is completed, all system software in the rack-type device is updated, and the overall software is enhanced.

Specification

Generally, a main package is released to upgrade box-type devices.

Generally, a rack package is released to upgrade rack-type devices.

Generally, a BIOS package is released to upgrade the BIOS subsystem on the device.

Generally, a CPLD package is released to upgrade the CPLD on the device.

Generally, an MCU package is released to upgrade the MCU on the device.

2. Restrictions and Guidelines

- To upgrade a main package of the device, download the installation package to the local device and run the **upgrade** command.
- After running the **upgrade** command to complete the upgrade of the installation package, you need to run the **reboot** command to restart the device to complete the update process. Otherwise, the installation package before the upgrade will be still running on the system.
- To upgrade a subsystem by using a rack package, store the rack package in the USB flash drive or SD card

of the device because the rack package is large, and then run the **upgrade** command. After completing the upgrade by using the installation package, run the **reboot** command to restart the device to complete the installation package update. Otherwise, the installation package running on the system is still the one used before the execution of the **upgrade** command.

- After completing the upgrade by using the installation package, run the **reboot** command to restart the device to complete the installation package update. Otherwise, the installation package running on the system is still the one used before the execution of the **upgrade** command.
- To upgrade the device by using a BIOS installation package, download the BIOS installation package to the local device and then run the **upgrade** command. After completing the upgrade by using the BIOS installation package, run the **reboot** command to restart the device to complete the installation package update. Otherwise, the BIOS installation package running on the system is still the one used before the execution of the **upgrade** command.
- To upgrade the device by using a CPLD installation package of the device, download the installation package to the local device and then run the **upgrade hot** command. After completing the upgrade by using the CPLD installation package, you do not need to run the **reboot** command and the installation package is updated.
- To upgrade the device by using an MCU installation package, download the MCU installation package to the local device and then run the **upgrade mcu** command. After completing the upgrade by using the MCU installation package, you do not need to run the **reboot** command and the installation package is updated.
- To upgrade the device by using an optical module installation package, download the optical module installation package to the local device and then run the **upgrade fiber-firmware** command. After completing the upgrade by using the optical module installation package, you do not need to run the **reload** command to restart the device.

3. Procedure

- Main Program and Rack Packages

(1) Enter the privileged EXEC mode.

privilege

(2) Upgrade the installation package corresponding to the device.

upgrade { *url* | **download** { **oob_ftp**://*path* [via mgmt interface-number] | **oob_tftp**://*path* [**via mgmt** i*nterface-number*] | ftp://path [**vrf** *vrf-name*] | **tftp**://*path* [**vrf** *vrf-name*] } } [slot slot-*id* | **slot all**] [**force**]

(3) Upgrade all Uboot and Rboot files on the device.

upgrade boot { *url* | **download** { **ftp:**//*path* | **oob_ftp:**//*path* [**via mgmt** *interface-number*] | **oob_tftp:**//*path* [**via mgmt** *interface-number*] | **tftp:**//*path* } } [**slot** s*lot-id* | **slot all**] [**vrf** *vrf-name*] [**force**]

(4) Upgrade the device by using a BIOS installation package.

upgrade { *url* | **download** { **oob_ftp:**//*path* [**via mgmt** *interface-number*] | **oob_tftp:**//*path* [**via mgmt** i*nterface-number*] | ftp://path [**vrf** *vrf-name*] | **tftp:**//*path* [**vrf** *vrf-name*] } } [**slot** s*lot-id* | **slot all**] [**force**] [**ignore_cpu_mem**]

(5) Upgrade the device by using a CPLD installation package.

upgrade hot { *url* | **download** { **oob_ftp:**//*path* [**via mgmt** *interface-number*] | **oob_tftp:**//*path* [**via mgmt** i*nterface-number*] | ftp://path [**vrf** *vrf-name*] | **tftp:**//*path* [**vrf** *vrf-name*] } } [**slot** s*lot-id* | **slot all**] [**force**] [**ignore_cpu_mem**]

(6) Upgrade the device by using an MCU installation package.

upgrade mcu { *url* | **download** { **oob_ftp:**//*path* [**via mgmt** interface-number] | **oob_tftp:**//*path* [**via mgmt** *interface-number*] | **ftp:**//*path* [**vrf** *vrf-name*] | **tftp:**//*path* [**vrf** *vrf-name*] } } [**slot** *slot-id* | **slot all**]

Note

Only S6510-48VS8CQ and S6250-48XS8CQ support this command.

(7) Restart the device to complete the upgrade process.

reboot

- Optical Module Installation Package

(1) Enter the privileged EXEC mode.
enable

(2) Upgrade the device using the optical module installation package.
Enter the interface configuration mode.

upgrade fiber-firmware { *url* } [**password**: *key*]

(3) The configuration takes effect without restart.

1.6.3 Configuring Version Auto-sync

1. Overview

This section describes how to configure a version auto-sync policy for a distributed system.

2. Restrictions and Guidelines

- To upgrade a new member in the system, check the version auto-sync upgrade configuration, such as preparation of the installation package.
- If a device member is not checked because the system is not configured with a version auto-sync policy, you can upgrade a specific member manually.
- After the path of an upgrade package used for version auto-sync is configured, the upgrade function module will determine the location of the installation package according to this path when new members (boards, chassis, or new VSU members) join the system. Each time the system is upgraded, the upgrade function module records the path of the installation package for this upgrade and uses it for version auto-sync. You can also use the **upgrade auto-sync package** command to set an upgrade package path.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure a version auto-sync policy.

upgrade automatic-sync policy [**none** | **compatible** | **coordinate**]

Specification

All the parameters apply to only the rack-type devices and VSUs.

(3) Configure the range of version auto-sync.

upgrade automatic-sync range [**vsu**]

(4) Configure the path of the upgrade package for version auto-sync.

upgrade automatic-sync package *url*

Specification

All the parameters apply to only the rack-type devices and VSUs.

1.7 Configuring Patch Version Upgrade

1.7.1 Overview

Besides the main program version, RGOS provides a patch mechanism to quickly fix the problems occurring on the main program version. A hot patch package can fix software bugs before the device is restarted; but an installed cold patch package can take effect only after the device is restarted. This patch package fixes a specific software version only. After the main program version is upgraded, the corresponding patch will become invalid automatically. RGOS provides a patching mechanism in addition to the main program version to quickly fix bugs in the main program version. The patch package can only fix the bugs in specific software versions. After the main program version is upgraded, the corresponding patch will expire automatically.

1.7.2 Configuration Restrictions and Guidance

- After the hot patch package is installed, the patch will take effect directly without restarting the device. If a cold patch package is installed, the device must be restarted for the patch to take effect.

1.7.3 Configuration Tasks

Patch version configuration includes the following tasks:

(1) Installing a Patch Package

(2) Activating a Patch Package by One Click

(3) Download and activate the patch package permanently by one click.

patch autoload *url*

(4) Uninstalling a Patch Package

1.7.4 Installing a Patch Package

1. Overview

This section describes how to install a patch package to fix software bugs. A patch package is used to upgrade a specific version of software only.

2. Restrictions and Guidelines

- Before installing a patch package, download it to a local directory of the device.

Caution

The downloaded patch package cannot be renamed. The name of the patch package is used as the operation keyword for the subsequent patch activation, deactivation, and uninstallation.

- The installed patch must be activated before it is used. You must activate the patch first temporarily and then permanently. If a patch is activated temporarily, the patch will roll back to the previous state after the device is restarted. You can activate a patch temporarily only when verifying the patch function temporarily.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Download a patch package.

install add *url* [**slot** *slot-id* | **slot all**]

(3) Install the patch package. Select either configuration.

- Run the general patch command to install the patch package. Run the following commands in sequence.

a Activate the installed patch temporarily.

install activate *package_name* [**slot** *slot-id* | **slot all**]

b Activate the patch permanently.

install commit [**slot** *slot-id* | **slot all**]

- Run the **patch** command to install the patch package. Run the following commands in sequence.

Note

The **patch** command is used to install patches for the dedicated version of DCN.

a Activate the installed patch temporarily.

patch active [**slot** *slot-id* | **slot all**]

b Activate the patch permanently.

patch running [**slot** *slot-id* | **slot all**]

(4) Activate the patch permanently by one click.

patch auto-running [**slot** *slot-id* | **slot all**]

(5) (Optional) Restart the device for the patch to take effect.

reboot

Run this command only after a cold patch is installed.

1.7.5 Activating a Patch Package by One Click

1. Overview

This section describes how to install a patch package to fix software bugs. A patch package is used to upgrade a specific version of software only.

2. Restrictions and Guidelines

- To fix software bugs without upgrading the main program and perform temporary verification, activate the patch package temporarily by one click.
- To fix software bugs without upgrading the main program, activate the patch package permanently by one click.

⚠ Caution

The downloaded patch package cannot be renamed. The name of the patch package name is used as the operation keyword for the subsequent patch activation, deactivation, and uninstallation.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Activate the patch package by one click.

- Activate the patch package temporarily by one click.

  upgrade { *url* | **download** { **oob_ftp:** *//path* [**via mgmt** *interface-number*] | **oob_tftp:** *//path* [**via mgmt** *interface-number*] | **ftp:** *//path* [**vrf** *vrf-name*] | **tftp**: //path [**vrf** *vrf-name*] } } **patch-active** [**slot** *slot-id* | **slot all**] [**force**]

- Activate the patch package permanently by one click.

  upgrade { *url* | **download** { **oob_ftp:** *//path* [**via mgmt** *interface-number*] | **oob_tftp:** *//path* [**via mgmt** *interface-number*] | **ftp:** *//path* [**vrf** *vrf-name*] | **tftp**: //path [**vrf** *vrf-name*] } } **patch-running** [**slot** *slot-id* | **slot all**] [**force**]

- Download and activate the patch package permanently by one click.

  patch autoload *url*

1.7.6 Uninstalling a Patch Package

1. Overview

This section describes how to deactivate the activated patches and remove the existing patches from the device, that is, the database information.

2. Restrictions and Guidelines

For users, only one patch is effective on the device. Therefore, only one patch to be activated is allowed on the device. To install another patch package of a function component, you need to uninstall the previous patch.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Uninstall a patch.

- Run the general patch command to uninstall a patch package. Run the following commands in sequence.

  a Deactivate an activated patch.

install deactivate *package_name* [**slot** *slot-id* | **slot all**]

b Uninstall a patch.

install remove *package_name* [**slot** *slot-id* | **slot all**]

- Run the **patch** command to uninstall a patch package. Run the following commands in sequence.

a Deactivate an activated patch.

patch deactivate *package_name* [**slot** *slot-id* | **slot all**]

b Uninstall a patch.

patch delete [**slot** *slot-id* | **slot all**]

Note

The **patch** command is used to install patches for the dedicated version of DCN.

1.8 Monitoring

Run the **display**commands to check the running status of a configured function to verify the configuration effect.

Run the **check** commands to check information.

Run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
display upgrade status	Displays the upgrade status of the device and the installation status of a patch.
display upgrade history	Displays the upgrade history.
check version	Checks the version matching.

Table 1-2 Patch Version Upgrade Monitoring

Command	Purpose
cleanup install storage [**slot** *slot-id* | **slot all**]	Clears all the patch packages not running currently and corresponding database information.
display install	Displays information about the hot patch package already installed on the device in a multi-patch solution.
display patch	Displays information about the hot patch package already installed on the device in a single-patch solution.

Command	Purpose
display install automatic-sync	Displays whether the device enables patch auto-sync.
display patch auto-sync	Displays information about all the patch packages that need to be synchronized automatically on the device.

1.9 Configuration Examples

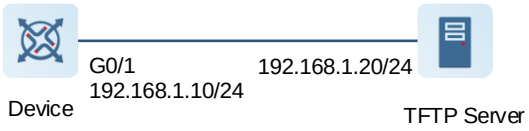
1.9.1 Configuring Subsystem Upgrade

1. Requirements

Upgrade a subsystem installation package to update all software in the device so that the overall software is enhanced and the known software bugs are fixed. Use Figure 1-1 as an example. The **upgrade download oob_tftp://***path* command is used to upgrade the subsystems of the device by using the installation package file stored on the TFTP server.

2. Topology

Figure 1-1 Configuring Subsystem Upgrade



3. Notes

- Configure an IP address for the Ethernet interface interface on the device so that the L3 route is reachable between the device and the TFTP server.
- Run the **upgrade download oob_tftp:**//*path* command to upgrade the installation package file stored on the TFTP server.

4. Procedure

(1) Configure an IP address for the Ethernet interface and default route on the device.

```
Hostname> privilege
Hostname# config-system terminal
Hostname(cfg)# interface gigabitethernet 0/1
Hostname(cfg-if-gigabitethernet 0/1)# ip address 192.168.1.10 255.255.255.0
Hostname(cfg-if-gigabitethernet 0/1)# gateway 192.168.1.1
Hostname(cfg-if-gigabitethernet 0/1)# ip route-config 0.0.0.0 0.0.0.0 gigabitethernet 0/1 192.168.1.20
Hostname(cfg)#end
```

(2) Run the **upgrade download oob_tftp:**//*path* command to upgrade the device.

```
Hostname# upgrade download oob_tftp://192.168.1.20/main.bin force via mgmt 0
You use the force command, Are you sure? Continue [Y/N]y
*Nov 23 13:21:39: %UPGRADE-6-INFO: Copy to /tmp/vsd/0/upgrade_rep/
```

```
*Nov 23 13:21:39: %UPGRADE-6-INFO: Please wait for a moment......
Press Ctrl+C to quit
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
< The terminal is locked by upgrade module >
Upgrade start
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!10%
< you can press Ctrl+C to unlock terminal >
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!20%
< you can press Ctrl+C to unlock terminal >
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!30%
< you can press Ctrl+C to unlock terminal >
!!!!!!!!!!!!!!!!!!!!!!!!!!!100%
Upgrade success
< The terminal is unlocked by upgrade module >
[Slot 0]
          Device type        : s6120
          Status              : success
```

(3) After upgrading the subsystems, run the **reboot** command to restart the device so that the configuration takes effect.

```
Hostname# reboot
Reboot system?(Y/N)y
*Dec 25 06:48:02: %SYSMON-5-RELOAD: Reset system from [CONSOLE] with reason [Management
port/Reset system by reboot command]
Hostname#*Dec 25 06:48:02: %DP-3-RESET_DEV: Reset device 1 due to
SYSMON/3766/CONSOLE/Management port/Reset system by reboot command.
*Dec 25 06:48:04: %DP-3-MACHINE_RESTART: MACHINE_RESTART.
```

5. Verification

(1) Run the **ping** command to check whether the L3 route between the device and the TFTP server is reachable.

```
Hostname# ping oob 192.168.1.10
Sending 5, 100-byte ICMP Echoes to 192.168.1.20, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/3 ms.
```

(2) Run the **display upgrade status** command to displaythe upgrade result.

```
Hostname# display upgrade status
upgrade global status: INIT
[Slot 0]
          Device type        : s6120
          Status              : success
```

(3) Check the running version on the device. If the version information changes, the upgrade is successful.

```
Hostname# display version
System description         : Ruijie Full 25G Routing Switch(S6510-48VS8CQ) By Ruijie Networks
```

```
System start time           : 2023-10-07 19:13:29
System uptime               : 0:00:04:23
System hardware version : 2.00
System software version : S6500_RGOS 12.5(1)B0506
System patch number         : NA
System software number    : M02523010052023
System serial number        : G1MB11B000057
System boot version         : 1.3.36(Master) 1.3.29(Slave)
System rboot version        : 1.2.8
System core version         : 5.4.231-Cavium-Octeon+
System cpu partition        : 3
Module information:
  Slot 0 : RG-S6510-48VS8CQ
    System uptime           : 0:00:04:23
    Hardware version        : 2.00
    Boot version            : 1.3.36(Master) 1.3.29(Slave)
    Rboot version           : 1.2.8
    Software version        : S6500_RGOS 12.5(1)B0506
    Software number         : M02523010052023
    Serial number           : G1MB11B000057
```

6. Common Errors

If an error occurs during the upgrade, the system displays the cause of the upgrade failure. You can also run the **display upgrade status** command to displaythe result of the last upgrade.

The following describes several types of common error messages:

- Invalid installation package: The possible cause is that the package is damaged or is not an installation package. You are advised to obtain the correct installation package and perform the upgrade again.
- Installation package not supported by the device: The possible cause is that you use the installation package of other devices. You are advised to obtain the correct installation package, verify the package information, and perform the upgrade again.
- Insufficient device space: Generally, this error occurs on a rack-type device. You are advised to check whether the device is equipped with a USB flash drive or SD card. Rack-type devices often have USB flash drives.

1.9.2 Configuring Auto-sync Upgrade

1. Requirements

On devices using a distributed structure such as rack-type devices and VSUs, auto-sync upgrade can upgrade the software version of matched members according to a configured auto-sync upgrade policy when software is upgraded or new members are added. This function ensures the coordination of the entire system. In the example of the VSU system using rack-type devices, we auto-sync an installation package.

2. Procedure

Configure an auto-sync upgrade policy based on the software version of the controlling device.

```
Hostname> privilege
```

```
Hostname# upgrade automatic-sync policy coordinate
```

Set the auto-sync upgrade range to the VSU system.

```
Hostname# upgrade auto-sync range vsu
```

3. Verification

Displaythe current auto-sync upgrade policy.

```
Hostname# display upgrade auto-sync
automatic-sync range          : vsu
automatic-sync policy         : coordinate
automatic-sync package        : flash:install_file/S6120_install.bin
```

Check version matching of device nodes.

```
Hostname# check version
  Dev Slot State
  --- ---- ------------
  1   4    Compatible
  1   FE6  Compatible
  1   M1   Compatible
```

4. Common Errors

- The url parameter in the command used for configuring auto-sync upgrade does not point to a valid upgrade package.

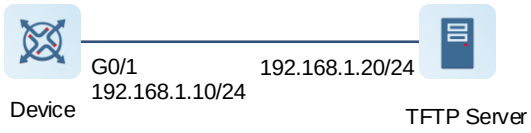
1.9.3 Configuring Patch-based Upgrade

1. Requirements

Install a patch package to fix software bugs without upgrading the main program version.

2. Topology

Figure 1-2 Configuring Patch-based Upgrade



3. Notes

(1) Configure the L3 route reachable to the device.

(2) Copy the patch package to the device.

(3) Run the patch installation command.

(4) Activate a hot patch.

a Activate a hot patch temporarily.

b Activate a hot patch permanently.

(5) Configure patch auto-sync.

4. Procedure

(1) Configure an IP address for the Ethernet interface and default route on the device.

```
Hostname> privilege
Hostname# config-system terminal
Hostname(cfg)#interface gigabitethernet 0/1
Hostname(cfg-if-gigabitethernet 0/1)# ip address 192.168.1.10 255.255.255.0
Hostname(cfg-if-gigabitethernet 0/1)# exit
Hostname(cfg)# ip route-config 0.0.0.0 0.0.0.0 gigabitethernet 0/1 192.168.1.20
Hostname(cfg)# end
```

(2) Install a patch package.

```
Hostname# install add tftp//192.1.1.20/smu_rf_hot1002_0118.bin
< The terminal is locked by patch moduleed by patch module >
Press Ctrl+C to quit
!
Operating, please wait for a moment.....
Operate finish!
Operate result information:
-------------------------------------
Slot              Result            Comment
1/0               Success           None
2/0               Success           None

< The terminal is unlocked by patch module >
```

(3) Activate a patch.

- Run the **install** command to activate the patch.

  a Run the **install active** command to activate the hot patch temporarily.

```
Hostname# install activate smu_rf_hot1002_0118.bin
< The terminal is locked by patch module >
Operating, please wait for a moment.....
Operate finish!
Operate result information:
-------------------------------------
1/0               Success           None
2/0               Success           None

< The terminal is unlocked by patch module >
```

  b Run the **install commit** command to activate the hot patch permanently.

```
Hostname# install commit
< The terminal is locked by patch module >
Operating, please wait for a moment...
Operate finish!
Operate result information:
-------------------------------------
```

```
Slot              Result            Comment
1/0               Success           None
2/0               Success           None

< The terminal is unlocked by patch module >
Hostname#
```

- Run the **patch** command to activate the patch.

  a Run the **patch active** command to activate the hot patch temporarily.

```
Hostname# patch active
< The terminal is locked by patch module >
Operating, please wait for a moment.....
Operate finish!
Operate result information:
-------------------------------------
1/0               Success           None
2/0               Success           None

< The terminal is unlocked by patch module >
```

  b Run the **patch running** command to activate the hot patch permanently.

```
Hostname# patch running
< The terminal is locked by patch module >
Operating, please wait for a moment...
Operate finish!
Operate result information:
-------------------------------------
Slot              Result            Comment
1/0               Success           None
2/0               Success           None

< The terminal is unlocked by patch module >
```

(4) Configure patch auto-sync.

```
Hostname# install sync enable
```

5. Verification

Run the **display install** command to displayinformation about the hot patch installed on the current device.

```
Hostname# display install
Install information:
 [Slot 1/0]
      Name                      Package        State        Flag        Effective time
      smu_rf_hot1002_0118.bin                  Running      Hot         2019-09-17 19:00:02

 [Slot 2/0]
      Name                      Package        State        Flag        Effective time
      smu_rf_hot1002_0118.bin                  Running      Hot         2019-09-17 19:00:02
```

Run the **display patch** command to displayinformation about the hot patch installed on the current device.

```
Hostname# display patch
 [Slot 1/0]
      Patch name              : smu_rf_hot1002_0118
      File name               : smu_rf_hot1002_0118.bin
      Patch state             : Install
      Patch flag              : Hot
      Effective time          :
      Last patch              :

 [Slot 2/0]
      Patch name              : smu_rf_hot1002_0118
      File name               : smu_rf_hot1002_0118.bin
      Patch state             : Install
      Patch flag              : Hot
      Effective time          :
      Last patch              :
```

6. Common Errors

- If you run the **install commit** command when a patch is not activated, an error is prompted. The **install commit** command takes effect only when the patch is in the active state.
- If you run the **patch running** command when a patch is not activated, an error is prompted. The **patch running** command takes effect only when the patch is in the active state.

1.9.4 Uninstalling a Patch Package

1. Requirements

One function component in one device can run one patch only. To install a new patch for this function component, you need to uninstall the previous patch before installation.

2. Notes

(1) Run the patch deactivation command.

(2) Run the patch uninstall command.

3. Procedure

(1) Uninstall a patch.

- Run the **install** command to uninstall a patch.

  a Run the **install deactivate** command to deactivate the patch.

```
Hostname# install deactivate smu_rf_hot1002_0118.bin
< The terminal is locked by patch module >
Operating, please wait for a moment.....
Operate finish!
Operate result information:
-------------------------------------
1/0                Success            None
```

```
2/0                Success          None

< The terminal is unlocked by patch module >
```

b Run the **install remove** command to uninstall the patch.

```
Hostname#install remove smu_rf_hot1002_0118.bin
< The terminal is locked by patch module >
Operating, please wait for a moment...
Operate finish!
Operate result information:
---------------------------------------
Slot               Result           Comment
1/0                Success          None
2/0                Success          None

< The terminal is unlocked by patch module >
```

- Run the **patch** command to uninstall a patch.

a Run the **patch deactive** command to deactivate the patch.

```
Hostname# patch deactive
< The terminal is locked by patch module >
Operating, please wait for a moment.....
Operate finish!
Operate result information:
---------------------------------------
1/0                Success          None
2/0                Success          None

< The terminal is unlocked by patch module >
```

b Run the **patch delete** command to uninstall the patch.

```
Hostname# patch delete
< The terminal is locked by patch module >
Operating, please wait for a moment...
Operate finish!
Operate result information:
---------------------------------------
Slot               Result           Comment
1/0                Success          None
2/0                Success          None

< The terminal is unlocked by patch module >
```

4. Verification

Run the **display install** command to displaythe patch uninstallation result.

```
Hostname#display install
[Slot 1/0]
```

```
    [No Install information]

[Slot 2/0]
    [No Install information]
```

Run the **display patch** command to displaythe patch uninstallation result.

```
Hostname# display patch
[Slot 1/0]
    [No Install information]

[Slot 2/0]
    [No Install information]
```

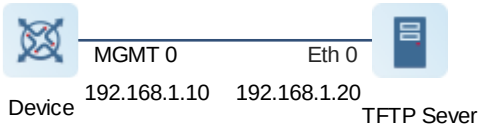
1.9.5 Installing a Patch Package by One Click

1. Requirements

Install a patch package to fix software bugs without upgrading the main program version.

2. Topology

Figure 1-3 Installing a Patch Packet with One Click



3. Notes

- Configure an L3 route reachable to the device.
- Run the patch installation command to load a patch and activate it permanently.
- Configure patch auto-sync.

4. Procedure

(1) Configure an IP address for the network management interface on the device.

```
Hostname> privilege
Hostname# config-system terminal
Hostname(cfg)#interface mgmt 0
Hostname(cfg-if-Mgmt 0)# ip address 192.168.1.10 255.255.255.0
Hostname(cfg-if-Mgmt 0)# gateway 192.168.1.1
```

(2) Install a patch package.

```
Hostname# patch autoload tftp://192.168.1.20/SP1.bin
Press Ctrl+C to quit
!
< The terminal is lock >
Operating,
Operate finish!
```

```
Operate result information:
-------------------------------------
Slot              Result            Comment
M1                 Success           None
3                  Success           None

< The terminal is unlock >
```

(3) Configure patch auto-sync.

```
Hostname# install sync enable
```

5. Verification

Run the **display patch** command to displayinformation about the hot patch installed on the current device.

```
Hostname# display patch
[Slot M1]
Patch package SP1 installed in the system, version:5.0.0.0
    ----------------------------------------
    Patch          : SP1.bin
    Status         : running
    Version        : 5.0.0.0
    Size           : 1770
    Install time: 1970-05-29 17:15:08
    Description : test SP1

[Slot 3]
Patch package SP1 installed in the system, version:5.0.0.0
    ----------------------------------------
    Patch          : SP1.bin
    Status         : running
    Version        : 5.0.0.0
    Size           : 1770
    Install time: 1970-05-29 17:15:07
    Description : test SP1
```

Contents

1 Configuring Uboot 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Accessing the Uboot Menu 1
1.1.3 Downloading XModem 2
1.1.4 Running the Main Program 2
1.1.5 Running the Rboot Program 2
1.1.6 Querying and Setting Other Functions 2
2 Configuring the SIMPLE CLI 4
2.1 Introduction 4
2.1.1 Overview 4
2.1.2 Accessing the SIMPLE CLI Operation Interface 4
2.1.3 Downloading Xmdown 5
2.1.4 Running the Rboot Program 5
2.1.5 Running the Main Program 5
2.1.6 Setting the Baud Rate 6
2.1.7 Reloading the System 6
2.1.8 Displaying the Version 6
2.1.9 Skipping to the Menu Interface 6

1 Configuring Uboot

1.1 Introduction

1.1.1 Overview

The Uboot menu includes all functions supported by the universal boot loader (Uboot), including booting the main program, booting Rboot, and updating Uboot or Rboot.

1.1.2 Accessing the Uboot Menu

1. Connecting to the Client

Connect the COM port of a PC to the serial port of the device through a serial port cable, start the HyperTerminal, and configure the following settings:

- Protocol type for connection: Serial
- Port for connection: COM
- Bits per second: 9600
- Data bit: 8
- Parity check: N/A
- Stop bit: 1
- Data flow control: N/A

2. Starting the Device

After the device is started, the following information is displayed:

```
Press Ctrl+B to enter Boot Menu, Press Ctrl+C to enter Rboot
```

Press **Ctrl+B**.

3. Accessing the Uboot Menu Interface

The interface is displayed as follows:

```
====== BootLoader Menu("Ctrl+Z" to upper level) ======
    TOP menu items.
************************************************
    0. XModem utilities.
    1. Run main.
    2. Run rboot.
    3. Scattered utilities.
************************************************
Press a key to run the command:
```

After the device displays the menu interface, select a menu item to select the corresponding function.

Press the first character of the menu item to perform the function of the menu item. If the menu item is a submenu, the submenu is accessed. The device displays the interface of the submenu.

Press **Ctrl+Z** to exit the current submenu and return to the previous-level menu.

1.1.3 Downloading XModem

On the menu interface, select **Upgrade bootloader**.

This is a submenu. After you access the submenu, the following information is displayed:

```
====== BootLoader Menu("Ctrl+Z" to upper level) ======
    XModem utilities.
***********************************************
    0. Upgrade bootloader.
    1. Upgrade rboot.
***********************************************
Press a key to run the command:
```

1. Upgrading bootloader

This function downloads a Uboot file to the flash memory to update the master Uboot of the system.

(1) On the menu interface, select **Upgrade bootloader**.

(2) Click **Send File** under **Transfer** in HyperTerminal and select a Uboot image file to be downloaded (for example, **UBOOT.BIN**).

(3) Click **Send**.

(4) Wait until the system displays a message indicating that the operation is complete.

2. Upgrading rboot

This function downloads an Rboot file to the flash memory to update Rboot of the system.

(1) On the menu interface, select **Upgrade rboot**.

(2) Click **Send File** under **Transfer** in HyperTerminal and select an Rboot image file to be downloaded (for example, **RBOOT.BIN**).

(3) Click **Send**.

(4) Wait until the system displays a message indicating that the operation is complete.

1.1.4 Running the Main Program

On the menu interface, select **Run main**. The program automatically skips to the main program from Uboot.

1.1.5 Running the Rboot Program

On the menu interface, select **Run rboot**. The program automatically skips to Rboot from Uboot.

1.1.6 Querying and Setting Other Functions

On the menu interface, select **Scattered utilities**.

This is a submenu. After you access the submenu, the following information is displayed:

```
====== BootLoader Menu("Ctrl+Z" to upper level) ======
    Scattered utilities.
***********************************************
    0. Display the bootloader version.
```

```
    1. reboot system.
    2. Set baudrate.
    3. Set default environment.
    4. Set debugging mode.
5. Run main without privilege cipher.
***********************************************
Press a key to run the command:
```

1. Display the bootloader Version

This function is used to display the version of the boot program on the current flash memory.

(1) On the menu interface, select **Display the bootloader version**.

(2) The version of Uboot is displayed.

Information similar to the following is displayed:

```
'master_bootloader' program information:
  Version:      1.3.13
'slave_bootloader' program information:
  Version:      1.3.13
```

2. Reloading the System

On the menu interface, select **Reboot system**. The system reboots automatically.

3. Setting the Baudrate

This is a submenu. In the submenu, select **Set baudrate**. The following information is displayed:

```
====== BootLoader Menu("Ctrl+Z" to upper level) ======
    Set baudrate.
***********************************************
    0. Change baudrate to 9600
    1. Change baudrate to 57600
    2. Change baudrate to 115200
***********************************************
Press a key to run the command:
```

- On the menu interface, select **Change baudrate to 9600**. The system changes the baud rate to **9600**.
- On the menu interface, select **Change baudrate to 57600**. The system changes the baud rate to **57600**.
- On the menu interface, select **Change baudrate to 115200**. The system changes the baud rate to **115200**.

4. Setting the Default Environmental Variable

This function is used to replace the environmental variable stored on the flash memory with the default environmental variable.

On the menu interface, select **Set default environment**. The following information is displayed:

```
Warning: The environment information maybe lose, Are you sure to continue? [yes/No]:
```

Enter **yes** or **no**. If you do not enter any character, **no** is used by default. When this menu is used to set the default environmental variable, the baud rate is not restored to the default value. The baud rate currently saved on the flash memory is still used.

5. Setting the Debugging Mode

This is a submenu. After you access the submenu, the following information is displayed:

```
====== BootLoader Menu("Ctrl+Z" to upper level) ======
    Set debugging mode.
************************************************
    0. Debug switch On.
    1. Debug switch Off.
************************************************
Press a key to run the command:
```

Select **Debug switch On** to enable the debugging mode or select **Debug switch Off** to disable the debugging mode.

The main program and Rboot output more boot logs in debugging mode than those in normal mode.

6. Starting the Main Program without Enabling Password

On the menu interface, select **Run main without privilege cipher**. Click **OK**. The main program is started. In the main program, you can enable the device without a password. This function is used when you forget the password. This function is effective for only the current operation. After reboot, the password is still required.

2 Configuring the SIMPLE CLI

2.1 Introduction

2.1.1 Overview

The SIMPLE command line interface (CLI) includes some functions supported by Uboot, which are implemented through command lines, including booting the main program, booting Rboot, and updating Uboot or Rboot.

2.1.2 Accessing the SIMPLE CLI Operation Interface

On the Uboot menu interface, press **Ctrl+Q** to access the SIMPLE CLI operation interface.

```
====== BootLoader Menu("Ctrl+Z" to upper level) ======
    TOP menu items.
************************************************
    0. XModem utilities.
    1. Run main.
    2. Run rboot.
    3. Scattered utilities.
************************************************
Press a key to run the command:
```

Enter **help**. The functions of commands are displayed as follows:

```
bootloader# help
Total commands:
```

```
help            Dump command list OR show a command's details
xmdown            Download programs through XModem.
runrboot        Run rboot program.
runmain          Run main program.
setbaud          Set BOOT/BOOTLOADER baudrate tools.
reboot           reboot tools.
version          Display current version information.
quit             Quit from CLI command line.
```

2.1.3 Downloading Xmdown

You can run this command to upgrade the main Uboot program and Rboot program on the flash memory.

Enter **help xmdown**. The usage details of the **xmdown** command are displayed.

```
bootloader# help xmdown
Syntax: xmdown (-boot | -rboot)
Usage Details:
    -boot:      Upgrade BootLoader.
    -rboot:   Upgrade Rboot.
Examples:
    xmdown -boot
    xmdown -rboot
```

1. xmdown -boot

This command is used to upgrade the main Uboot program on the flash memory.

(1) Enter **xmdown –boot** and press **Enter**.

(2) Click **Send File** under **Transfer** in HyperTerminal and select a Uboot image file to be downloaded (for example, **UBOOT.BIN**).

(3) Click **Send**.

(4) Wait until the system displays a message indicating that the operation is complete.

2. xmdown -rboot

This command is used to upgrade the Rboot program on the flash memory.

(1) Enter **xmdown –rboot** and press **Enter**.

(2) Click **Send File** under **Transfer** in HyperTerminal and select an Rboot image file to be downloaded (for example, **RBOOT.BIN**).

(3) Click **Send**.

(4) Wait until the system displays a message indicating that the operation is complete.

2.1.4 Running the Rboot Program

Enter the **runrboot** command. The program automatically skips to Rboot from Uboot.

2.1.5 Running the Main Program

Enter the **runmain** command. The program automatically skips to the main program from Uboot.

2.1.6 Setting the Baud Rate

After the **setbaud** command is run, the baud rate of the current device is changed and the changed baud rate is saved to the environmental variable partition of the flash memory.

Enter the **help setbaud** command. The usage details of the **setbaud** command are displayed:

```
Hostname#help setbaud
Syntax: setbaud (-h | -m | -l)
Usage Details:
      -h:              Set to 115200
      -m:              Set to 57600
      -l:              Set to 9600
Examples:
      setbaud -l
      setbaud -m
setbaud –h
```

- setbaud -l: The baud rate of the current device is changed into **9600** and the changed baud rate is saved to the environmental variable partition of the flash memory.
- setbaud -m: The baud rate of the current device is changed into **57600** and the changed baud rate is saved to the environmental variable partition of the flash memory.
- setbaud -h: The baud rate of the current device is changed into **115200** and the changed baud rate is saved to the environmental variable partition of the flash memory.

2.1.7 Reloading the System

Enter the **Reboot** command and press **Enter**. The system reboots automatically.

2.1.8 Displaying the Version

Enter the **version** command and press **Enter**. The system displays the versions of the master Uboot and slave Uboot. For example:

```
bootloader# version
'master_bootloader' program information:
   Version:        1.3.13
'slave_bootloader' program information:
   Version:        1.3.13
```

2.1.9 Skipping to the Menu Interface

Enter the **quit** command and press **Enter**. The system returns to the menu interface.

Contents

1 Configuring Rboot 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Accessing the Rboot Menu 1

1.1.3 Downloading Tftp utilities 2

1.1.4 Downloading X/Y/ZModem 4

1.1.5 Running the Main Program 6

1.1.6 Querying and Setting Other Functions 6

1 Configuring Rboot

1.1 Introduction

1.1.1 Overview

The Rboot menu includes all functions supported by Rboot, including booting the main program, updating the universal boot loader (Uboot), updating Rboot, and reinstalling the system.

Specification

The support for the configuration varies with specific products.

1.1.2 Accessing the Rboot Menu

1. Connecting to the Client

Connect the COM port of a PC to the serial port of the device through a serial port cable, start the client software, and configure the following settings:

- Protocol type -for connection: Serial
- Port for connection: COM
- Bits per second: 9600
- Data bit: 8
- Parity check: N/A
- Stop bit: 1
- Data flow control: N/A

2. Starting the Device

After the device is started, the following information is displayed:

```
Press Ctrl+B to enter Boot Menu, Press Ctrl+C to enter Rboot
```

Press **Ctrl+C**.

When Rboot is started, the current version is displayed.

```
Rboot                                   Version: 1.1.22
-------------------------------------------------
```

3. Accessing the Rboot Menu Interface

The interface displays the following:

```
====== Rboot Menu(Ctrl+Z to upper level) ======
      TOP menu items.
************************************************
      0. Tftp utilities.
      1. X/Y/ZModem utilities.
```

```
    2. Run main.
    3. SetMac utilities.
    4. Scattered utilities.
***********************************************
Press a key to run the command:
```

After the device displays the menu interface, select a menu item to select the corresponding function.

Press the first character of the menu item to perform the function of the menu item. If the menu item is a submenu, the submenu is accessed. The device displays the interface of the submenu.

Press **Ctrl+Z** to exit the current submenu and return to the previous-level menu.

1.1.3 Downloading TFTP utilities

This is a submenu. After you access the submenu, the following information is displayed:

```
====== Rboot Menu(Ctrl+Z to upper level) ======
    Tftp utilities.
***********************************************
    0. Upgrade uboot/bios program.
    1. Upgrade rboot program.
    2. Upgrade main program.
    3. Upgrade the entire device by distribute package.
    4. Burn the total FlashROM by this downloaded file.
***********************************************
Press a key to run the command:
```

1. Upgrading Uboot/BIOS

This function downloads a Uboot or BIOS file to the flash memory to update the master or slave Uboot of the system or the master or slave BIOS of the system, depending on the specific requirement.

(1) Start the Trivial File Transfer Protocol (TFTP) server on the PC or supervisor module.

(2) Be sure to put the Uboot or BIOS upgrade image file to be downloaded in the TFTP server directory of the PC or supervisor module.

(3) Connect the device to the PC through an Ethernet cable.

(4) On the menu interface, select **Upgrade uboot/bios program**.

(5) Enter **Local IP** (IP address of the switch), **Remote IP** (IP address of the PC), and **Filename** (name of the Uboot or BIOS file to be downloaded) based on prompts and press **Enter**.

(6) Wait until the system displays a message indicating that the operation is complete.

2. Upgrading Rboot

This function downloads an Rboot file to the flash memory to update Rboot of the system.

(1) Start the TFTP server on the PC or supervisor module.

(2) Be sure to put the Rboot upgrade image file to be downloaded in the TFTP server directory of the PC or supervisor module.

(3) Connect the device to the PC through an Ethernet cable.

(4) On the menu interface, select **Upgrade rboot program**.

(5) Enter **Local IP** (IP address of the switch), **Remote IP** (IP address of the PC), and **Filename** (name of the Rboot file to be downloaded) based on prompts and press **Enter**.

3. Upgrading the Main Program

The function is used to parse a main program package and upgrade the main program. If the main program package contains Uboot or Rboot, Uboot or Rboot is upgraded depending on the version.

(1) Start the TFTP server on the PC or supervisor module.

(2) Be sure to put the upgrade image file of the main program to be downloaded in the TFTP server directory of the PC or supervisor module.

(3) Connect the device to the PC through an Ethernet cable.

(4) On the menu interface, select **Upgrade main program**.

(5) Enter **Local IP** (IP address of the switch), **Remote IP** (IP address of the PC), and **Filename** (name of the main program file to be downloaded) based on prompts and press **Enter**.

(6) Wait until the system displays a message indicating that the operation is complete.

4. Upgrade the Software Package

This function is used to parse a software package, retrieve the main program from the software package, and upgrade the device by using the main program according to the upgrade procedure of the main program.

(1) Start the TFTP server on the PC or supervisor module.

(2) Be sure to put the upgrade image file of the software package to be downloaded in the TFTP server directory of the PC or supervisor module.

(3) Connect the device to the PC through an Ethernet cable.

(4) On the menu interface, select **Upgrade the entire device by distribute package**.

(5) Enter **Local IP** (IP address of the switch), **Remote IP** (IP address of the PC), and **Filename** (name of the program file to be downloaded) based on prompts and press **Enter**.

(6) Wait until the system displays a message indicating that the operation is complete.

5. Upgrading the Flash ROM File

This function is used to refresh the ROM files (including the slave Uboot, master Uboot, and Rboot) in the flash memory to burn such files into the flash memory again. After upgrade, the baud rate is changed to **9600** and **SetMac** needs to be set again.

(1) Start the TFTP server on the PC or supervisor module.

(2) Be sure to put the ROM upgrade image file to be downloaded in the TFTP server directory of the PC or supervisor module.

(3) Connect the device to the PC through an Ethernet cable.

(4) On the menu interface, select **Burn the total FlashROM by this downloaded file**.

(5) Enter **Local IP** (IP address of the switch), **Remote IP** (IP address of the PC), and **Filename** (name of the program file to be downloaded) based on prompts and press **Enter**.

(6) Wait until the system displays a message indicating that the operation is complete.

1.1.4 Downloading X/Y/ZModem

This is a submenu. After you access the submenu, the following information is displayed:

```
====== BootLoader Menu("Ctrl+Z" to upper level) ======
    XModem utilities.
************************************************
    X/Y/ZModem utilities.
************************************************
    0. XModem utilities.
    1. YModem utilities.
    2. ZModem utilities.
    3. Local utilities.
************************************************
Press a key to run the command:
```

1. Upgrading XModem

This function is used to download Uboot, BIOS, Rboot, main program, software package, or ROM through XMODEM and install it on the device.

(1) On the menu interface, select **XModem utilities**.

(2) The device displays the following submenus:

```
====== Rboot Menu (Ctrl+Z to upper level) ======
    XModem utilities.
************************************************
    0. Upgrade uboot/bios program.
    1. Upgrade rboot program.
    2. Upgrade main program.
    3. Upgrade the entire device by distribute package.
    4. Burn the total FlashROM by this downloaded file.
************************************************
Press a key to run the command:
```

(3) Select a required upgrade type.

(4) Select **Transfer** > **Send File** in HyperTerminal, select **XMODEM** for the protocol, and select a file to be downloaded by the upgrade type.

(5) Click **Send**.

(6) Wait until the system displays a message indicating that the operation is complete.

2. Upgrading YModem

This function is used to download Uboot, BIOS, Rboot, main program, software package, or ROM through YMODEM and install it on the device.

(1) On the menu interface, select **YModem utilities**.

(2) The device displays the following submenus:

```
====== Rboot Menu (Ctrl+Z to upper level) ======
    YModem utilities.
```

```
**********************************************
    0. Upgrade uboot/bios program.
    1. Upgrade rboot program.
    2. Upgrade main program.
    3. Upgrade the entire device by distribute package.
    4. Burn the total FlashROM by this downloaded file.
**********************************************
Press a key to run the command:
```

(3) Select a required upgrade type.

Note

This menu is similar to the TFTP menu. For details, see the TFTP menu.

(4) Select **Transfer** > **Send File** in HyperTerminal, select **YMODEM** for the protocol, and select a file to be downloaded by upgrade type.

(5) Click **Send**.

(6) Wait until the system displays a message indicating that the operation is complete.

3. Upgrading ZModem

This function is used to download Uboot, BIOS, Rboot, main program, software package, or ROM through ZMODEM and install them on the device.

(1) On the menu interface, select **ZModem utilities**.

(2) The device displays the following submenus:

```
====== Rboot Menu (Ctrl+Z to upper level) ======
    ZModem utilities.
**********************************************
    0. Upgrade uboot/bios program.
    1. Upgrade rboot program.
    2. Upgrade main program.
    3. Upgrade the entire device by distribute package.
    4. Burn the total FlashROM by this downloaded file.
**********************************************
Press a key to run the command:
```

(3) Select a required upgrade type.

(4) Select **Transfer** > **Send File** in HyperTerminal, select **ZMODEM** for the protocol, and select a file to be downloaded by upgrade type.

(5) Click **Send**.

(6) Wait until the system displays a message indicating that the operation is complete.

4. Upgrading Local Utilities

This function is used to download the Uboot, BIOS, Rboot, main program, software package, and ROM to a USB flash drive or flash memory partition and install them on the device.

(1) On the menu interface, select **Local utilities**.

(2) The device displays the following submenus:

```
====== Rboot Menu (Ctrl+Z to upper level) ======
    Local utilities.
************************************************
    0. Upgrade uboot/bios program.
    1. Upgrade rboot program.
    2. Upgrade main program.
    3. Upgrade the entire device by distribute package.
    4. Burn the total FlashROM by this downloaded file.
************************************************
Press a key to run the command:
```

(3) Select a required upgrade type.

(4) Based on the prompt, enter a file path.

a If a USB flash drive is used, the file path is the location of the **usb0:** file. If there are multiple USB flash drives or partitions, the file paths are usb1: file location, usb2: file location, and so on.

b If a flash memory partition is used, the file path is the location of the **flash:** file.

(5) Wait until the system displays a message indicating that the operation is complete.

1.1.5 Running the Main Program

On the menu interface, enter **3**. The program is booted to access the main program.

1.1.6 Querying and Setting Other Functions

This is a submenu. After you access the submenu, the following information is displayed:

```
====== Rboot Menu(Ctrl+Z to upper level) ======
    Scattered utilities.
************************************************
    0. Show Version.
    1. Reload System.
    2. Set Baudrate.
    3. Format Flash.
    4. Exit menu.
    5. Set debugging mode.
    6. Run main without privilege cipher.
    7. Set Download Gateway.
************************************************
Press a key to run the command:
```

Note

Options in the submenu increases with the update of the Rboot version but some Rboot items of some earlier products may not be listed.

1. Displaying the Boot Version

This function is used to display the versions of Uboot/BIOS and Rboot on all flash memories.

(1) On the menu interface, select **Show Version**.

(2) The system displays versions of Uboot/BIOS and Rboot.

Example:

```
The First    MasterBoot Version: 1.3.13
The First    SlaveBoot    Version: 1.3.13
The First    Rboot        Version: 1.0.5
The Second MasterBoot Version: 1.3.13
The Second SlaveBoot    Version: 1.3.13
The Second Rboot         Version: 1.0.5
```

Note

Some devices may not display all the six boot programs above and NA is displayed for the inexistent boot programs.

2. Reloading the System

On the menu interface, select **Reload System**. The system reboots automatically.

3. Setting the Baud Rate

This is a submenu. After you access the submenu, the following information is displayed:

```
====== BootLoader Menu("Ctrl+Z" to upper level) ======
    Set baudrate.
************************************************
    0. Change baudrate to 9600
    1. Change baudrate to 57600
    2. Change baudrate to 115200
************************************************
Press a key to run the command:
```

- On the menu interface, select **Change baudrate to 9600**. The system changes the baud rate to **9600**.
- On the menu interface, select **Change baudrate to 57600**. The system changes the baud rate to **57600**.
- On the menu interface, select **Change baudrate to 115200**. The system changes the baud rate to **115200**.

Note

The configuration applies to the boot program. In the main program, the baud rate of the boot program may change with the configuration of the main program.

4. Formatting the Main Storage Partition

Warning

If this function is used, the main program and all configurations of the device and all files in the flash memory partition are lost. Therefore, do not use this function unless necessary.

On the menu interface, select **Format Flash**. The main partition is formatted and a partition table is recreated.

5. Setting the Debugging Mode

This is a submenu. After you access the submenu, the following information is displayed:

```
====== Rboot Menu (Ctrl+Z to upper level) ======
    Set debugging mode.
************************************************
    0. Debug switch On.
    1. Debug switch Off.
************************************************
Press a key to run the command:
```

Select **Debug switch On** to enable the debugging mode or select **Debug switch Off** to disable the debugging mode.

The main program outputs more logs in debugging mode than those in normal mode.

6. Running the Main Program Without Enabling Password

On the menu interface, select **Run main without privilege cipher**. Click **OK**. The main program is started. In the main program, you can enable the device without a password. This function is used when you forget the password. This function is effective for only the current operation. After reboot, the password is still required.

Some devices do not support startup without a password under Rboot (but supports startup without a password under Uboot). In this case, the following information is displayed:

```
    Not supported in Rboot.
    Please reload, press Ctrl+B to enter Uboot,
    and select the same item in Uboot menu.
```

For such devices, access Uboot during startup and select the **Run main without privilege cipher** menu.

7. Downloading TFTP Across Gateways

This function is used to configure a gateway and a mask so that TFTP can be downloaded across gateways.

On the menu interface, select **Set Download Gateway**. The following information is displayed:

```
Plz enter the Local GatewayIP [<NULL>]:
```

Enter the IP address of a gateway in the format of X.Y.Z.W. The following message is displayed:

```
Plz enter the IP Netmask [<NULL>]:
```

Enter a subnet mask in the format of X.Y.Z.W. The following message is displayed:

```
    Set Local GatewayIP to <set IP address of the gateway>
Set IP Netmask to <set subnet mask>
```

Then return to the main menu and select items in **Tftp utilities**. The gateway and subnet mask are displayed during downloading and TFTP items are downloaded across gateways.

```
    TFTP: Detect Gateway <set IP address of the gateway> ...
    TFTP: Detect Netmask <set subnet mask> ...
```

The gateway configuration is not retained. After you leave Rboot, the configuration is lost.

To clear the gateway configuration in Rboot, enter a null address in **Set Download Gateway**.

```
Plz enter the Local GatewayIP [<NULL>]:
Plz enter the IP Netmask [<NULL>]:
    Cleanup Local GatewayIP config.
    Cleanup IP Netmask config.
```

Contents

1 Configuring Time Range 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuring Basic Features 1
1.2.1 Overview 1
1.2.2 Restrictions and Guidelines 1
1.2.3 Procedure 2
1.3 Monitoring 2
1.4 Configuration Examples 2
1.4.1 Configuring Basic Features of Time Range 2

1 Configuring Time Range

1.1 Introduction

1.1.1 Overview

Time range is a time control service. Users can create a time range and reference the time range in a service to control the time of the service. For example, to make an access control list (ACL) take effect in a time range within a week, configure a time range and associate the ACL with the time range. In this case, the ACL is effective only in the specified time range.

1.1.2 Principles

1. Basic Concepts

A time range is defined in the following modes:

- Absolute time range

  An absolute time range indicates the time range from the specified start time to the end time, for example, from 12:00:00 on January 1, 2000 to 12:00:00 on January 1, 2001.

- Period

  A period indicates a periodic time range, for example, from 8:00 Monday to 17:00 Friday.

2. Using an Absolute Time Range

When a function is enabled in a service that references a time range, the service determines whether the current time of the system is in an absolute time range. If yes, the function can be enabled or disabled in the current time.

3. Using a Period

When a function is enabled in an application based on a time range, the application determines whether the current time of the system is in a period. If yes, the function can be enabled or disabled in the current time.

1.2 Configuring Basic Features

1.2.1 Overview

This section describes how to create a time range and configure an absolute time range or a period.

1.2.2 Restrictions and Guidelines

- Before modifying an absolute time range or a period for a service, you are advised to disassociate the time range. After the absolute time range or period is modified, associate the time range again.
- In a time range, you can configure multiple effective periods.
- If an absolute time range and a period are configured in a time range, the effective time range is a period in the absolute time range. If multiple periods are configured, the effective time is the multiple periods configured in the absolute time range.

1.2.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a time range.

timerange*time-range-name*

(4) (Optional) Configure effective time of the time range. Configure at least one of the following tasks.

- Configure an absolute time range.

  absolute { **start** *hh:mm DD/MM/YY* | **end** *hh:mm DD/MM/YY* }*

  No absolute time range is configured. The maximum time range is used by default.

- Configure a period.

  periodic *day-of-the-week time* **to** [*day-of-the-week*] *time*

  No period is configured by default. The current time is considered to be within a period.

1.3 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-1 Time Range Monitoring

Command	Purpose
display timerange [*time-range-name*]	Displays a time range.

1.4 Configuration Examples

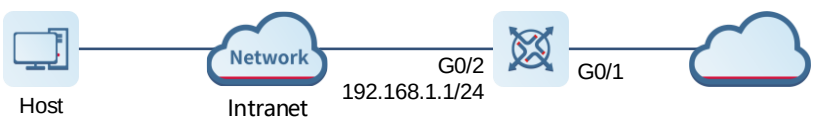
1.4.1 Configuring Basic Features of Time Range

1. Requirements

In the traffic control, an enterprise needs to control the traffic type and traffic in a specific period. For example, an enterprise prohibits employees from accessing the Internet during work and allows them to do so after work. In this case, an ACL can be configured to associate with a time range. During the working hours (9:00 to 12:00 and 14:00 to 18:00), PCs in the Intranet are prohibited from accessing the Internet. In other time ranges, the PCs are allowed to access the Internet. The PCs can access one another at any time.

2. Topology

Figure 1-1 Configuring Basic Functions of Time Range



3. Notes

- First configure system time for the device. The effective time of an ACL is based on the system time of the device.
- A time range cannot cross 0:00, that is, to create a time range from 22:00 to 7:00 the next day, create two time ranges.

4. Procedure

Configure system time in the following format: hh:mm:ss MMDDYY.

```
Hostname> privilege
Hostname # clock set 10:00:00 12 1 2012
```

Set the time zone of the device to GMT+8.

```
Hostname# config-system terminal
Hostname(cfg)# clock timezone beijing 8
```

Create a time range named **time1**, which is effective from 9:00 to 12:00 and from 14:00 to 18:00 every day.

```
Hostname(cfg)# timerange time1
Hostname(cfg-timerange)# periodic daily 9:00 to 12:00
Hostname(cfg-timerange)# periodic daily 14:00 to 18:00
Hostname(cfg-timerange)# exit
```

(1) Configure an ACL to be associated with the time range.

In the time range **time1**, configure the ACL to reject any traffic from the Intranet PCs in the network segment 192.168.1.0/24 to the Internet.

```
Hostname(cfg)# ip access-control-list extended 100
Hostname(cfg-extacl)# 10 deny ip 192.168.1.0 0.0.0.255 any time-range time1
```

Configure the ACL to allow the Intranet PCs in the network segment 192.168.1.0/24 to access the Internet in non-working hours.

```
Hostname(cfg-extacl)# 20 permit ip any any
Hostname(cfg-extacl)# exit
```

Apply the ACL to interface GigabitEthernet 0/1.

```
Hostname(cfg)# interface GigabitEthernet 0/2
Hostname(cfg-if-GigabitEthernet 0/1)# ip access-apply 100 in
```

5. Verification

Run the **display clock** command to displaythe time of the device.

```
Hostname# display clock
10:14:01 beijing Sat, Dec 1, 2012
```

Run the **display time-range** command to displaythe time range configurations.

```
Hostname# display timerange time1

time-range entry: time1 (active)
   periodic Daily 9:00 to 12:00
   periodic Daily 14:00 to 18:00
```

Run the **display access-lists** command to displaythe ACL configurations. When the time of the device is in the time range **time1**, the displayed status of the ACL is active; when the time of the device is not in the effective time of ACL, the displayed status of the ACL is inactive.

```
Hostname# display access-lists
ip access-control-list extended 100
10 deny ip 192.168.1.0 0.0.0.255 any timerange time1 (active)
20 permit ip any any
```

Run the **display ip access-apply** command to displaythe calling of the ACL on the interface.

```
Hostname# display ip access-apply
ip access-apply 100 in
Applied On interface GigabitEthernet 0/1.
```

Set the system time to working hours so that PCs in the Intranet cannot ping the Internet during working hours.

```
Hostname# display clock
14:52:26 beijing Tue, Dec 15, 2020
Hostname# ping 14.215.177.38
Sending 5, 100-byte ICMP Echoes to 14.215.177.38, timeout is 2 seconds:
   < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5).
```

6. Configuration Files

Device configuration file

```
hostname Device
!
ip access-control-list extended 100
  10 permit ip host 10.10.10.20 any
  20 permit ip any any
!
timerange time1
  periodic Daily 9:00 to 12:00
  periodic Daily 14:00 to 18:00
!
interface GigabitEthernet 0/2
  unset switch-port
  ip access-apply 100 in
  ip address 192.168.1.1 255.255.255.0
!
End
```

Contents

1 Configuring UFT Mode Management....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.2 Restrictions and Guidelines....2
1.3 Configuring a UFT Mode....2
1.3.1 Overview....2
1.3.2 Restrictions and Guidelines....2
1.3.3 Procedure....3
1.4 Monitoring....3
1.5 Configuration Examples....4
1.5.1 Switching the UFT Mode of a Switch in Standalone Mode....4

1 Configuring UFT Mode Management

Note

S6980-64QC does not support management in the forwarding table mode.

1.1 Introduction

1.1.1 Overview

The unified forwarding table (UFT) provides the function of dynamically allocating switch hardware forwarding entry resources so that users can select UFT modes to adapt to requirements of the scenarios.

1.1.2 Principles

1. UFT Mode

- **Default**

  Default UFT mode. In this mode, the hardware entries of a switch apply to most user scenarios, including user access to the gateway and large L2 gateway cores.

- **Alpm**

  Alpm mode. This mode is applicable to scenarios with many network routes.

- **Arp-fastswitch**

  Arp-fastswitch mode. This mode is applicable to fast ARP switching during an active/standby switchover.

- **Bridge**

  Layer 2 forwarding mode. This mode greatly reduces the Address Resolution Protocol (ARP)/Neighbor Discovery (ND) and Multiprotocol Label Switching (MPLS) capacity and allocates most capacity to MAC address table. It is applicable to scenarios of pure Layer 2 services.

- **Dual-stack-alpm**

  Dual-stack-alpm mode. This mode is applicable to dual-stack scenarios.

- **Fpem-alpm**

  Fpem-alpm mode. This mode is applicable to scenarios with many FP and route entries.

- **Gateway**

  Gateway mode. This mode is applicable to scenarios with mostly Layer 3 services.

- **Gateway-IPv4**

  IPv4 gateway mode. This mode does not support IPv6 addresses, but allocates resources occupied by IPv6 addresses exclusively to IPv4 addresses to increase the IPv4 routing capacity. This mode is applicable to scenarios with many IPv4 routes.

- **Ifa**

  ifa mode. This mode is applicable to scenarios requiring IFA scenarios.

- **Ipv6-64-alpm**

Ipv6-64-alpm mode. This mode is applicable to scenarios with many IPv4 and IPv6 routes with 0- to 64-bit subnet mask.

- **Route**

  Dual-stack routing mode. This mode supports IPv4 and IPv6 dual-stack routes. It is applicable to traditional Layer 3 core networks and the forwarding of many routes.

- **Route-v4max**
  IPv4 routing mode. This mode is applicable to scenarios with many IPv4 routes.
- **Route-v6max**
  IPv6 routing mode. This mode is applicable to scenarios with many IPv6 routes.
- **Sr-default**
  Segment routing (SR) default mode. This mode is applicable to scenarios in which SR-related functions (such as SR, SRP MPLS, SRv6, and Policy-based Tunnel Selection (PBTS) traffic diversion) are enabled in the default mode.
- **Vxlan**
  VXLAN mode. This mode is applicable to VXLAN scenarios.

1.2 Restrictions and Guidelines

- Different devices support different UFT modes.
- After different UFT modes are configured, you need to save the configurations and restart the device to validate the mode.
- When the sr-default mode is configured, the all parameter needs to be carried, for example, switch-mode sr-default all. When you switch from the sr-default mode to other modes, the all parameter also needs to be carried, for example, switch-mode default all.
- Before configuring the sr-default mode, disassociate services (such as security ACL, policy-based QoS, and policy-based mirroring), to which the ternary content-addressable memory (TCAM) has been applied. Otherwise, associated services cannot take effect or a system log indicating insufficient resources is displayed after mode switching due to TCAM resource restrictions.
- If chassis-type devices are started without cards, the configuration of the sr-default or sr-route mode takes effect only after line cards are inserted and the mode is reconfigured.

1.3 Configuring a UFT Mode

1.3.1 Overview

Different UFT modes are configured to meet requirements of different networking scenarios.

1.3.2 Restrictions and Guidelines

- After configuration, you need to save the configurations and restart the device to validate them. If the configuration is not saved and the device is restarted, then the original configuration will still be running on the device. If only the line card is restarted, the line card may not work normally.
- Change the UFT mode to a non-default mode and save the configurations. If you upgrade the device, the UFT function may cause line cards to automatically restart during the first restart. If SYSLOG prompts that the UFT mode is not supported, and the actual operation mode of the device is different from the

configuration mode, then you need to reconfigure or delete the configuration, and restart the device.

- If the system log shows that the UFT mode is not supported and the actual operating mode of the device differs from the configured mode, you need to reconfigure the UFT mode or delete the configurations and restart the device. If the CLI configuration mode before the restart is different from the mode in the config.text file, the device will restart automatically once after starting. The causes of inconsistent modes include but are not limited to the following:
  - The non-default UFT mode is saved before the device is upgraded.
  - The config.text file is manually deleted or modified.
  - The UFT mode modified in CLI is not saved.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Switch the UFT mode. The configuration tasks below are mutually exclusive. Configure one of them.

- In standalone mode, switch the UFT mode.

  switch-mode *mode-type* [**overlay**] **slot** *slot-id*

  The UFT mode is the default mode by default.

- In virtual switching unit (VSU) mode, switch the UFT mode.

  switch-mode *mode-type* [**overlay**] **switch** *switch-id* **slot** *slot-id*

  The UFT mode is the default mode by default.

- (Optional) Pre-configure the UFT mode globally.

  switch-mode *mode-type* [**overlay**] **global**

  The UFT mode is the default mode by default.

- (Optional) Configure the UFT mode globally.

  switch-mode *mode-type* [**overlay**] **all**

  The default UFT mode is the default mode.

1.4 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-1 UFT Monitoring

Command	Purpose
display switch-mode status	Displays information about the UFT mode of a switch.
display switch-mode all	Displays software and hardware resources allocated by the device in the UFT mode.

Note

- The above monitoring command is compatible with chassis-type devices and box-type devices, as well as the standalone mode and VSU mode.
- For the stand-alone mode, the keyword **device** is invisible. For a chassis-type device, you can set **slot** to specify a line card.

1.5 Configuration Examples

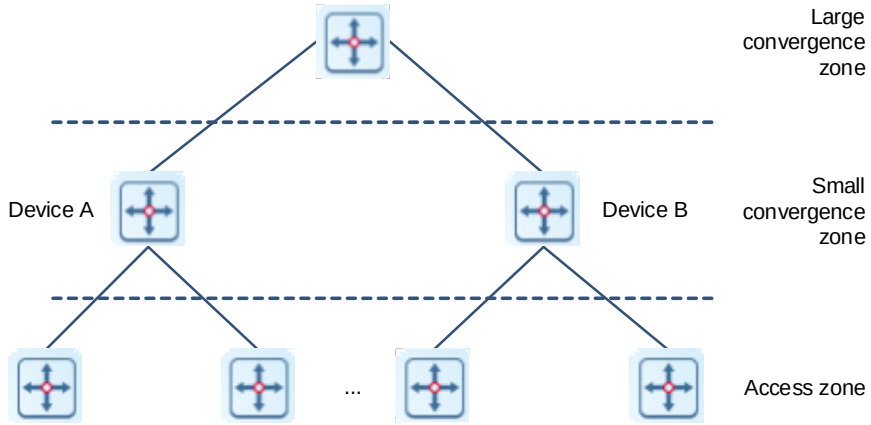
1.5.1 Switching the UFT Mode of a Switch in Standalone Mode

1. Requirements

Take a common campus network as an example. Core devices may be deployed in a small convergence zone as small convergence devices of access devices. L2 functions are enabled on the devices. The small convergence zone must be supported by a large MAC address table. Therefore, you need to set the UFT mode of switches in the small convergence zone to the bridge mode and adjust the forwarding entries to support the pure L2 mode to meet the requirement for the large-capacity MAC address table.

2. Topology

Figure 1-1 Topology for UFT Mode Switching



3. Procedure

Switch the UFT mode of a line card in slot 3 of device A to the bridge mode.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# switch-mode bridge slot 3
Please save current config and restart your device!
```

4. Verification

Run the **display switch-mode status** command to display the configurations.

```
DeviceA# display switch-mode statusDevice No   Slot No      Switch-Mode-Next        Switch-Mode-Current
Status
---------   -------      -------------------- -------------------- ------
```

1	3	bridge	bridge	ok

5. Configuration Files

Configuration file of device A

```
hostname DeviceA
!
switch-mode bridge slot 3
!
end
```

Contents

1 Configuring Supervisor Module Redundancy 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.2 Restrictions and Guidelines 3

1.3 Configuration Task Summary 3

1.4 Configuring RDND 4

1.4.1 Overview 4

1.4.2 Configuring Manual Active/Standby Switchover 4

1.4.3 Configuring an Automatic Synchronization Period 5

1.4.4 Resetting Supervisor Modules 5

1.5 Monitoring 6

1 Configuring Supervisor Module Redundancy

1.1 Introduction

1.1.1 Overview

Redundancy (RDND) is a mechanism that improves device availability by backing up the operating status of services on the supervisor module (SM) in real time.

On a network device with the control plane separated from the forwarding plane, the control plane runs on a SM and the forwarding plane runs on a line card. During the running of the device, the control plane information of the active device is backed up to the standby device in real time. When the active device is suspended as planned (for example, software is upgraded) or unexpectedly (for example, a software or hardware failure occurs), the standby device can quickly take over the work and user configurations are not lost. In this way, the network can continue to normally run. During switchover, the forwarding plane continues with forwarding; during restart of the control plane, forwarding does not stop or links do not flap.

1.1.2 Principles

1. Basic Concepts

- Active/Standby supervisor module

  On a chassis-type device with two SMs, the system elects one as the active SM for normal use purpose, and another as the standby SM for standby purpose. When the active SM fails or actively requires switchover, the standby SM becomes the new active SM. Generally, the standby SM does not participate in device management but monitors the running status of the active SM.

- Global active/standby/candidate

  If there are only two SMs in a virtual switching unit (VSU) system that consists of chassis-type devices, one is used as the active SM to manage the whole VSU system and the other is used as the standby SM of the chassis. If there are three or more SMs in a VSU system, one is selected as the active SM to manage the whole VSU system, another is used as the standby SM to back up the VSU system, and the other modules become candidate SMs. When the active and standby SMs are faulty, the candidate SMs replace them and run in the VSU system.

  In a VSU system that consists of box-type devices, each box-type device is equivalent to a SM and a line card. The system also selects a device as the global active device, another as the global standby device, and other devices as global candidate devices. When the active and standby devices fail, the candidate devices take their place and run in the VSU system.

Note

- Generally, candidate SMs or devices do not participate in backup.
- A candidate SM or device cannot directly replace the active device. When a candidate SM or device has to become the active device, you must restart the candidate SM or device.
- To differentiate the active and standby SMs inside the chassis from those in a VSU system, the active, standby, and candidate SMs in the VSU system are called "global active", "global standby", and "global candidate" SMs respectively. The mechanism of SM redundancy works on the global active and standby

SMs. Therefore, the active and standby concepts mentioned in the VSU environment below refer to the global active and global standby concepts.

- Prerequisites for redundancy of supervisor modules

  The SM redundancy works properly only when the software and hardware of all devices in the device system are compatible.

  During startup, batch synchronization must be performed between the active device and the standby device to ensure their state consistency. Before this process, the SM redundancy does not fully play its role.

- Redundancy state of supervisor modules

  During active-standby backup, the active device has three states as follows:

  - alone, namely alone state. In this state, only one device is running in the system or active/standby switchover is not completed and no redundancy is created between the new active device and the new standby device.
  - batch, namely batch backup state. Redundancy is created between the active device and the standby device and batch backup is in progress.
  - realtime state, namely real-time backup state. Upon batch backup, the active and standby devices are in this state. Real-time backup is performed between the active device and the standby device. Only in this state manual switchover is allowed.

2. Selection of Active and Standby Devices

- Automatic selection of active and standby devices for chassis-type devices

  Users are allowed to insert or remove SMs during running of devices. Based on the current condition of the system, a device automatically selects an engine for running, without affecting normal data exchange. The active SM makes a choice in the following circumstances:

  - If only one SM is inserted into the M1 slot or the M2 slot during startup of the device, the device selects this module as the active one.
  - If two supervisor modules are inserted during startup of the device, by default, the module in the M1 slot is selected as the active one, the module in the M2 slot as the standby one for backup purpose and a prompt is displayed.
  - If only one SM is inserted into the M1 slot or M2 slot during startup of the device and another is inserted during the running of the device, the module inserted later is used as the standby one for backup purpose and outputs prompts.
  - Assume that two SMs are inserted during startup of the device and either is removed during the running of the device (or one SM fails). If the removed or faulty module is the standby SM, the device only displays a prompt for the removal or failure of the standby SM. If the removed or faulty module is the active SM, the other module becomes the active SM and displays a prompt.

- Manual selection of the active and standby SMs

  Users can manually select the active and standby SMs based on different environments as follows:

  - In standalone mode, users can manually perform active/standby switchover, which takes effect after the SMs are reset.

- In VSU mode, users can manually perform active/standby switchover to make the global standby SM become the global active module. If only two SMs exist in a VSU system, the original global active SM becomes the new global standby one after reset. If more than two SMs exist, one global candidate module is selected as the new global standby one and the original global active SM becomes a global candidate one after reset.

3. Information Synchronization of Supervisor Modules

- Status synchronization

  The active SM synchronizes its running status to the standby SM in real time. Thus, the standby module can take over the functions of the active module at any time, without any perceivable change.

- Configuration synchronization

  During the running of the device, two system configuration files exist: **current-config**, which is dynamically generated during running and changes with the service configuration; **startup-config**, which is imported during the startup of the device. The **save** command is run to write **current-config** into **startup-config** or the **copy** command is run to copy the configuration files.

  For some functions that are not directly related to uninterrupted forwarding, the synchronization of system configuration files ensures the consistency of user configurations during switchover.

  In the case of the redundancy of dual SMs, the active module periodically synchronizes the **startup-config** and **current-config** files to the standby module and all candidate modules. Configurations are synchronized in the following operations:

  - When a user switches from the global configuration mode to privileged EXEC mode, the **current-config** file is synchronized.
  - When a user runs the **save** or **copy** command to save configurations, the **startup-config** file is synchronized.
  - For configurations issued over SNMP, the **startup-config** file is not automatically synchronized and the **current-config** file must be synchronized via CLI configuration.

1.2 Restrictions and Guidelines

- For chassis-type devices, the system provides the active-standby backup mechanism. The system supports plug-and-play if the active and standby SMs meet redundancy conditions.
- For box-type devices, each device is equivalent to a SM and a line card. The VSU system of multiple box-type devices also provides the active-standby backup mechanism.

⚠ Caution

For box-type devices, box-type stacked VSUs are used to express the redundancy mechanism.

1.3 Configuration Task Summary

RDND configuration includes the following tasks:

All of the following configuration tasks are optional. Select the task as required.

- (Optional) Configuring Manual Active/Standby Switchover

- (Optional) Configuring an Automatic Synchronization Period
- (Optional) Resetting Supervisor Modules

1.4 Configuring RDND

1.4.1 Overview

- The device automatically selects the active and standby SMs based on the current system, or users do so manually.
- In the SM redundancy environment, the active SM synchronizes status information and configuration files to the standby SM in real time.

1.4.2 Configuring Manual Active/Standby Switchover

1. Overview

If more than two SMs exist in the system, you can manually perform active/standby switchover to change the standby SM into the active SM and select a new standby SM from candidate SMs (the original active SM becomes a candidate SM after reset).

2. Restrictions and Guidelines

- To ensure smooth data forwarding during switchover, perform batch synchronization performed between the active supervisor module and the standby supervisor module so that the two modules are in the same state. That is, manual switchover is allowed only when the redundancy of SMs is in the real-time backup state. In addition, manual active/standby switchover is temporarily disabled by service modules during synchronization to ensure that configuration files are completely synchronized. Therefore, the following conditions must be met for the manual switchover:
  - The operation is performed on the active SM and a standby SM exists.
  - The hot backup of all VSDs in the system is in the real-time state.
  - The active/standby switchover of all VSDs in the system is not temporarily disabled by service modules.
- If devices are virtualized into multiple VSDs, manual switchover can be performed only when the redundancy of the SMs of all VSDs allows switchover.
- You are not advised to include this command for batch command execution, that is, avoid exiting to the privileged EXEC mode immediately after running commands in batches through a script. Otherwise, some asynchronously executed commands may not be synchronized to the standby device, leading to configuration loss upon active/standby switchover.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Perform a active/standby switchover.

redundancy forceswitch

1.4.3 Configuring an Automatic Synchronization Period

1. Overview

By configuring an automatic synchronization cycle, you can adjust the interval for the active SM to synchronize the configuration files (**startup-config** and **current-config**) to change the automatic synchronization period.

2. Restrictions and Guidelines

It is recommended that a short synchronization period be configured so that the active device can frequently synchronize configurations to other member devices when configurations are modified. In this way, the configurations are not lost when services and data are forcibly switched to the standby device due to the failure of the active device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the redundancy configuration mode.

redundancy-mode

(4) Configure a period for automatically synchronizing configuration files.

auto matic-sync period-time *synchronization-interval-time*

The default period for automatically synchronizing configuration files is 1 hour.

1.4.4 Resetting Supervisor Modules

1. Overview

- Resetting only the standby SM does not affect data forwarding. During the resetting of the standby SM, forwarding is not interrupted or user session information is not lost.
- In standalone mode, running the **redundancy reload shelf** command makes the SMs and line cards in the whole chassis reset; in VSU mode, this command is used to reset the device with a specified ID. If two or more devices exist in the system and the device where the global active SM resides is reset, the system performs active/standby switchover.

2. Restrictions and Guidelines

In VSU mode, if the SMs of the system are not in the real-time backup state, resetting the device where the global active SM resides will cause the whole VSU system to reset.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Reset a SM.

redundancy reload { **standby** | **device** [*switch-id*] }

1.5 Monitoring

Run the **display**commands to check the running status of a configured function to verify the configuration effect.

Table 1-1 RDND Monitoring

Command	Purpose
display redundancy status	Displays the current redundancy status of dual SMs.

Contents

1 Configuring Module Hot Swapping 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Working Principle 1

1.2 Configuration Task Summary 1

1.3 Clearing and Modifying Module/Device Configurations 2

1.3.1 Overview 2

1.3.2 Procedure 2

1.4 Monitoring 3

1.5 Configuration Examples 3

1.5.1 Clearing the Configuration of an Offline Module 3

1 Configuring Module Hot Swapping

1.1 Introduction

1.1.1 Overview

Module hot swapping is a maintenance function provided by chassis-type devices. With this function, users can insert or remove a module when a device is operating.

The module hot swapping function supports automatic installation, unloading, reset, and information query of various hot-swappable modules (including Supervisor Engines, line cards, switch fabric modules, and multi-service card) on chassis-type devices.

1.1.2 Working Principle

1. Automatic Installation of an Inserted Module

Users can insert or remove a module (in hot swapping mode) when a device is operating. The operation of other modules is not affected. After a module is inserted into a slot, the management software of the device tries to install the driver of the module automatically. The configuration of a removed module is retained for future configuration. If a removed module is inserted again, the module is automatically started and its configuration becomes effective.

Note

Modules here refer to Supervisor Engines, line cards, switch fabric modules, and multi-service cards. Supervisor Engines can be inserted into only the Supervisor Engine slot (M1 or M2), line cards and multi-service cards can be inserted into line card slots, and switch fabric modules can be inserted into only switch fabric module slots.

2. Resetting an Online Module

The management software of a device provides the online module reset function so that users can reset an online module to rectify software failures occurring on the module.

Caution

Resetting an online module may interrupt some services on the device.

1.2 Configuration Task Summary

Note

The module hot swapping function is automatically implemented and no manual configuration is required.

The module hot swapping configuration includes the following tasks:

- Clearing and Modifying Module/Device Configurations

1.3 Clearing and Modifying Module/Device Configurations

1.3.1 Overview

In different circumstances, different configurations need to be cleared. The specific application scenarios are described as follows:

- Clearing the configurations of a module; A user inserts a board card into a slot on a device. Later, the user removes the board card and deletes the configurations of the board card and its port because the board card is no longer used.
- Clearing the configuration of a VSU member device: A user uses a member device in a VSU. Later, the user removes the device from the VSU and deletes the configurations of the member device in the VSU.
- Deleting the MAC address retained in the configuration file: A user hopes to change the MAC address of the system and uses the new MAC address elected by the system. Generally, the MAC address used by the system is stored in the Supervisor Engine or flash memory of the chassis. In VSU mode, however, the system automatically stores the used MAC address in the configuration file to avoid interruption caused by the change of the MAC address. If a valid MAC address exists in the configuration file after restart of the device, the MAC address is used in preference.
- Modifying the MAC address retained in the configuration file; The function is used when the system MAC address of a device needs to be modified. In gateway mode (the **auth-mode gateway** command is configured in the system), the gateway MAC address is bound on some peripheral devices. If the gateway is replaced, users can run the **sysmac** command to set the gateway MAC address to the MAC address of the new gateway so that configurations do not need to be modified for downstream devices with the gateway MAC address bound. The **sysmac** command is available only when the system is configured to work in gateway mode.

1.3.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Clear the configurations of a module.

remove configuration module *slot-number*

This command is applicable to only chassis-type devices.

Note

This command is supported by only the S6920-4C, and S6510-4C.

Warning

Do not run the command on online line cards. Otherwise, network loops may be caused after the anti-loop configurations of the online line cards are deleted.

(4) (Optional) Modify the MAC address retained in the configuration file.

sysmac *mac-address*

No MAC address is specified for the system by default.

1.4 Monitoring

This section describes the **display** commands used for checking the running status of a configured function to verify the configuration effect.

You can run the **reset** command to reset a module.

⚠ Caution

Running the **reset module** command may reset a module and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
display version module detail [*slot-number*] **display version module detail** [*device-id* / *slot-number*] (in VSU mode)	Displays the details of a module.
display version slots [*slot-number*] **display version slots** [*device-id* / *slot-number*] (in VSU mode)	Displays the online status of a module.
display manuinfo	Displays the asset information of all independent components in the current system.
display sysmac	Displays the current MAC address of the system.
display alarm	Displays system-level alarms.
display intp tipc status	Displays the communication status of the Network Time Protocol (NTP) synchronization module.

1.5 Configuration Examples

1.5.1 Clearing the Configuration of an Offline Module

1. Requirements

A user inserted a board card into slot 1 before. The board card is no longer used because the network topology is changed. The user hopes to delete the port configuration of the board card so that the configuration file of the device can become concise.

2. Notes

Run the module configuration clearing command to delete the configuration of the board card.

3. Procedure

```
Device# privilege
Device# config-system terminal
Device(cfg)# remove configuration module 1
```

4. Verification

Run the **display version slots** command to verify that the configuration of the board card in slot 1 is removed.

```
Device# display version slots
Dev   Slot   Port Configured Module     Online Module        Software Status
---   ----   ---- ------------------     -----                --------------
1     1      0    none                  none                 none
1     2      24   M8606-24SFP/12GT      M8606-24SFP/12GT     none
1     3      2    M8606-2XFP            M8606-2XFP           cannot startup
1     4      24   M8606-24GT/12SFP      M8606-24GT/12SFP     ok
1     M1     0    N/A                   M8606-CM             master
1     M2     0    N/A                   none                 none
```

Contents

1 Configuring Process Restarting 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Restrictions and Guidelines 1
1.3 Configuration Task Summary 1
1.4 Configuring RAS-CMDK 2
1.4.1 Overview 2
1.4.2 Restarting a Process of the Member Device in a Specified Slot on a Specified Device 2
1.4.3 Starting a Process of the Member Device in a Specified Slot on a Specified Device 2
1.4.4 Stopping a Process of the Member Device in a Specified Slot on a Specified Device 2
1.4.5 Displaying Processes that can Be Restarted on the Member Device in a Specified Slot on a Specified Device 3
1.5 Monitoring 3

1 Configuring Process Restarting

1.1 Introduction

1.1.1 Overview

The command line interface (CLI) reboot process module (CMDK) provides a means of restarting a back-end process of a device on the CLI for users. Therefore, when the function of a service fails on a device, the user can restart a specified process of the member device in a specified slot on a specified device in a cluster rather than restart the device, to improve device availability.

The CMDK has the following features:

- Improve the device availability. When a process in a device becomes abnormal, users can manually restart a specified process.
- Support process restart in the virtual switching unit (VSU) environment. The CMDK can restart a specified process of the member device in a specified slot on a specified device. Therefore, users can restart a process on the master device or master supervisor module.

1.1.2 Principles

First, a process determines whether the current device is in the standalone or VSU environment. Then, the process sends a process kill message to a target node. After receiving the message, the target node restarts the whole process by running the command in the startup script.

1.2 Restrictions and Guidelines

Configuration commands are run on only the master device or master supervisor module. Other member devices do not have the permission to run the commands.

1.3 Configuration Task Summary

The RAS-CMDK configuration includes the following tasks: The following configuration tasks are optional. Configure the task as required.

- Restarting a Process of the Member Device in a Specified Slot on a Specified Device
- Starting a Process of the Member Device in a Specified Slot on a Specified Device
- Stopping a Process of the Member Device in a Specified Slot on a Specified Device
- Displaying Processes that can Be Restarted on the Member Device in a Specified Slot on a Specified Device

1.4 Configuring RAS-CMDK

1.4.1 Overview

The CLI CMDK provides a means of restarting a back-end process of a device on the CLI for users. Therefore, when the function of a service fails on a device, the user can restart a specified process of the member device in a specified slot on a specified device in a cluster rather than restart the device, to improve device availability.

1.4.2 Restarting a Process of the Member Device in a Specified Slot on a Specified Device

1. Overview

This section describes how to restart a specified process on a specified device.

2. Restrictions and Guidelines

Only valid processes excluded from the whitelist can be restarted.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Restart a process of the member device in a specified slot on a specified device.

cmdk device *device-id* **slot** *slot-id* **module** *module-name* **restart**

1.4.3 Starting a Process of the Member Device in a Specified Slot on a Specified Device

1. Overview

This section describes how to start a specified process on a specified device.

2. Restrictions and Guidelines

Only valid processes can be started.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Start a process of the member device in a specified slot on a specified device.

cmdk device *device-id* **slot** *slot-id* **module** *module-name* **start**

1.4.4 Stopping a Process of the Member Device in a Specified Slot on a Specified Device

1. Overview

This section describes how to stop a specified process on a specified device.

2. Restrictions and Guidelines

Only valid processes excluded from the whitelist can be stopped.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Stop a process of the member device in a specified slot on a specified device.

cmdk device *device-id* **slot** *slot-id* **module** *module-name* **stop**

1.4.5 Displaying Processes that can Be Restarted on the Member Device in a Specified Slot on a Specified Device

1. Overview

This section describes how to display processes that can be restarted on the member device in a specified slot on a device.

2. Restrictions and Guidelines

The processes that can be restarted on the member device in a specified slot are displayed only on the master device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Display the processes that can be restarted on the member device in a specified slot on a specified device.

cmdk device *device-id* **slot** *slot-id* **detail**

1.5 Monitoring

Run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 RAS-CMDK Monitoring

Command	Purpose
debugging cmdk on	Enables the CMDK debugging mode.
debugging cmdk off	Disables the CMDK debugging mode.

Contents

1 Configuring Python....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.2 Restrictions and Guidelines....1
1.3 Configuring Basic Features....1
1.3.1 Overview....1
1.3.2 Procedure....1
2 Configuring APS....2
1.1 Introduction....2
1.2 Overview....2
1.3 Principles....2
1.4 Restrictions and Guidelines....2
1.5 Configuring Basic Features....2
1.5.1 Overview....2
1.5.2 Procedure....2

1 Configuring Python

1.1 Introduction

1.1.1 Overview

Python is an object-oriented interpretive computer programming language and is pure free software. Its source code and interpreter Cpython comply with GNU general public license (GPL) protocol. The Python shell component can debugging and run Python scripts through CLI commands.

1.1.2 Principles

1. Python Script Debugging

Python script debugging refers to debugging or running a script to be run on a device. A Python script is debugged with two methods:

- One is to copy the Python script to a device and debugging the script on the Python console.
- The other is to copy the Python script to a device and run the script. Check whether the expected effect is achieved. If not, modify and debugging the script.

Regardless of methods, the user runs an open source Python command and the debugging method is consistent with that of the open source Python. The Python shell component redirects the input and output of the open source Python process to the current terminal.

2. Permission Control

Permission control indicates that Python CLI commands are not allowed to operate directories or files other than **flash:** and **tmp:** and risky operations such as popen and system are not allowed in a script.

Permissions are controlled at the entry of Python. So long as a Python script is run on a device, permissions are controlled to control the risk caused by running of the Python script.

1.2 Restrictions and Guidelines

Permission control: Python commands and scripts can be run on only files in the **flash:** and **tmp:** directories.

1.3 Configuring Basic Features

1.3.1 Overview

This section describes how to run a Python script.

1.3.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Run the Python script.

python *file-name* [args]

2 Configuring APS

1.1 Introduction

1.2 Overview

The Autonomous Programming System (APS) is a system that enables the system with open programmability through the unified application API.

With limited functions and known services, the conventional network device is not dynamic and flexible. With popularization and growth of the network, diversity and difference requirements arise. Some customers expect openness of different levels from devices, which allow them to develop customized functions and deploy private management policies to implement automated O&M and reduce costs.

APS provides network devices with open programmability. A Python application API allows customers to run commands as scheduled through the Python script.

1.3 Principles

APS allows users to manages devices through the open API.

1.4 Restrictions and Guidelines

- Permission control. The python command and script only allow operation on the files in the **flash:** and **tmp:** directories.

1.5 Configuring Basic Features

1.5.1 Overview

The APS provides CLI commands and API.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the APS mode.
aps

(3) Run the Python script.
aps run python *flash:test.py*

Contents

1 Configuring License Management 1
1.1 Introduction 1
1.1.1 License Management Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 4
1.3 Configuring License Management 4
1.3.1 Overview 4
1.3.2 Installing License File 4
1.3.3 Backing Up License File 5
1.3.4 Configuring Grace Period Warning 6
1.3.5 Updating License File 6
1.3.6 Uninstalling License 7
1.3.7 Unbinding License 7
1.4 Monitoring 8

1 Configuring License Management

1.1 Introduction

1.1.1 License Management Overview

A user needs to install a correct license file to use some extension functions of a device. The extension functions provided by Ruijie General Operating System (RGOS) include the virtual switch device (VSD), Transparent Interconnection of Lots of Links (TRILL), Fibre Channel over Ethernet (FCoE), user connections supported by the access point (AP), and user concurrency supported by virtual private network over Secure Sockets Layer (SSLVPN). A user can use the general and extension functions of the RGOS only after obtaining a license.

Note

Generally, all features of the RGOS are pre-installed in the factory. For feature licenses and license types, see their configuration guides. RGOS features or basic functions that are not expressly specified can be used directly without licenses.

1.1.2 Principles

1. Basic Concepts

- Software to be licensed

  A software function used by only a licensed user.

- Feature license

  As a license for a special feature, it is obtained through a license file, hardware entity, or legal contract. This license specifies the maximum number of supported users, the maximum number of supported instances, and valid period.

- Product authorization key (PAK)

  It is a manner of granting a license to a user by providing a license code.

- Purchase voucher

  A voucher of purchasing a license. The voucher contains the PAK and license download address. The legitimate owner of the purchase voucher logs in to the license download address, and uses the PAK to sign up. The license file is downloaded from the displayed link or is directly sent to the registered email.

- Host ID

  A unique serial number for identifying each device.

Caution

A license matches the host ID of any device added to a VSU. When the device is removed from the VSU, the feature enabled by the license becomes invalid.

- License file

  A file that controls the features of a licensed software. A license file is generated by using a special tool based on contract information, and each license file contains a digital signature to prevent tampering. The license file serves a product. That is, one license file is used on only one device and matches the device's host ID. License is not transferred. Licensed functions are available when the license file is installed.

- Temporary license

  A temporary license becomes invalid after the valid period expires.

- Evaluation license

  An evaluation license is temporary, and has been installed on the device before delivery. It is used to provide users with features for trial use. This type of license is independent of the host ID.

- Permanent license

  This license has no time limit and is permanently effective upon purchase by the user.

- License stacking

  Different licenses can be used on one device. For example, if a device provides both the FCoE and TRILL functions, users can purchase the licenses of the two functions and use them on the same device.

- Single-Instance license

  Only one license can be installed for one feature at a time.

⚠ Caution

Licenses for the features like user concurrency supported by VSD, TRILL, FCoE, and SSLVPN are all single-instance licenses.

- Multi-instance license

  Multiple license files can be installed for the same feature on a device.

⚠ Caution

Only the license for the feature "user connections supported by the AP" is a multi-instance license.

- Grace period warning

  A grace period warning is generated in forms of log and trap messages 180 days before a license expires. Setting a grace period warning prevents licensed software from being affected when an aging license suddenly stops working upon expiration.

2. License Acquisition and Use

- Use of License

  A license must be obtained from Ruijie's official website or marketing channel. It is device-specific. You log in to the website specified in the purchase voucher and provide the PAK and host ID to obtain the license file, which is directly downloaded or given through an email. After acquisition, you need to install the license file. Afterwards, you can use the features of the license.

  RGOS licenses include permanent license and temporary license. When a user first enables a temporary license to use a feature, the valid period of the license starts, and after the period ends, the feature will be disabled. To continue using this feature, this user can purchase another license (permanent or temporary) from Ruijie's official website or marketing channel.

Caution

In the VSU environment, multiple devices form one virtual device. In this scenario, if one device obtains the license for a feature, the whole VSU is licensed to use this feature. But when the devices in the VSU are separated, the license is checked by device.

The VSD function provided by the RGOS helps divide one RGOS device into multiple virtual devices. For external users, each virtual device works like an independent device. That is, when the license for a feature is obtained, the license for this feature is effective to all VSD domains. In addition, licenses are installed and managed in the default VSD only.

- License check

  After a device starts to run, the license for each feature needs to be checked. If the corresponding license is properly installed, this feature is ready for application. Otherwise, this feature becomes invalid and cannot be used.

  Note

  Checking the licenses for various features occurs at different times. The licenses for some features are checked during startup, and the licenses for other features are checked in real time.

- Loss of License File

  License files are stored in the **/data** directory of a device and is not lost after software upgrade.

  If the memory or file system is damaged during product use but the license file has been backed up, you can install the backup license file again after system recovery. If the license file is not backed up, you can visit Ruijie's website to obtain the license file again. This process is the same as the previous acquisition process (you do not need to purchase the license again).

  Licensing is device-based. After a license is provided for a specified device, the license file is verified only on this device. Upon maintenance or replacement of a device, the host ID may change and the obtained license file may not be identified. In this case, contact Ruijie after-sales personnel.

3. Backup, Update, and Uninstallation of a License File

- License File Backup

  When a fault such as damage of file system storage media occurs on a device, the license file on this device may be lost after you troubleshoot this fault. Therefore, you need to back up the license file in advance to reinstall it after you troubleshoot the fault.

- License File Update

  If the existing license of the system fails to meet the feature requirements, you can visit Ruijie's website to purchase a desired license and then update the license file locally.

- License File Uninstallation

  If you do not need a feature, you can uninstall the license for this feature to improve the utilization of resources like the memory. To reuse this feature, reinstall the license. It is recommended that license files be kept properly.

1.2 Configuration Task Summary

Configuration of license management includes the following tasks:

(1) Installing License File

(2) Backing Up License File

(3) (Optional) Grace Period Warning

(4) (Optional) Updating License File

(5) (Optional) Uninstalling License

(6) (Optional) Unbinding License

1.3 Configuring License Management

1.3.1 Overview

This section describes how to enable or disable a feature by installing or uninstalling a license file.

1.3.2 Installing License File

1. Overview

This section describes how to install a license file to enable a licensed feature.

2. Restrictions and Guidelines

- Ruijie provides a paper purchase voucher that contains a PAK when you purchase a license.
- Log in to Ruijie's website and obtain a license file as prompted.
- If you use a feature without being licensed, the CLI window displays a prompt, indicating that the feature is not licensed and is unavailable, and provides a website to download a license file.
- After downloading a license file from the specified website, upload this license file to the device (or store it in a USB flash drive) for installation. You can install the license file offline. Copy the license file to the file system of the device. Use conventional file system operation commands. For example, download the license file through Trivial File Transfer Protocol (TFTP) or copy the license file to a USB flash drive.
- After a license file is installed on a device, this license file is automatically backed up in the **/data** directory of the system (the license file is suffixed with **.lic**). The backup license file is deleted when the installed license is uninstalled.
- In a VSU environment, run the **license install** command on a master supervisor module to install the license file on all devices, or run this command on a non-CM supervisor module to install the license file only on the device from which you enter commands.
- In a VSU environment, run the **license auto-install** command to automatically match compliant devices to install the license file. In a non-VSU environment, run the following commands to do the same:
- In a VSU environment, when the license for a feature is obtained by one device by running the **license install** command, the whole VSU system is licensed to use this feature. But when the devices in the VSU are separated, the licensed device can still use the license.
- In a VSU environment, run the **license install** command on a CM supervisor module to install a license on all boards, or run the **license auto-install** command on a CM supervisor module to install a license on a

matched board only. Run the **license install** command on a non-CM supervisor module to install a license on the local device, or run the **license auto-install** command on a non-CM supervisor module to automatically match and install a license on all devices.

- The **license auto-install** command enables the auto-matching function only in a VSU environment, and has the same function as the **license install** command in other environments.
- In a VSD environment, licenses are managed in global configuration mode. When the license for a certain feature is obtained, the license for this feature is effective to all VSD domains. In addition, licenses are installed and managed in the default VSD only.
- Different devices do not share the same license file.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Obtain **host-id** of the device.

display license hostid

The name of a license file cannot be modified.

This command can be executed without permission.

(3) (Optional) Install the license file.

license install { **flash:** | **usb0:** } *filename*

license auto-install { **flash:** | **usb0:** } *filename*

1.3.3 Backing Up License File

1. Overview

This section describes how to back up the license files of one or all features in the system.

2. Restrictions and Guidelines

- Evaluation license files cannot be backed up.
- Sufficient storage space is required to store backup license files. Generally, one license file is 4 KB to 10 KB.
- Backup license files are normal files.
- When all license files in the system are backed up, a .tar file is generated.
- You can run the **dir** command to save the license files to other storage devices, for example, a USB flash drive.
- After backing up license files, you can run the **dir** command to confirm the generated license files, and compare them with the license file names displayed in the **installed license** fields of permanently licensed features in the command output of the **display license all-license** or **display license file** command, to check whether licenses are correctly backed up.
- The **installed license** field displayed in the **display license all-license** command and **display license file** command is available only for multi-instance licenses. The name of a file internally backed up in the multi-instance license system is determined by the ID of the multi-instance license. One single-instance license file exists at most in the system at a time. Therefore, the internally backed up single-instance license file is named after the feature.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Back up the license files in the system.

license { **copy-all** | **copy-file** *filename* } { **flash:** | **usb0:** } [*target-filename*]

1.3.4 Configuring Grace Period Warning

1. Overview

This section describes how to evaluate the expiration date of a license and how to provide early warnings in form of log messages to remind users of taking actions in advance.

2. Restrictions and Guidelines

- Each licensed feature is separately configured.
- The grace period warning is affected by the device time.
- A permanent license file needs no grace period warning.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Set the grace period for reminding the user of the expiration of a license.

license grace-period *filename time*

The grace period before the expiration of a license is 180 days by default.

1.3.5 Updating License File

1. Overview

This section describes how to update the license file for a feature of the system. Generally, this function is performed to update an evaluation license file into a temporary license file.

2. Restrictions and Guidelines

- A formal permanent license file does not need to be updated.
- A later license version cannot be updated into an earlier one.
- A license file contains a generation time field. The value of this field is the time when the license file is generated on the website. When the value is closer to the current time, the license file has a later version.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Update the license file.

license update { **flash:** | **usb0:** } *filename*

1.3.6 Uninstalling License

1. Overview

This section describes how to uninstall one or all feature licenses in the system when the features are not needed.

Note

This function becomes invalid when the corresponding license is uninstalled.

2. Restrictions and Guidelines

- If the licensed feature is running, the uninstallation does not take effect immediately, but takes effect after the feature is enabled or the device is restarted next time.
- You are advised to back up a license file before uninstallation to reinstall an uninstalled license when needed.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Uninstall a license in the system.

license uninstall { **all** | *license* [*filename*] }

license auto-uninstall *devid license* [*filename*]

This command can be executed without permission.

1.3.7 Unbinding License

1. Overview

If you want to unbind a license from a device on the license website, unbind the license on the device first.

2. Restrictions and Guidelines

- Upon unbinding a license, you obtain a verification code, which is used to unbind the license on the license website.
- After the license is unbound, the corresponding license file cannot be reinstalled.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Unbind a license in the system.

license unbind *pak*

license auto-unbind *device-id pak*

This command can be executed without permission.

1.4 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Table 1-1 Monitoring

Command	Purpose
display license all-license	Displays the list of all installed license files on the device.
display license file *file-license*	Displays the information about a specified license file on the device.
display license dev-license	Displays license configurations of all devices.
display license usage	Displays licenses in use.
display license hostid	Displays the device SN used for licensing.
display license dev-hostid	Displays the device SN of each device
display license unbind-code	Displays the verification code of a license unbound from the device.
display license dev-unbind-code	Displays the verification code of a license unbound from each device in the environment.

Contents

1 Configuring Hung Task Checking 1

1.1 Overview 1

1.2 Configuration Task Summary 1

1.3 Configuring Basic Parameters 1

1.3.1 Overview 1

1.3.2 Procedure 1

1.4 Monitoring 1

1.5 Configuration Examples 2

1.5.1 Configuring Basic Parameters 2

1 Configuring Hung Task Checking

1.1 Overview

Hung task checking is performed in the kernel to detect processes that are stalled into the D state (TASK_UNINTERRUPTIBLE state) and take measures as needed, which prevents these processes from staying in the D state indefinitely and failing in waking up.

1.2 Configuration Task Summary

Configuring hung task checking includes the following tasks:

- Configuring Basic Parameters

1.3 Configuring Basic Parameters

1.3.1 Overview

After you enable the hung task checking and device reset function and configure the checking interval and the maximum number of alerts, the kernel determines whether a process stays in D state for a long period by checking whether a process switchover occurs within the specified interval.

1.3.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the hung task checking and device reset function.

hungtask panic { **enable** | **disable** }

The hung task checking and device reset function is enabled by default.

(4) Configure the hung task checking interval.

hungtask timeout-secs *secs-value*

The default checking interval is 3,600 seconds.

(5) Configure the maximum number of alerts that a device reports for hung task checking.

hungtask warnings *warnings-value*

The default number of alerts that a device reports for hung task checking is 10.

1.4 Monitoring

Run the **display** command to check the configuration.

Caution

System resources are occupied when debugging information is output. Therefore, disable debugging immediately after use.

Table 1-1 Monitoring

Command	Purpose
display hungtask config	Displays the configuration of the hung task checking function.
display hungtask status	Displays the basic parameters of the hung task checking function.

1.5 Configuration Examples

1.5.1 Configuring Basic Parameters

1. Requirements

Configure and enable the hung task checking and device reset function, and set the checking interval to 600 seconds and the maximum number of alerts to 20.

2. Procedure

Configure and enable the hung task checking and device reset function, and set the checking interval to 600 seconds and the maximum number of alerts to 20.

```
Hostname> privilege
Hostname# config-system terminal
Hostname(cfg)# hungtask panic enable
Hostname(cfg)# hungtask timeout-secs 600
Hostname(cfg)# hungtask warnings 20
```

Contents

1 Configuring TCAM Mode 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.2 Restrictions and Guidelines 2

1.3 Configuring Basic Features 2

1.3.1 Overview 2

1.3.2 Restrictions and Guidelines 2

1.3.3 Procedure 2

1.4 Monitoring 3

1.5 Configuration Examples 3

1.5.1 Configuring TCAM Modes 3

1 Configuring TCAM Mode

Note

Ternary Content Addressable Memory (TCAM) mode is not supported by the S6980-64QC, S6580-48CQ8QC or S6580-24DC8QC.

1.1 Introduction

1.1.1 Overview

TCAM is a type of content-addressable memory that is used for quick searching of access control list (ACL) rules and route table entries. In this chapter, TCAM resources are used for quick queries of ACL rules. TCAM resources of switches are used for different types of services and in various ways. In most cases where resources are dynamically allocated, TCAM resources are integrated in specific scenarios due to limited TCAM resources. TCAM modes support flexibly allocating TCAM resources.

1.1.2 Principles

1. TCAM Resource Allocation

By default, TCAM services are not integrated. In this case, a service consumes at least one or two TCAM resources. For example, TCAM services are not available in a dual-stack scenario due to various constraints on TCAM resources. TCAM resources can be integrated in scenarios where multiple TCAM services exist and do not conflict with each other. You can configure a proper TCAM mode based on a scenario to meet the requirements for TCAM resources.

Take the following services to be integrated as examples. Non-IPv6 Classification, Policing, and Prioritization (CPP) services are referred to as CPP V4 services and IPv6 CPP services are referred to as CPP V6 services. The in keyword represents that the service takes effect on the inflow traffic, and the out keyword represents that the service takes effect on the outflow traffic.

- CPP V4 service and sub-interface route-sample V4 in service
- CPP V6 service and sub-interface route-sample V6 in service
- ACL V4 in flow statistics (counter-only) service and QoS Remark V4 in service
- ACL V6 in flow statistics (counter-only) service and QoS Remark V6 in service

2. TCAM Modes

TCAM modes allow you to select a proper mode for a specific scenario as needed.
The following modes are supported. The modes take effect only after the configuration is saved and the device restarts.

- Default mode

  Default mode is configured for a switch by default, where multiple TCAM services do not integrate TCAM resources. This mode is suitable for scenarios with a moderate number of TCAM services.

- ACL-MAX mode

You can configure ACL-MAX mode for a switch. In this mode, TCAM services can maximize the use of TCAM resources, and the non-conflicting TCAM services will integrate TCAM resources.

- ACL-SLICES-ALL

  You can configure ACL-SLICES-ALL mode for a switch. In this mode, TCAM services can maximize the use of TCAM resources, and TCAM services will integrate TCAM resources.

- VXLAN

  You can configure VXLAN mode for a switch. In this mode, TCAM services will integrate TCAM resources in Virtual Extensible LAN (VXLAN) scenarios.

- VXLAN-HASH

  You can configure VXLAN-HASH mode for a switch. In this mode, TCAM services will integrate TCAM resources in VXLAN scenarios that support VXLAN hash balancing.

- ACL-DEFAULT-VXLAN-HASH

  You can configure ACL-DEFAULT-VXLAN-HASH mode for a switch. In this mode, TCAM services will integrate TCAM resources in common scenarios. This mode applies to the default mode scenario that supports VXLAN hash balancing.

1.2 Restrictions and Guidelines

- After the TCAM mode is configured, you must save the configuration and restart the system to make the configuration take effect.
- If you change the mode to Default mode, you must also save the configuration and restart the system to make the configuration take effect.

1.3 Configuring Basic Features

1.3.1 Overview

TCAM modes allow you to select a proper mode for a specific scenario as needed.

1.3.2 Restrictions and Guidelines

- After the TCAM mode is configured, you must save the configuration and restart the system to make the configuration take effect.
- If you change the mode to Default mode, you must also save the configuration and restart the system to make the configuration take effect.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the TCAM mode.

tcam-mode *mode-type*

The Default mode is configured for the switch by default.

The following TCAM modes are supported:

- default: Default mode, which is applicable to most scenarios.
- acl-max: ACL-MAX mode, which is suitable for scenarios where non-conflicting TCAM services integrate TCAM resources and maximize the use of resources.
- acl-slices-all: ACL-SLICES-ALL mode, which is suitable for scenarios where TCAM services integrate TCAM resources and maximize the use of resources (supported only by S6510 devices).
- vxlan: VXLAN mode, which is suitable for VXLAN scenarios.
- vxlan-hash: VXLAN-HASH mode, which is suitable for VXLAN scenarios that support VXLAN hash balancing.
- acl-default-vxlan-hash: ACL-DEFAULT-VXLAN-HASH mode, which is suitable for Default mode scenarios that support VXLAN hash balancing.

Note

The S6920-4C supports only the following values: default and acl-max. The S6510-48VS8CQ and S6250-48XS8CQ support only the following values: default, acl-max, and acl-slices-all.

1.4 Monitoring

Table 1-1 Monitoring

Command	Purpose
display tcam-mode status	Displays the TCAM mode of a switch.

1.5 Configuration Examples

1.5.1 Configuring TCAM Modes

1. Procedure

The following example configures the ACL-MAX mode for the device.

```
Device# config-system terminal
Device(cfg)# tcam-mode acl-max
Please remove ACL related config, save current config and restart your device!
Device(cfg)# end
Devic# save

Building configuration...

[OK]
Device# display current-config

Building configuration...
```

```
Current configuration : 2597 bytes


...
tcam-mode acl-max
```

2. Verification

Run the **display tcam-mode status** command to display the configurations.

```
Device(cfg)# display tcam-mode status
config-mode: acl-max
save-mode: acl-max
running-mode: acl-max
```

Contents

1 Handling the Parity Check Error and ECC Error....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.2 Configuring the Auto-restart and Recovery Command....1
1.2.1 Overview....1
1.2.2 Procedure....1
1.3 Configuring the System Log Printing Command....2
1.3.1 Overview....2
1.3.2 Procedure....2
1.4 Monitoring....2

1 Handling the Parity Check Error and ECC Error

1.1 Introduction

1.1.1 Overview

Parity check and error-correction code (ECC) errors occur when an occasional bit flip happens in the register of an application-specific integrated circuit (ASIC) chip, which may cause exceptions in the chip.

In most cases, the errors are caused by electrostatic or electromagnetic interference, such as radiation (neutrons from cosmic rays), electromagnetic interference (EMI), and electrostatic discharge (ESD). These events can randomly cause bit flips in one or more storage units. The sources of the parity check error could be the power supply, cable, optical system, generator, nuclear facility (radiation), and solar radiation. The parity check error can also be caused by hardware failures that may result from electrical surge (one possible cause is no proper grounding), ESD, chip overheating or cooling, and production defects.

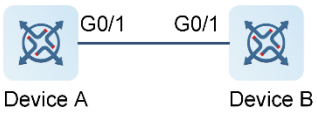
To resolve this type of data error, the chip manufacturers introduce a soft error recovery (SER) mechanism for the software, which maintains a software cache recording table entries in the chip. When a parity check error is reported in the chip, an interrupt handler is triggered to deal with the error, and the correct content in the cache is written into the ASIC chip.

If the software cannot fix the error, the device is restarted to perform recovery, which affects normal services. In this case, a command is required to control the restart.

1.1.2 Principles

In this example, a forwarding exception occurs due to the parity check error. In the following figure, Device A communicates with Device B through the physical interface G0/1. A parity check error that causes a forwarding exception occurs on Device A and cannot be fixed. If the auto-restart and recovery command is configured on Device A, Device A is restarted to resolve the issue.

Figure 1-1 Physical Connection Between Switches



1.2 Configuring the Auto-restart and Recovery Command

1.2.1 Overview

By default, the device does not restart and the system logs are not printed when uncorrectable parity check and ECC errors occur.

A command is provided to allow the administrator to configure the auto-restart and recovery function when uncorrectable errors occur.

⚠ Caution

If the auto-restart and recovery function is enabled, the device or board restarts for recovery upon an uncorrectable error, which may affect data forwarding.

1.2.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the auto-restart and recovery function.

parity-ecc-error auto-recover enable

The auto-restart and recovery function is disabled by default.

1.3 Configuring the System Log Printing Command

1.3.1 Overview

By default, the device does not restart and the system logs are not printed when uncorrectable parity check and ECC errors occur.

A command is provided to allow the administrator to configure the system log printing function when uncorrectable errors occur.

1.3.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the system log printing function.

parity-ecc-error syslog on

The system log printing function is disabled by default.

1.4 Monitoring

Run the **display** command to display the configurations.

Table 1-1 Monitoring

Command	Purpose
display current-config	Displays configurations.

Alarm Management Configuration

1. Configuring Alarm Management

Contents

1 Configuring Alarm Management 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.2 Configuration Task Summary 2

1.3 Configuring Alarm Filtering 2

1.3.1 Overview 2

1.3.2 Restrictions and Guidelines 2

1.3.3 Procedure 2

1.4 Configuring Alarm Suppression 2

1.4.1 Overview 2

1.4.2 Procedure 3

1.5 Monitoring 3

1.6 Configuration Examples 4

1.6.1 Configuring Alarm Management 4

1 Configuring Alarm Management

1.1 Introduction

1.1.1 Overview

A network device failure could cause exceptions in services. Therefore, when a failure occurs, you need to locate and resolve the fault immediately. An alarm is reported when the device detects a fault. You can view the alarm containing the fault information to locate the fault.

As the network grows in size and complexity, the failure of a service module may cause a large number of alarms. If all alarms are sent to the network management server or recorded and output through syslog, unnecessary system resources are consumed, and fault information is not collected effectively.

The alarm management function helps manage the statistics and report of alarms and events. In this way, alarms can be reported in a more effective and accurate manner, which assists the network administrators in locating and diagnosing faults efficiently.

- The alarm filtering function can filter out unwanted alarms and helps you obtain valid information about faults.
- The alarm suppression function can control the number of flapping alarms that are reported in a short period and filter out the unstable alarms, which ensures that alarms are reported in an effective way.

1.1.2 Principles

1. Basic Concepts

- Active Alarm

  When detecting a fault, the service module sends a fault notification to the alarm management module. This notification is displayed as an active alarm on the device.

- Historical Alarm

  When detecting that a fault is fixed, the service module sends a fault recovery notification to the alarm management module. In this case, the alarm is not displayed on the device. The cleared alarms are recorded as historical alarms.

- Event

  When an object or a state on a device changes, an event notification is generated. An alarm can persist while an event represents only a change on the device.

2. Report of Active Alarms

An active alarm indicates a persistent fault. The active alarm persists if the fault is not fixed.

When an alarm is reported by the service module, the alarm management module checks whether the alarm exists in the alarm database first. If the alarm already exists, the alarm is not handled. If the alarm does not exist in the database, the alarm management module checks whether the number of active alarms on the device has reached the upper limit. If not, the alarm is sent to the network management server or recorded and output

through syslog. If the number has reached the upper limit, the alarm is discarded and the alarm information is recorded in internal logs.

3. Record of Historical Alarms

Historical alarms record the information about cleared alarms, which helps locate faults.

When the service module reports alarm clearance, the alarm management module deletes this active alarm, records the clearance time, and creates an alarm record. The statistical count and sequence numbers of alarms are also updated. If the maximum number of alarm records is reached, the earliest historical alarm record in the database is deleted.

1.2 Configuration Task Summary

The alarm management configuration includes the following tasks:

- (Optional) Configuring Alarm Filtering
- (Optional) Configuring Alarm Suppression

1.3 Configuring Alarm Filtering

1.3.1 Overview

Redundant alarms often hinder the O&M personnel from locating the fault. You can configure the alarm filtering function to filter out unwanted alarms.

1.3.2 Restrictions and Guidelines

- You can configure different conditions to filter out alarms. When an alarm that meets one filter condition is generated, the alarm is filtered out.
- To cancel the filtering of a specific alarm, ensure that the alarm does not match any filter condition that you configure.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the alarm filtering function and configure the filter condition.

alarm mask { **alarm-name** *alarm-name* | **all** | **slot** *slot-id* | **type** *alarm-type* }

The alarm filtering function is disabled by default.

1.4 Configuring Alarm Suppression

1.4.1 Overview

The underlying status may fluctuate on some service modules. These fluctuations cause the alarm to oscillate between an active state and a cleared state in a short period. During this process, false alarm clearance occurs

due to unstable underlying status. In this case, if each alarm or alarm clearance is reported, the alarm notifications may flood, which affects user experience.

The alarm suppression function suppresses flapping alarms. In the suppression period, alarms of the same source are not reported repeatedly, which improves the accuracy of the alarm notification.

If important alarms may be suppressed in a scenario, you can disable the alarm suppression function to obtain complete information about all alarms.

1.4.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable the alarm suppression function.

alarm jitter suppress disable

The alarm suppression function is enabled by default.

1.5 Monitoring

Run the **display** command to display the configuration.

Run the **cleanup** command to clear information.

⚠ Note

Running the **cleanup** command may lose vital information and thus interrupt services.

Table 1-1 Alarm Monitoring and Maintenance

Command	Purpose
cleanup alarm { **active** | **history** } [**sequence-number** *alarm-number*]	Clears alarm information.
cleanup alarm statistic	Clears all alarm statistics.
cleanup event { **all** | **sequence-number** *event-number* }	Clears event information.
display alarm active [**alarm-name** *alarm-name* | **slot** *slot-id* | **type** *alarm-type*] [**all**]	Displays the information about active alarms.
display alarm config	Displays the global configuration of the alarm module.
display alarm config alarm-name *alarm-name*	Displays the configuration of a specific alarm.
display alarm config list	Displays the information about all alarms.

Command	Purpose
display alarm history [**alarm-name** *alarm-name* | **slot** *slot-id* | **type** *alarm-type*]	Displays the information about historical alarms.
display alarm mask [**alarm-name** *alarm-name* | **slot** *slot-id* | **type** *alarm-type*] [**all**]	Displays the information about suppressed alarms.
display alarm statistic [**alarm-name** *alarm-name* **slot** *slot-id*]	Displays statistics of the alarm module.
display event { **all** | **event-name** *event-name* | **slot** *slot-id* | **type** *event-type* }	Displays the information about events.

1.6 Configuration Examples

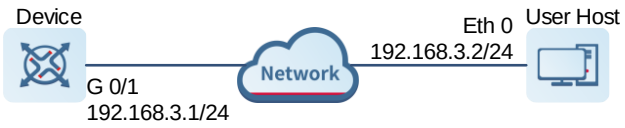
1.6.1 Configuring Alarm Management

1. Requirements

The network administrator can configure the alarm management function to obtain valid information about faults on the device and locate the faults in a timely manner.

2. Topology

Figure 1-1 Topology



3. Note

- Enable the alarm filtering function.
- Enable the alarm suppression function.

4. Procedure

(1) Enable the alarm filtering function and configure the filter condition.

```
Device> privilege
Device# config-system terminal
Device(cfg)# alarm mask alarm-name PortLoopbackAlarm
Device(cfg)# alarm mask slot 0
Device(cfg)# alarm mask type bus
```

(2) Enable the alarm suppression function.

```
Device> privilege
Device# config-system terminal
Device(cfg)# no alarm jitter suppress disable
```

1. Verification

Run the **display alarm config** command to display the global configuration of the alarm module.

```
Device# display alarm config
Alarm state       : ALL READY
Mask flag         : enable
Active count      : 65536
History count     : 2000
Event count       : 0
Mask config set : 0
Active sequence num   : 65536
History sequence num : 65542
Event sequence num    : 0
Old master nodeid     : 0
Slot exist flag       : 0x40000000 (dev1) 0x40000000 (dev2)
Slot online flag      : 0x40000000 (dev1) 0x40000000 (dev2)
Slot mask flag        : 0x40000000 (dev1)
Type mask flag        : 0x8000 0x0 0x0 0x0
Jitter suppress flag : disable
```

2. Configuration Files

```
!
alarm mask alarm-name PortLoopbackAlarm
alarm mask slot 0
alarm mask type bus
!
```

Virtualization Configuration

1. Configuring VSU

Contents

1 Configuring VSU....1
1.1 Introduction....1
1.1.1 Basic Concepts....1
1.1.2 VSU Application....4
1.1.3 VSU Topology....7
1.1.4 DAD....9
1.1.5 Traffic Forwarding....11
1.1.6 System Management....12
1.1.7 System Upgrade....13
1.1.8 Quick Blinking Location....13
1.2 Restrictions and Guidelines....14
1.3 Configuration Task Summary....14
1.4 Setting Up a VSU....14
1.4.1 Overview....14
1.4.2 Restrictions and Guidelines....15
1.4.3 Prerequisites....15
1.4.4 Procedure....15
1.5 Configuring BFD-based DAD....16
1.5.1 Overview....16
1.5.2 Restrictions and Guidelines....16
1.5.3 Prerequisites....16
1.5.4 Procedure....16

1.6 Configuring AP-based DAD 17
1.6.1 Overview 17
1.6.2 Restrictions and Guidelines 17
1.6.3 Prerequisites 17
1.6.4 Procedure 18
1.7 Configuring the VSL 18
1.7.1 Overview 18
1.7.2 Restrictions and Guidelines 19
1.7.3 Prerequisites 19
1.7.4 Procedure 19
1.8 Configuring VSU Attributes 19
1.8.1 Overview 19
1.8.2 Configuration Tasks 20
1.8.3 Configuring Basic VSU Attributes 20
1.8.4 Configuring Traffic Balancing 21
1.8.5 Switching to Standalone Mode 22
1.8.6 Configuring the Restart Mode Upon Link Fault Rectification for the Device in Recovery Mode 22
1.8.7 Configuring Restart-free Auto-Recovery Upon Master Device Malfunction for the Device in Recovery Mode 23
1.9 Configuring Quick Blinking Location 23
1.9.1 Overview 23
1.9.2 Restrictions and Guidelines 23
1.9.3 Procedure 23
1.10 Configuring the Management Interface Creation Mode 24

1.10.1 Overview 24
1.10.2 Restrictions and Guidelines 24
1.10.3 Procedure 24
1.11 Monitoring 24
1.12 Configuration Examples 25
1.12.1 Setting Up a VSU 25
1.12.2 Configuring BFD-based DAD 27
1.12.3 Configuring AP-based DAD 30
1.13 Common Misconfigurations 34

1 Configuring VSU

Note

VSU is supported by only the S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ.

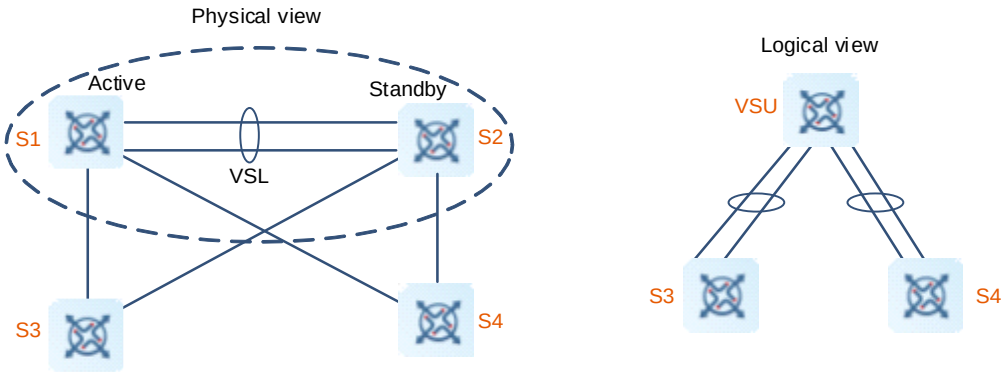
1.1 Introduction

Virtual Switching Unit (VSU) is an N:1 network device technology independently developed by our company. It simplifies the device operation & maintenance (O&M) and network topology by virtualizing multiple network devices into a single logical device for management and use. In addition, the VSU technology connects peripherals to different member devices in the VSU through aggregated links to achieve cross-device link redundancy and improve reliability and scalability of networks.

1.1.1 Basic Concepts

As shown in Figure 1-1, the left part shows the physical view of devices that form a VSU, which is equivalent to one logical device.

Figure 1-1 Schematic Diagram of VSU



1. VSU

The VSU is a single logic entity composed of multiple redundant devices in a conventional network structure. The devices on the access layer, the distribution layer, and the core layer can form the VSU.

2. Domain ID

A domain ID is the unique identifier of a VSU and distinguishes different VSUs. Only two devices that share the same domain ID can compose a VSU.

3. Device ID

Each device in a VSU is called a member device, and each member device has a unique device ID, i.e. switch ID. This device ID is used to manage the member device and configure interfaces on the member device. When

adding the device to a VSU, you need to configure an ID for a device and ensure that the ID is unique in the same VSU. If device IDs of member devices conflict in a VSU, only one device is retained according to rules.

4. Device priority

Priority is an attribute of a member device, and is used to elect a role. A higher priority indicates a higher probability of being elected as the active device. To elect a device as the active device, increase its priority. Member devices have two types of priorities.

- Configuration priority: priority stored in the configuration file, which is changed at any time, and takes effect after the configurations are saved and the VSU is restarted.
- Running priority: configuration priority stored in the configuration file at the startup, which does not change during VSU running.

5. Device role

Member devices are assigned with three roles based on their functions:

- Active device: Manages and controls the entire VSU. Only one active device exists in a VSU domain.
- Standby device: Serves as a backup of the active device and participates in data forwarding only. All received data packets are forwarded to the global active device for processing. When the active device is faulty, the standby device automatically switches to the active mode and takes over functions of the original active device.
- Candidate device: Serves as a backup of the standby device and participates in data forwarding only. When the standby device is faulty, the system automatically elects a new standby device from candidate devices to take over functions of the original standby device. When the active device is faulty and the standby device switches to the active mode, the system automatically elects a new standby device from candidate devices.

6. Role election

A VSU is composed of at least the active device and standby device. When it is composed of more than two devices, the other devices are candidate devices. Role election applies to the following scenarios:

- VSU setup
- Device fault or exit from a VSU
- VSU splitting
- VSU combination

Roles are elected based on the following rules:

- The active device is elected by priority. The device priorities are ranked as follows:

  Current active device (no active device exists upon startup) > Device with a higher priority > Device with a smaller ID > Device with a smaller MAC address

- A device directly connected to the active device is preferentially elected as the standby device to avoid dual active devices. The device priorities are ranked as follows:
  Device directly connected to the active device > Device with a higher priority > Device with a smaller MAC address

Note

- When a device joins an existing VSU in hot mode, the system does not switch between the active and standby roles, even if the device has a higher priority than the active and standby devices in the running VSU.
- The startup sequence of member devices may affect the election of the active device. A member device may not join the VSU in time due to slow startup (the current VSU directly converges if it does not discover a neighbor within 5 minutes). In this case, the member device joins the VSU in hot mode, and the system does not switch active/standby role even if the member device has a higher priority than the active device in the current VSU.
- A candidate device is elected as a standby device only.

7. Virtual switching link

The VSU is a network entity of multiple devices that need to share control information and data flows. Virtual Switching Link (VSL) is a special link used to transmit control information and data flows between the devices in the VSU. Figure 1-1 shows the locations of VSLs in the VSU.

A VSL usually exists in the form of aggregation port group, and a data flow transmitted by the VSL is load-balanced among the aggregation port members based on a traffic balancing algorithm.

- VSL traffic

Control flows transmitted by the VSL between devices are classified into the following two types:

  - Protocol packets received by a member device, which are forwarded through the VSL to the global active device for processing.
  - Protocol packets processed by the global active device, which are forwarded through the VSL to the interfaces of other member devices and then sent to the peer devices by these interfaces.

The data flows transmitted through the VSL between devices include:

  - Flooding data flows in a VLAN
  - Data flows that need to be forwarded across devices and transmitted through the VSL

A VSL also transmits internal management packets of the VSU, for example, protocol information of hot backup devices, and packets that carry configurations to be delivered from the active device to other member devices.

- CRC error detection of VSL port

When a VSL port has many consecutive Cyclic Redundancy Check (CRC) errors, disable this port and switch to another VSL port. CRC errors are handled according to the following rules:

  - The VSL port is checked for CRC errors every 5 seconds by default. If the number of CRC errors incremented between one check and the latest check is greater than the threshold of CRC errors (the *error-number* parameter), one CRC error occurrence is counted.
  - When the number of consecutive CRC errors exceeds the configured number (the *time-number* parameter), the port is abnormal.
  - If there are multiple VSLs and CRC errors occur in only one link, VSL is switched. If there are multiple VSL links and CRC errors occur in the last normal VSL, the link is not switched to prevent topology splitting.
  - The values of *error-number* and *time-number* vary with the scenario. The default values of *error-number* and *time-number* are **3** and **10** respectively. In scenarios with high tolerance for CRC errors, reduce the

two recommended parameter values; otherwise, increase them.

- VSL recovery

By default, a device automatically restarts and joins the VSU again after a VSL fault is rectified. If the automatic restart function is disabled, you must enable this function or directly restart the device after the VSL fault is rectified to rebuild the VSU.

1.1.2 VSU Application

Compared with conventional networks, VSU has the following advantages:

- Low cost: lightweight and dynamic network capacity expansion for protecting the equipment investment in the existing network, and reducing the construction cost.
- Simple O&M: reduced management workload, simplified networking, and lowered O&M difficulty.
- High reliability: device and link redundancy mechanisms for improved network reliability.

1. Flexible expansion

Capacity expansion applies to the following scenarios:

- Port expansion: The port density of the existing access switch cannot meet the requirement of increasing access users, as shown in Figure 1-2.
- Forwarding capacity expansion: The forwarding capacity of the existing core switch cannot meet the requirement of increasing services, as shown in Figure 1-3.
- Bandwidth expansion: The uplink bandwidth cannot meet the demand in peak hours, as shown in Figure 1-4.

The above problems are solved by adding a switch to form a VSU with the original switch. Physically, the two switches seem to be one, and the original switch backs up and sends the current configurations in batches to the new switch. Therefore, this change has little impact on network planning and configuration. The VSU features the following:

- Higher port density: The VSU forwards data as a whole, and the new switch is equivalent to a port expansion card of the original switch to increase port density.
- Higher forwarding capacity: If one switch has a forwarding capacity of 14,400 Mpps, the entire VSU has a forwarding capacity of 28800 Mpps after one switch is added.
- Larger bandwidth: If a 1000 M switch needs to perform 10 G uplink connection but provides only two 10G ports, up to 20 Gbps uplink bandwidth is supported. By adding one switch of the same type to set up a VSU, you can configure four 10 G ports into one aggregation group to increase the uplink bandwidth to 40 Gbps.

Figure 1-2 Port Expansion over VSU

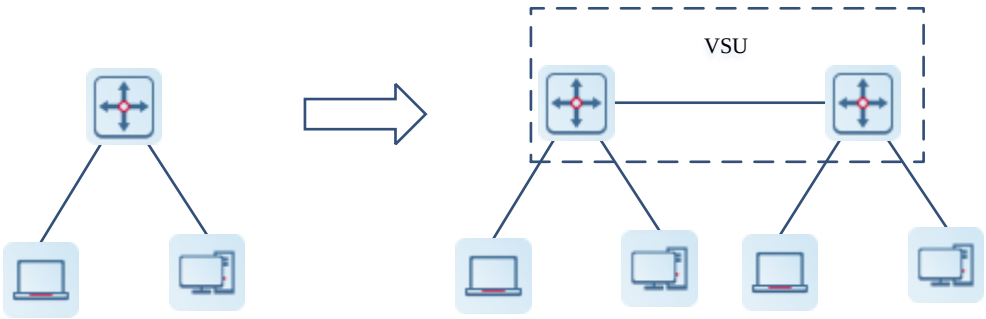


Figure 1-3 Forwarding Capacity Expansion over VSU

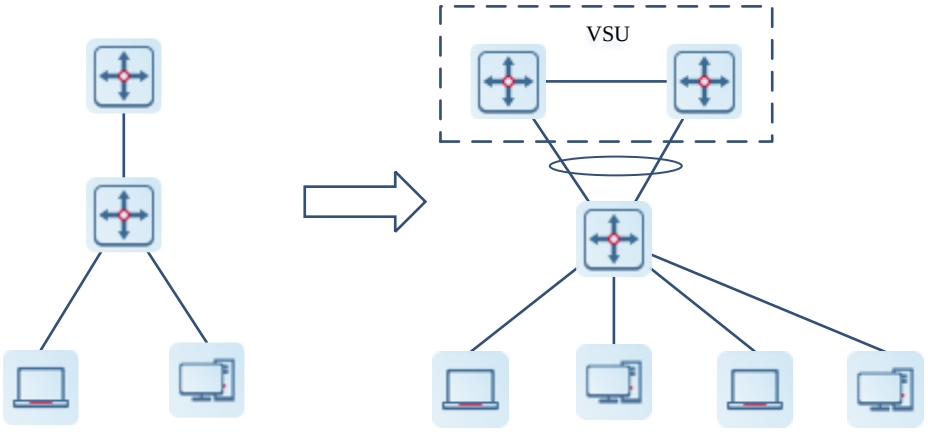
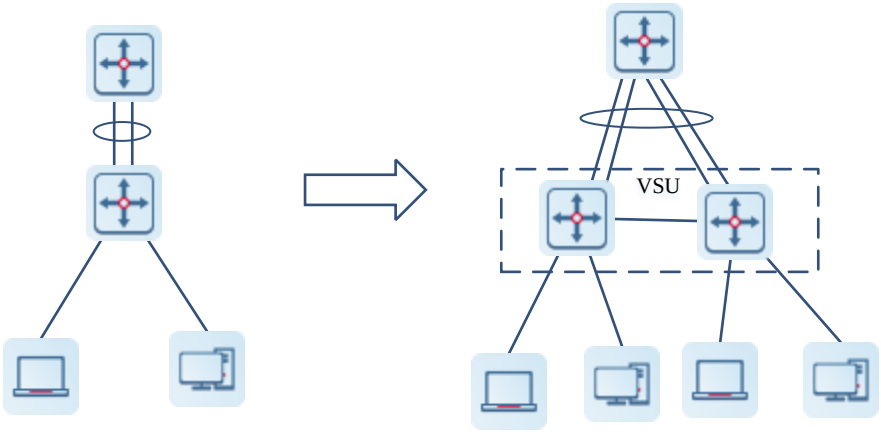


Figure 1-4 Bandwidth Expansion over VSU



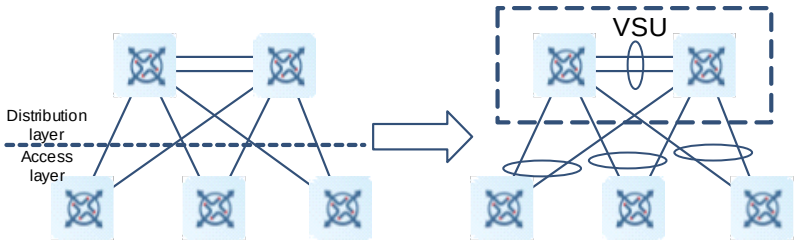
2. Simplified management

Simplified management is reflected in the following aspects:

- Device: After multiple devices form a VSU, an administrator can manage them together, without connecting to them for separate configuration and management. If a VSU is deployed between two floors/buildings, only one distribution device is needed between the floors/buildings, and this reduces the management and O&M costs.
- Networking: Figure 1-5 shows a common networking mode. Typically, each access switch is connected to two distribution switches to improve network reliability. The entire network topology is complex and has a loop. You also have to enable protocols such as Multiple Spanning Tree Protocol (MSTP) and Virtual Router Redundancy Protocol (VRRP) to eliminate the loop and support gateway backup. To meet the traffic balancing requirement, multi-loop network technologies such as Ethernet ring protection switching (ERPS) also need to be enabled. As a result, the configuration is complex, and the maintenance is difficult and prone to errors. After a VSU is formed, MSTP, VRRP, and ERPS are not required. Only the common link aggregation function can balance traffic, eliminate loops, and provide cross-device link redundancy, thus

simplifying the topology and making deployment easier.

Figure 1-5 Simplified Networking over VSU



3. Hot backup

As shown in Figure 1-6, two switches form a VSU to provide the following redundancy to adapt to high reliability scenarios:

- Device redundancy: redundancy between member devices in the VSU. Failure of the standby device does not affect the entire system, but failure of the active device makes the system switch to the standby device. This switching is similar to the dual-engine switching of a standalone device. Such switching is completed within milliseconds, without interrupting services (TCP flows rely on the retransmission mechanism to ensure service continuity).
- Link redundancy: A switch connects to surrounding devices through aggregated links, which provides cross-device link redundancy, implements load balancing, and fully utilizes all bandwidth. If one member link fails, the system switches to another member link within milliseconds.

Figure 1-6 Hot Backup over VSU

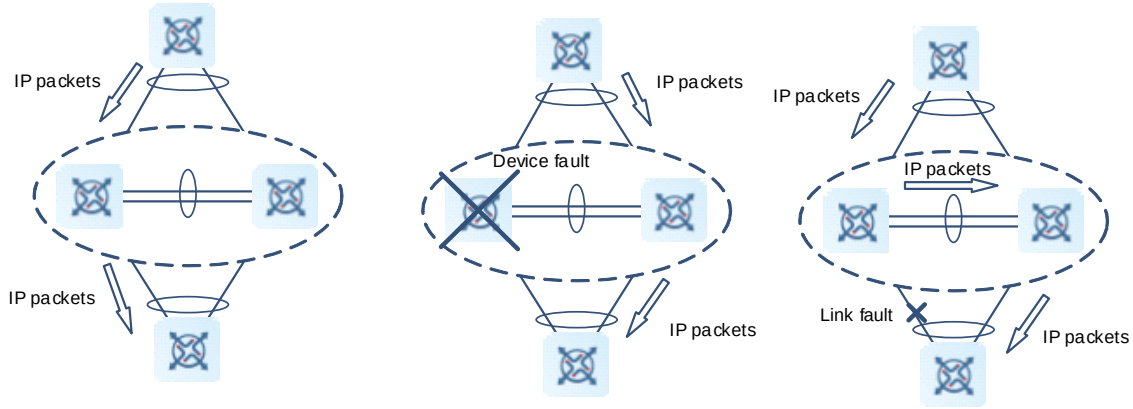
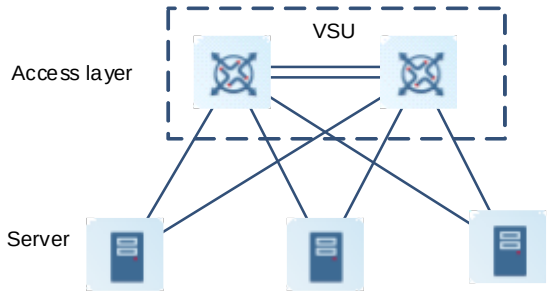


Figure 1-7 shows the typical application scenarios. Access servers with high availability requirements are connected with devices on the access layer by binding multiple network interface cards of a single server into an aggregation port. As the aggregation port allows access to only one access device, the risk of failure of the single device increases. In this case, setting up a VSU enables the server to connect to member devices in the VSU over the aggregation port. This deployment prevents network interruption caused by the single point of failure of the access device or the failure of a single link, and improves network reliability of services carried by the server.

Figure 1-7 Server Access Link Redundancy



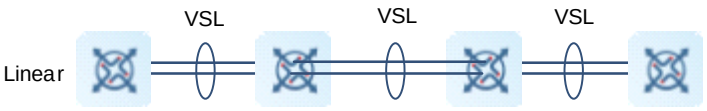
1.1.3 VSU Topology

A VSU supports linear topology and ring topology.

1. Linear topology

As shown in Figure 1-8, devices are connected through a VSL to form a line, and this topology is therefore called linear topology. The linear topology is simple. It uses a few ports and cables. Two devices are connected with a communication link only. Therefore, the VSL is less reliable.

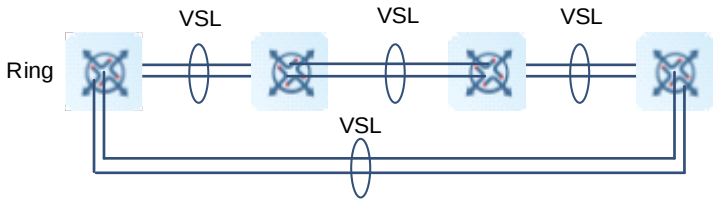
Figure 1-8 Linear Topology



2. Ring topology

In the ring topology shown in Figure 1-9, the two communication links backed up by each other form link redundancy, which improves the reliability of the VSU.

Figure 1-9 Ring Topology



Note

- The ring topology is recommended for the VSU, to ensure that the failure of any single device or any single VSL link does not affect the normal operation of the entire VSU.
- Besides, you are advised to configure multiple VSLs to improve reliability. You are advised to configure at least two VSLs, and to deploy VSLs across line cards for chassis devices

3. Topology convergence

During setup of the VSU, the management scope is determined through topology convergence. The process is as follows:

(1) The member devices discover their neighbors through a topology discovery protocol to determine the list of devices included in the VSU.

(2) The global active device is elected to manage the entire VSU.

(3) The global standby device is elected as the backup for the active device.

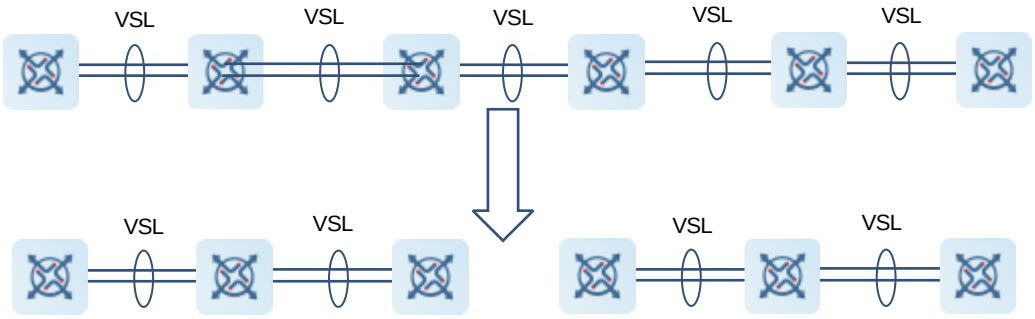
Note

The startup time varies with a device, and the first topology convergence time is different accordingly.

4. Topology splitting

The linear topology is split if the Virtual Switching Link-aggregation port (VSL-AP) link is disconnected. As shown in Figure 1-10, the VSU is split into two groups. In this case, there are two devices with the same configurations on the network, causing the network to malfunction. This topology splitting is resolved by deploying the dual-active detection (DAD) function (see section 1.1.4 DAD).

Figure 1-10 Topology Splitting



5. Topology combination

Topologies are combined when two VSUs with the same VSU domain ID are connected through a VSL-AP link. During topology combination, one of the VSUs automatically restarts in an attempt to join the other VSU, as shown in Figure 1-11.

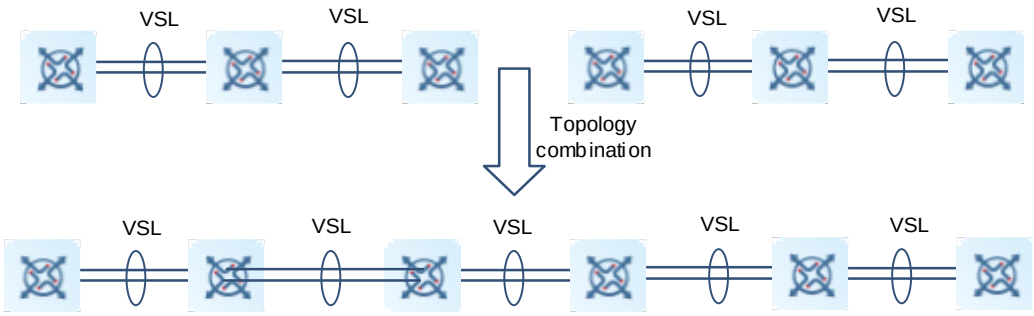
Topologies are combined based on the principle of minimizing the service impact caused by topology combination. The topology combination rules are as follows (topology combination starts from the first rule; if the optimal topology is not selected according to the first rule, the system goes on with the next rule for judgment):

(1) The user configurations are superior to other configurations. A VSU member device with the highest priority prevails.

(2) If the active device cannot be determined according to the preceding rule, a device with a smaller device ID prevails (in a case with two global active devices).

(3) If the active device cannot be determined according to the preceding rule, a device with a smaller MAC address prevails (in a case with two global active devices).

Figure 1-11 Topology Combination



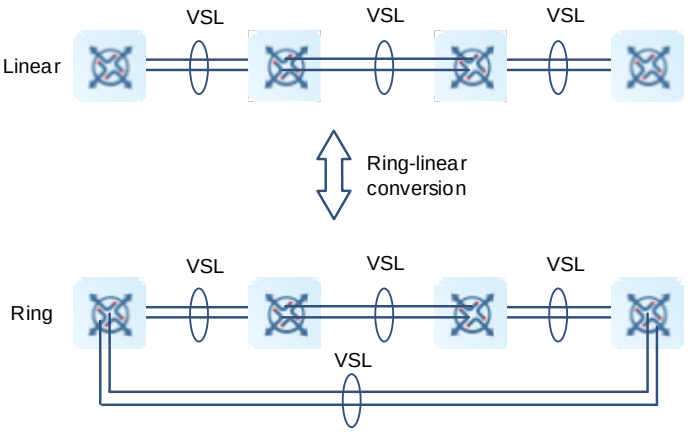
Note

During topology combination of two VSUs, they must be elected. The VSU that fails the election automatically restarts and joins the other VSU in hot mode.

6. Topology conversion

As shown in Figure 1-12, when one VSL-AP link is disconnected, the ring topology is converted into a linear one. The whole VSU can still run normally without network disconnection. However, to avoid the failure of other VSL-AP links or nodes, troubleshoot the VSL fault in time for VSL recovery. After the VSL-AP link is recovered, the linear topology is converted back into a ring one.

Figure 1-12 Conversion Between the Ring Topology and Linear Topology



1.1.4 DAD

When the VSL is disconnected, the standby device becomes the active device. If the original active device is still running, two active devices are main roles and an IP address conflict and other problems occur in the LAN due to their same configurations. In this case, the VSU must detect dual device devices, and take recovery measures. The VSU supports DAD in two ways:

- Bidirectional forwarding detection-based (BFD-based) DAD
- Aggregate interface-based DAD

1. Detection rules

Determine the desired active device according to the following rules one by one:

(1) The healthier device prevails. (A greater sum of bandwidths of up physical ports excluding the administration port and VSL port indicates a healthier device).

(2) The global active device with a higher priority prevails.

(3) The global active device in the VSU with more physical devices prevails.

(4) The global active device with the smaller ID prevails.

(5) The global active device with the smaller MAC address prevails.

(6) The global active device with the longer startup time prevails.

Caution

If DAD is not configured, the network is interrupted after the topology is split.

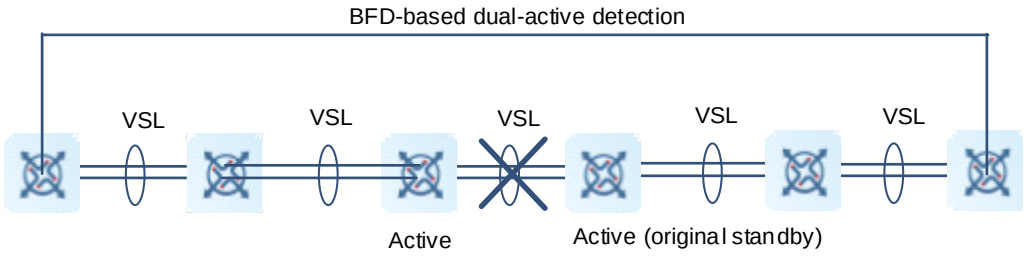
2. Keepalive Timer

To detect the link status of dual hosts, you can configure the Keepalive Timer. When it is detected that the link is deactivated and exceeds the keepalive interval, the dual host detection protocol will become invalid and dual host detection cannot be performed.

3. BFD-based DAD

The VSU supports the BFD-based DAD function. Figure 1-13 shows the connection topology. A dedicated link for DAD is added between two edge devices. When the VSL between the global active and standby devices is disconnected, two active devices coexist. If the BFD-based DAD function is configured, the two active devices send BFD-based DAD packets to each other through the BFD link, and find two identical active devices. Finally, the VSU including one of the active devices is shut down based on rules (see "Topology combination" for details) to enter the Recovery mode to avoid network exceptions.

Figure 1-13 BFD-based DAD



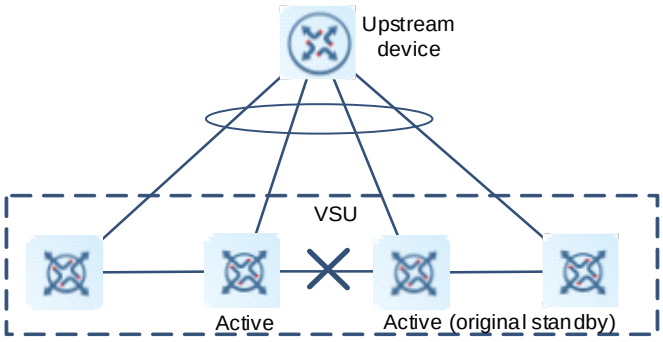
Caution

- When there is only one pair of BFD links, you are advised to deploy them at both ends of the topology.
- Extended BFD is used for the BFD-based detection, and the DAD ports cannot be configured by the existing BFD configurations and display commands.

4. AP-based DAD

The VSU also supports the aggregate interface-based DAD mechanism. Figure 1-14 shows the connection topology. The VSU and the upstream device must support the aggregate interface-based DAD function. When the VSL port is disconnected, dual active devices coexist and send detection packets to each AP member port. The upstream device forwards the detection packets from one active device to the other. The AP is composed of four member ports, which are connected to different devices in the VSU. When the topology is split, the four member ports all send and receive detection packets, and find two identical active devices. Finally, the VSU including one of the active devices is shut down based on rules (see "Topology combination" for details) to enter the Recovery mode to avoid network exceptions.

Figure 1-14 AP-based DAD



Specification

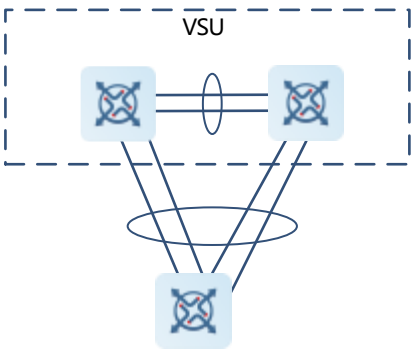
In the preceding topology, the upstream device must be a our company device and can forward detection packets.

1.1.5 Traffic Forwarding

1. Cross-device Aggregation

An AP binds multiple physical links to form a logical link. The VSU supports an AP across member devices. As shown in Figure 1-15, two devices form a VSU and switch A is connected to the VSU as an AP. For switch A, the AP connections shown in the figure are equivalent to those of a common AP group.

Figure 1-15 Cross-device AP



A peripheral device is physically connected to each device in the VSU when a cross-device AP is configured. The advantages are as follows:

- Reserved bandwidth for the VSL: For cross-device AP traffic, the AP member of the same device is preferentially selected as the traffic egress, to avoid unnecessary traffic from being transmitted over the VSL.
- Improved network reliability: If a chassis fails, a member port on a normal device still works properly.

Table 1-1 lists the possible cross-device AP faults and their impact:

Table 1-1 Cross-device AP Faults and Principle Description

Fault Scenario		Principle Description
Link fault	A single link of the AP is faulty.	● The cross-device AP redistributes traffic among the remaining normal links. ● A VSL is needed to forward traffic whose egress port is a port on the device with the faulty link.
	All links of the AP are faulty.	● The AP state changes to link down just as a common AP is disconnected.
Device fault	The active device is faulty.	● The original standby device is switched to the active device. ● Member ports on other member devices continue to work properly ● The cross-device AP redistributes traffic among the remaining normal links.
	The standby/candidate device is faulty.	● The AP member ink connected to the member device is disconnected but other member links still work properly. ● The cross-device AP redistributes traffic among the remaining normal links.

2. Traffic balancing

In a VSU, traffic may have multiple egresses. The AP and Equal-cost Multi-path Routing (ECMP) have respective traffic balancing algorithms, for example, traffic is balanced based on the source and destination MAC addresses. For details, see the description about the link AP module in the interface configuration guide. In this configuration guide, packets received by the local device are preferentially forwarded by the local device. In this way, packets are forwarded to other devices not through a VSL.

1.1.6 System Management

1. Console access

The console of the VSU active device manages multiple devices in the system. The consoles of the standby and candidate devices do not support command line input. However, you can configure VSU-related

commands on the active device for a specified member device, or log in to the console of the active device through the serial port of the standby device. You can also redirect to the master supervisor module of a device by running the **session** command.

2. Line card naming

For chassis-type devices in VSU mode, device ID (switch ID) is added to the ID (name) of a line card. That is, the line card ID changes from 1-dimensional to 2-dimensional. For example, line card 1/1 represents the line card in slot 1 on a member device with the ID of 1.

3. Interface naming

In VSU mode, the same slot ID may appear on multiple devices. Therefore, device ID (switch ID) is added to an interface name. For example, GigabitEthernet 1/1/1 indicates GE port 1 in slot 0 of a device with the ID of 1; GigabitEthernet 2/1/2 indicates GE port 2 in slot 0 of a device with the ID of 2.

4. Access to the file system

In VSU mode, you can access to the file systems on other member devices from the active device. The detailed access method is the same as that of the local file system. The only difference lies in URL prefixes.

5. Log management

All member devices of the VSU can display syslogs. Syslogs generated by the active device are displayed on the console of the active device, and their format is identical to that of syslogs displayed in standalone mode. Syslogs of other member devices are also displayed on the console of the active device, but their format is different from that of syslogs in standalone mode, because the device ID is added to the syslogs. For example, a syslog generated in standalone mode is "%VSU-5-DTM_TOPO_CVG: Node discovery done. Topology converged." Accordingly, a syslog generated by a member device with the device ID of 3 is "%VSU-5-DTM_TOPO_CVG:(3) Node discovery done. Topology converged."

1.1.7 System Upgrade

Usually, the main program versions of member devices in the VSU need to be consistent. But if such numerous member devices are upgraded one by one in standalone mode, this consumes much time and effort, and easily causes errors. The VSU offers a comprehensive system upgrade solution below:

- Check the main program versions of all member devices to determine whether the versions are consistent.
- If inconsistency is identified, the VSU synchronizes the main program of the active device to all member devices via Trivial File Transfer Protocol.

1.1.8 Quick Blinking Location

In the network cabling environment, the equipment room and the operating console are often in different places. If there are many devices in the environment, network administrators are hard to locate specific devices. Quick blinking location provides network administrators with a method for locating devices by quick blinking. By enabling this function for a device on the console, you easily find the corresponding device in the equipment room.

Note

A state of the original Status indicator is not displayed when quick blinking location is enabled, and is displayed only when it is disabled.

1.2 Restrictions and Guidelines

- A VSU can be set up by different products of the same series, but not by products of different series.
- A VSU cannot be set up from chassis-type devices with supervisor modules of different series.
- Some line cards are incompatible with each other. For example, a VSU cannot be set up by devices with CB line cards and those with ED/DB line cards.
- Most devices must use 10G ports as the VSL ports, depending on product specifications.
- When the Switched Port Analyzer (SPAN) function is configured, a VSL port is used as neither the source port nor the destination port of SPAN.

1.3 Configuration Task Summary

The VSU configuration includes the following tasks:

(1) Setting Up a VSU

(2) Configuring BFD-based DAD

(3) (Optional) Configure the host detection link keepalive interval.

dual-active keepalive *interval*

By default, the dual-active link detection keepalive interval is 30 seconds.

(4) Configuring Aggregate Interface-based DAD

(5) Configuring the VSL

(6) (Optional) Configuring VSU Attributes

All the following configuration tasks are optional and may be selected as needed.

- Configuring Basic VSU Attributes
- Configuring Traffic Balancing
- Switching to Standalone Mode
- **Configuring the Restart Mode Upon Link Fault Rectification for the Device in Recovery Mode**
- Configuring Restart-free Auto-Recovery Upon Master Device Malfunction for the Device in Recovery Mode

(7) (Optional) Configuring Quick Blinking Location

(8) Configuring the Management Interface Creation Mode

1.4 Setting Up a VSU

1.4.1 Overview

Devices start up in standalone mode by default. You need to configure the same domain ID on all devices used to set up a VSU, and assign a unique virtual machine ID for each device.

1.4.2 Restrictions and Guidelines

- If the current device works in VSU mode and cannot directly switch to another VSU domain, first switch to standalone mode and then to another VSU.
- The VSU configurations are applicable to a single physical device, and the configurations are stored in the special configuration file **config_vsu.dat**. Therefore, the **display current-config** command does not display the VSU configurations. You can display the current VSU configurations by running the **display switch virtual config** command only.
- In standalone mode, the VSU running information is blank. When you run the **display switch virtual**, a prompt is displayed, indicating that the device works in standalone mode and there is no VSU running information.

1.4.3 Prerequisites

The VSL between VSU devices is normal.

1.4.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the VSU domain ID and enter the VSL interface configuration mode.

switch virtual domain *domain-id*

The default domain ID is 100.

Only devices with the same domain ID can form a VSU.

(4) Configure the device ID of the VSU.

switch *switch-id*

The default device ID is 1.

(5) (Optional) Configure the device priority.

switch *switch-id* **priority** *priority-number*

The default device priority is 100. A larger value indicates a higher priority.

(6) (Optional) Configure the device alias.

switch *switch-id* **description** *device-name*

No device alias is configured by default.

(7) (Optional) Specify the manner of saving the VSU configuration file of a device.

switch cfg_mode { **normal** | **single** }

normal is selected by default. That is, the VSU configurations are separately saved in **config_vsu.dat**.

(8) Return to the global configuration mode.

exit

(9) Enter the VSL port configuration mode.

vsl-port

(10) Add the member ports to the VSL.

port-member interface *interface-type interface-number*

No member port is configured by default.

The VSL member port is a two-dimensional port in standalone mode. The port type depends on the actual product version.

(11) Return to the privileged EXEC mode.

end

(12) Switch the device from the standalone mode to the VSU mode.

switch convert mode virtual

A device works in standalone mode by default.

1.5 Configuring BFD-based DAD

1.5.1 Overview

The BFD-based DAD is configured to prevent coexistence of two active devices.

1.5.2 Restrictions and Guidelines

- BFD detection ports must be direct physical routed ports in different devices.
- The configured port type is unlimited. DAD links are used to transmit only BFD packets and thus need less traffic. Therefore, you are advised to configure a 1000M or 100M port as the DAD port.
- After a Layer 3 (L3) port configured for DAD is switched to a Layer 2 (L2) port (by running the **switch-port** command on the port), the BFD DAD configuration is automatically cleared.
- You are advised to directly connect only the active and standby devices in BFD mode.
- When the VSUs detect a dual-active conflict and one VSU enters the **Recovery** mode, you can rectify the fault by eliminating the VSL failure rather than by directly resetting the VSU in the **Recovery** mode. Otherwise, a dual-active conflict may be incurred in the network.
- Excluded ports in **Recovery** mode must be routed ports rather than VSL ports. After an excluded port is switched from a Layer-3 port to a Layer-2 port (by running the **switch-port** command on the port), the excluded port configuration associated with this port is automatically cleared.
- When dual active devices are detected, one of them must enter the **Recovery** mode. In **Recovery** mode, all service ports of the device must be disabled. To ensure normal use of certain special ports (for example, management ports for remote login to devices), you can configure excluded ports that are not disabled in Recovery mode.

1.5.3 Prerequisites

A VSU has been set up.

1.5.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the BFD detection port configuration mode.

interface *interface-type interface-number*

(4) Configure the BFD detection port as a Layer 3 port.

unset switch-port

A port is a Layer 2 port by default.

(5) Return to the global configuration mode.

exit

(6) Enter the VSU domain configuration mode.

switch virtual domain *domain-id*

(7) Enable the DAD function and specify the BFD-based detection method.

dual-active detection bfd

The DAD function is disabled by default.

(8) Configure a BFD detection port.

dual-active bfd interface *interface-type interface-number*

No BFD detection port is configured by default.

(9) (Optional) Configure the list of excluded ports for Recovery mode.

dual-active exclude interface *interface-type interface-number*

No excluded port is configured by default.

(10) (Optional) Configure the host detection link keepalive interval.

dual-active keepalive *interval*

By default, the dual-active link detection keepalive interval is 30 seconds.

1.6 Configuring Aggregate Interface-based DAD

1.6.1 Overview

The aggregate Interface-based DAD method is configured to prevent coexistence of two active devices.

1.6.2 Restrictions and Guidelines

- When the VSUs detect a dual-active conflict and one VSU enters the **Recovery** mode, you can rectify the fault by eliminating the VSL failure rather than by directly resetting the VSU in the **Recovery** mode. Otherwise, a dual-active conflict may be incurred in the network.
- Excluded ports for the **Recovery** mode must be routed ports rather than VSL ports. After an excluded port is switched from a Layer 3 port to a Layer 2 port (by running the **switch-port** command on the port), the excluded port configuration associated with this port is automatically cleared.

1.6.3 Prerequisites

A VSU has been set up.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VSU domain configuration mode.

switch virtual domain *domain-id*

(4) Enable the DAD function and specify the aggregate interface-based detection method.

dual-active detection aggregateport

The DAD function is disabled by default.

(5) Configure an aggregate interface-based detection port.

dual-active interface *interface-type interface-number* [**vlan** *vlan-id*]

No aggregate interface-based detection port is configured by default.

When the aggregate interface is a trunk port and the native VLAN is beyond the VLAN range allowed by the aggregate interface-based detection port, configure a detection VLAN for the aggregate interface-based detection port. The configured VLAN must fall in the VLAN range allowed by the trunk port, and be created on the device in advance.

(6) In link aggregation mode of the upstream and downstream devices, enable the function of forwarding aggregate interface-based DAD packets.

dad relay enable

The DAD packets are not forwarded by default.

(7) (Optional) Configure the list of excluded ports for Recovery mode.

dual-active exclude interface *interface-type interface-number*

No excluded port is configured by default.

When dual active devices are detected, one of them must enter the **Recovery** mode. In **Recovery** mode, all service ports of the device must be disabled. To ensure normal use of certain special ports (for example, management ports for remote login to devices), you can configure excluded ports that are not disabled in **Recovery** mode.

(8) (Optional) Configure the host detection link keepalive interval.

dual-active keepalive *interval*

By default, the dual-active link detection keepalive interval is 30 seconds.

1.7 Configuring the VSL

1.7.1 Overview

To modify a VSL member port (switch to or back from a common port, or add or delete a VSL) during setup or running of a VSU, you can log in to the VSU console through a serial port or Telnet.

1.7.2 Restrictions and Guidelines

- In practice, dynamic negotiation is used by the VSL AP to prevent link connection faults. A VSL port pool is first configured, and all VSL ports of the same member device are added to the same AP after negotiation. Ports connected to the same device are in the same AP.
- During VSU running, configured VSL member links take effect immediately. VSL ports need to be configured on all devices.
- Chassis-type devices allow 10G SFP ports or ports with higher bandwidth to act as VSL ports. Box-type devices depend on the actual product version.
- Do not set member ports (four 10G ports) of a 40G port as VSL member ports.
- To prevent a loop from being instantaneously generated when a VSL member port exits the VSL AP, the system automatically sets the member port to the **shutdown** state when you run the configuration command. After the VSL member port exits the VSL AP, you can reconnect the link and run the **unset shutdown** command to enable the port again. When you configure the VSL port, the system runs the **shutdown** command to shut down the port first. If the configuration fails and you need to continue using the port as a common port, you can run the **unset shutdown** command to enable the port again. An added member port ID must be three-dimensional.
- If a port is configured as an NLB reflex port, this port can be switched to a VSL member port only after the NLB reflex port configuration is deleted.
- If the VSU topology is split when a VSL port is switched to a common port, the VSL port must not be deleted. You can disconnect the physical port before deleting the VSL port.

1.7.3 Prerequisites

A VSU has been set up.

1.7.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VSL port configuration mode.

vsl-port

(4) Add the member ports for the VSL.

port-member interface *interface-type interface-number*

No VSL member port is configured by default.

The type of a VSL member port depends on the actual product version.

1.8 Configuring VSU Attributes

1.8.1 Overview

In VSU mode, you can configure enhancements to increase the reliability and usability of the VSU.

1.8.2 Configuration Tasks

The VSU attribute configuration includes the following tasks:

All the following configuration tasks are optional and may be selected as needed.

- Configuring Basic VSU Attributes
- Configuring the VSL
- Configuring Traffic Balancing
- Switching to Standalone Mode
- Configuring the Restart Mode Upon Link Fault Rectification for the Device in Recovery Mode
- Configuring Restart-free Auto-Recovery Upon Master Device Malfunction for the Device in Recovery Mode

1.8.3 Configuring Basic VSU Attributes

1. Overview

During VSU setup or running, you can log in to the console of the active or standby device of the VSU to modify some parameters.

2. Restrictions and Guidelines

All configuration commands take effect only after the device is restarted, except the device alias modification command, which takes effect immediately.

3. Prerequisites

A VSU has been set up.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VSU domain configuration mode.

switch virtual domain *domain-id*

(4) Configure the basic VSU attributes as required.

- Modify the VSU domain ID of a specified device.

  switch *switch-id* **domain** *new-domain-id*

  The default VSU domain ID is 100.

- Modify the ID of the specified device.

  switch *switch-id* **renumber** *new-switch-id* [**force**]

  The default device ID is 1.

- Modify the priority of the specified device.

  switch *switch-id* **priority** *priority-number*

  The default device priority is 100.

- Modify the alias of the specified device.

switch *switch-id* **description** *device-name*

No device alias is configured by default.

- Set the CRC error parameter.

  switch crc errors *error-number* **times** *time-number*

  By default, one CRC error occurrence is recorded if the number of CRC errors incremented between two checks is 3 or more. If 10 consecutive CRC error occurrences is recorded, the port is considered abnormal.

- Specify the manner of saving the VSU configuration file of a device.

  switch cfg_mode { **normal** | **single** }

  normal is selected by default. That is, the VSU configurations are separately saved in **config_vsu.dat**.

1.8.4 Configuring Traffic Balancing

1. Overview

In the VSU, if egresses are distributed on multiple devices, traffic balancing allows forwarding traffic preferentially on the local device.

2. Restrictions and Guidelines

- By default, aggregate interface-based and ECMP-based Local Forward First (LFF) are enabled.
- In VSU mode, the cross-device aggregate interface-based LFF and ECMP-based LFF over Layer 3 ports are enabled by default. To set up a VSU with Layer 3 switches, you are advised to configure the aggregate interface-based load balancing based on IP addresses (**src-ip**, **dst-ip**, and **src-dst-ip**).

3. Prerequisites

A VSU has been set up.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VSU domain configuration mode.

switch virtual domain *domain-id*

(4) Disable the aggregate interface-based LFF.

unset switch virtual aggregateport-lff enable

In VSU mode, the aggregate interface-based LFF is enabled by default.

(5) Disable the ECMP-based LFF.

unset switch virtual ecmp-lff enable

In VSU mode, the ECMP-based LFF is enabled by default.

1.8.5 Switching to Standalone Mode

1. Overview

This function switches devices in a VSU to the standalone mode.

2. Restrictions and Guidelines

You can switch devices in a VSU to the standalone mode by using a saved standalone configuration file, or by clearing the VSU configuration.

3. Prerequisites

A VSU has been set up.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Switch a device to the standalone mode.

switch convert mode standalone [*switch-id*]

A device works in standalone mode by default.

1.8.6 Configuring the Restart Mode Upon Link Fault Rectification for the Device in Recovery Mode

1. Overview

This function disables the automatic restart function for the Recovery mode.

2. Restrictions and Guidelines

If the automatic restart function is disabled, you must enable this function again or manually restart the device to recover the device in recovery mode.

3. Prerequisites

A VSU has been set up.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VSU domain configuration mode.

switch virtual domain *domain-id*

(4) Disable the automatic restart function for the **Recovery** mode.

unset recovery auto-restart enable

By default, the automatic restart function is enabled for the Recovery mode after the link fault is rectified.

1.8.7 Configuring Restart-free Auto-Recovery Upon Master Device Malfunction for the Device in Recovery Mode

1. Overview

After this function is configured, the device in recovery mode automatically recovers and becomes the master device when the master device malfunctions.

2. Restrictions and Guidelines

This command can be run only in VSU mode. The command configuration needs to be saved so that it takes effect immediately.

3. Prerequisites

A VSU has been set up.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VSU domain configuration mode.

switch virtual domain *domain-id*

(4) Configure restart-free auto-recovery upon master device malfunction for the device in recovery mode.

dual-active automatic-recovery enable

By default, the device in recovery mode does not recover to serve as the master device and remains in recovery mode.

1.9 Configuring Quick Blinking Location

1.9.1 Overview

When quick blinking location is enabled, the device status indicator blinks quickly.

1.9.2 Restrictions and Guidelines

- Quick blinking location is automatically disabled 30 minutes after it is started.
- The configuration takes effect immediately and cannot be saved. If you restart the device or perform an active/standby switch, quick blinking location is disabled.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enable/Disable quick blinking location.

led-blink { **enable** | **disable** } [**device** *switch-id*]

Quick blinking location is disabled by default. If no device ID is specified, the command applies to all devices in the VSU domain.

1.10 Configuring the Management Interface Creation Mode

1.10.1 Overview

You can configure the management interface creation mode for devices in a VSU as needed, for example, you can create one global management interface or create one management interface for each member device. One management interface is created for each member device in a VSU by default. When one global management interface is created, all member devices of a VSU share the same management interface configuration. If the link of the global management interface is up, physical management interfaces of member devices have at least one link in the up state. If the link of the global management interface is down, physical management interfaces of all member devices are in the down state.

1.10.2 Restrictions and Guidelines

- When configuring this function, you need to run the **save** command to save the configuration and restart the device to make the configuration take effect.
- This command can be configured and take effect only in VSU mode.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the creation of one global management interface for the VSU.

mgmt_mode unique

The default management interface creation mode of a VSU is the normal mode, indicating that one management interface is separately created for each member device in the VSU.

(4) Return to the privileged EXEC mode.

end

(5) Save the global configuration.

save

(6) Restart devices in the VSU.

reboot

1.11 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-2 VSU Monitoring

Command	Purpose
display switch virtual [**topology** | **config** | **role balance**]	Views the topology, configurations, roles, and forwarding and balancing policies of the running VSU.
display switch virtual dual-active { **bfd** | **aggregateport** | **summary** }	Displays the information about the current DAD.
display switch virtual link [**port**]	Displays the information about the current VSL.
display switch id	Displays the ID of the current device.
display mgmt_mode	Displays the management interface creation mode of devices in a VSU.

1.12 Configuration Examples

1.12.1 Configuring a VSU

1. Requirements

As shown in Figure 1-16, connect Device A to Device B through two 10G cables to form a VSU with basic configurations.

2. Topology

Figure 1-16 Configuring a VSU



3. Notes

- Configure the domain ID, device ID, device priority, device name, and VSL ports for the VSU.
- Configure the VSU switching mode.

4. Procedure

(1) On Device A, set the domain ID to 100, device ID to 1, device priority to 200, VSU device name to Device A, and VSL ports to tenGigabitEthernet 1/1 and tenGigabitEthernet 1/2.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# switch virtual domain 100
DeviceA(cfg-vs-domain)# switch 1
DeviceA(cfg-vs-domain)# switch 1 priority 200
DeviceA(cfg-vs-domain)# switch 1 description DeviceA
DeviceA(cfg-vs-domain)# switch crc errors 10 times 20
DeviceA(cfg-vs-domain))# exit
DeviceA(cfg)# vsl-port
```

```
DeviceA(cfg-vsl-port)# port-member interface tengigabitethernet 1/1
DeviceA(cfg-vsl-port)# port-member interface tengigabitethernet 1/2
DeviceA(cfg)# exit
DeviceA# switch convert mode virtual
```

(2) On Device B, set the domain ID to 100, device ID to 2, device priority to 100, VSU device name to Device B, and VSL ports to tenGigabitEthernet 1/1 and tenGigabitEthernet 1/2.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# switch virtual domain 100
DeviceB(cfg-vs-domain)# switch 2
DeviceB(cfg-vs-domain)# switch 2 priority 100
DeviceB(cfg-vs-domain)# switch 2 description DeviceB
DeviceB(cfg-vs-domain)# switch crc errors 10 times 20
DeviceB(cfg-vs-domain))# exit
DeviceB(cfg)# vsl-port
DeviceB(cfg-vsl-port)# port-member interface Tengigabitethernet 1/1
DeviceB(cfg-vsl-port)# port-member interface Tengigabitethernet 1/2
DeviceB(cfg-vsl-port)# exit
DeviceB# switch convert mode virtual
```

5. Verification

(1) Run the **display switch virtual role** command to verify whether the role of Device A is **Active** and that of Device B is **Standby**. If yes, the VSU is set up; if not, the VSU fails to be set up.

```
DeviceA# display switch virtual role
Switch_id    Domain_id    Priority    Position    Status    Role          Description
1(1)         100(100)     200(200)    LOCAL       OK        ACTIVE        DeviceA
2(2)         100(100)     100(100)    REMOTE      OK        STANDBY       DeviceB
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
switch virtual domain 100
!
switch 1
switch 1 priority 200
switch 1 description DeviceA
switch crc errors 10 times 20
!
port-member interface TenGigabitEthernet 0/1
port-member interface TenGigabitEthernet 0/2
!
switch convert mode virtual
!
end
```

- Device B configuration file

```
hostname DeviceB
!
switch virtual domain 100
!
switch 2
switch 2 priority 100
switch 2 description DeviceB
switch crc errors 10 times 20
!
port-member interface TenGigabitEthernet 0/1
port-member interface TenGigabitEthernet 0/2
!
switch convert mode virtual
!
end
```

7. Common Errors

- Different VSU domain IDs (**domain-id**) are configured.
- The VSL ports are not up.

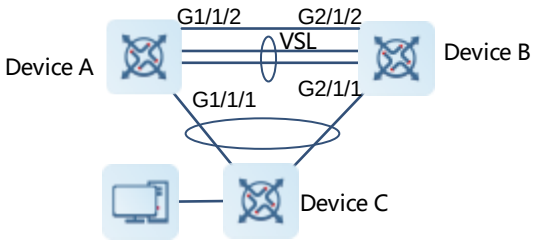
1.12.2 Configuring BFD-based DAD

1. Requirements

As shown in Figure 1-17, Device A and Device B form a VSU. To prevent an IP address conflict in the LAN caused by two identical active devices in case of VSL disconnection, configure the BFD-based detection mechanism to quickly detect two active devices and to actively switch one of them to the Recovery mode (the service forwarding function is disabled).

Topology

Figure 1-17 Configuring BFD-based DAD



2. Notes

- Enable the BFD-based DAD.
- Designate a BFD detection port, which must be configured as a Layer 3 port.

Note

As Device A and Device B form the VSU, the preceding configuration can be performed on either Device A or Device B. Device A is used as an example.

3. Procedure

On Device A, configure GigabitEthernet 1/1/2 and GigabitEthernet 2/1/2 as routed ports, enable the BFD-based DAD, and set the BFD-based detection ports to GigabitEthernet 1/1/2 and GigabitEthernet 2/1/2.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 1/1/2
DeviceA(cfg-if-GigabitEthernet 1/1/2)# unset switch-port
DeviceA(cfg)# interface GigabitEthernet 2/1/2
DeviceA(cfg-if-GigabitEthernet 2/1/2)# unset switch-port
DeviceA(cfg-if)# switch virtual domain 1
DeviceA(c config-vs-domain)# dual-active detection bfd
DeviceA(cfg-vs-domain)# dual-active bfd interface GigabitEthernet 1/1/2
DeviceA(cfg-vs-domain)# dual-active bfd interface GigabitEthernet 2/1/2
```

4. Verification

Disconnect all VSLs, log in to the console of Device A, and run the **display switch virtual** command to check whether **Status** is set to **OK** (Device A runs normally when dual active devices are detected).

```
VSU# display switch virtual
Switch_id       Domain_id       Priority        Position        Status      Role            Description
----------------------------------------------------------------------------------------------
1(1)            1(1)            100(100)        LOCAL           OK          ACTIVE          DeviceA
```

Log in to the console of Device A, and run the **display switch virtual dual-active bfd** command to check whether the BFD enabling status is **Yes** and the status of the BFD-based detection port GigabitEthernet 2/1/2 of Device B is **Down**.

```
VSU-RECOVERY-2# display switch virtual dual-active bfd
BFD dual-active detection enabled: Yes
BFD dual-active interface configured:
  GigabitEthernet 1/1/2: UP
  GigabitEthernet 2/1/2: DOWN
```

Log in to the console of Device B, and run the **display switch virtual** command to check whether the value of **Status** changes to **Recovery** and the device name is suffixed with **-RECOVERY-2**.

```
VSU-RECOVERY-2#display switch virtual
Switch_id       Domain_id       Priority        Position        Status      Role            Description
----------------------------------------------------------------------------------------------
2(2)            1(1)            90(90)          LOCAL           Recovery    ACTIVE          DeviceB
```

Log in to the console of Device B, and run the **display switch virtual dual-active summary** command to check whether the value of **In dual-active recovery mode** is **Yes** (Device B works in recovery mode).

```
VSU-RECOVERY-2# display switch virtual dual-active summary
BFD dual-active detection enabled: Yes
Aggregateport dual-active detection enabled: No
Interfaces excluded from shutdown in recovery mode:
In dual-active recovery mode: Yes
```

5. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 1/1/2
 unset switch-port
!
interface GigabitEthernet 2/1/2
 unset switch-port
!
switch virtual domain 1
dual-active detection bfd
dual-active bfd interface GigabitEthernet 1/1/2
dual-active bfd interface GigabitEthernet 2/1/2
!
switch 1
switch 1 priority 100
switch 1 description DeviceA
switch crc errors 10 times 20
!
port-member interface TenGigabitEthernet 0/1
port-member interface TenGigabitEthernet 0/2
!
switch convert mode virtual
!

end
```

- Device B configuration file

```
hostname DeviceB
!
interface GigabitEthernet 1/1/2
 unset switch-port
!
interface GigabitEthernet 2/1/2
 unset switch-port
!
switch virtual domain 1
dual-active detection bfd
dual-active bfd interface GigabitEthernet 1/1/2
dual-active bfd interface GigabitEthernet 2/1/2
!
switch 2
switch 2 priority 90
switch 2 description DeviceB
switch crc errors 10 times 20
```

```
!
port-member interface TenGigabitEthernet 0/1
port-member interface TenGigabitEthernet 0/2
!
switch convert mode virtual
!

end
```

6. Common Errors

- The BFD-based detection port is not configured as a Layer 3 port.
- Either BFD-based or aggregate interface-based DAD is selected.

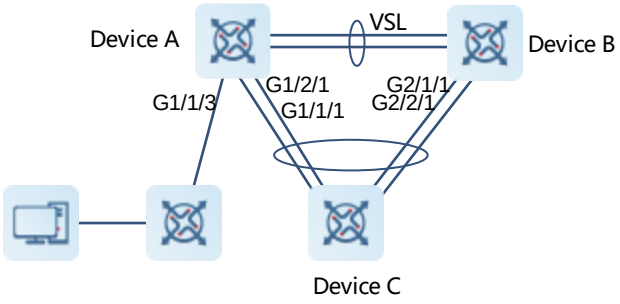
1.12.3 Configuring Aggregate Interface-based DAD

1. Requirements

As shown in Figure 1-18, Device A and Device B form a VSU, and Device C is connected to Device A and Device B separately over an AP. To prevent an IP address conflict in the LAN caused by two identical active devices in case of VSL disconnection, configure the aggregate interface-based detection mechanism to quickly detect two active devices and to actively switch one of them to the Recovery mode (the service forwarding function is disabled).

2. Topology

Figure 1-18 Configuring Aggregate Interface-based DAD



3. Notes

- Configure the ports connecting Device C to Device A and Device C to Device B as the same AP group.
- Enable the aggregate interface-based DAD.
- Configure the aggregate interface-based detection port.
- Configure the ports connecting Device C to Device A and Device B as the same AP group. Enable the function of forwarding aggregate interface-based DAD packets of this port group.

Note

As Device A and Device B form the VSU, the preceding configuration can be performed on either Device A or Device B. Device A is used as an example.

4. Procedure

On Device A, add GigabitEthernet 1/1/1, GigabitEthernet 1/2/1, GigabitEthernet 2/1/1, and GigabitEthernet 2/2/1 to the AP group port-group 1.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface range GigabitEthernet 1/1/1-2
DeviceA(cfg-if-list)# port-group 1
DeviceA(cfg-if-list)# interface range GigabitEthernet 2/1/1-2
DeviceA(cfg-if-list)# port-group 1
```

On Device A, enable the aggregate interface-based DAD, and set the detection port to aggregatePort 1.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# switch virtual domain 1
DeviceA(cfg-vs-domain)# dual-active detection aggregateport
DeviceA(cfg-vs-domain)# dual-active interface aggregatePort 1
```

On Device C, add GigabitEthernet 0/1, GigabitEthernet 0/2, GigabitEthernet 0/3, and GigabitEthernet 0/4 to the AP group port-group 1.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface range GigabitEthernet 0/1-4
DeviceC(cfg-if-list)# port-group 1
```

On Device C, configure the function of forwarding DAD packets.

```
DeviceC> privilege
DeviceC# interface aggregatePort 1
DeviceC(cfg-if-AggregatePort 1)# dad relay enable
```

5. Verification

Run the **display switch virtual dual-active summary** and **display switch virtual dual-active aggregateport** commands to check whether the configurations are correct.

```
DeviceA# display switch virtual dual-active summary
Aggregateport dual-active detection enabled: Yes
Interfaces excluded from shutdown in recovery mode:
In dual-active recovery mode: NO
DeviceA# display switch virtual dual-active aggregateport
Aggregateport dual-active detection enabled: Yes
Aggregateport dual-active interface configured:
  AggregatePort 1: UP
    GigabitEthernet 1/1/1: UP
    GigabitEthernet 1/2/1: UP
    GigabitEthernet 2/1/1: UP
    GigabitEthernet 2/2/1: UP
```

```
DeviceA# display switch virtual dual-active su
```

```
BFD dual-active detection enabled: No
Aggregateport dual-active detection enabled: Yes
Interfaces excluded from shutdown in recovery mode:
In dual-active recovery mode: Yes
DeviceA# display switch virtual
Switch_id       Domain_id       Priority       Position       Status       Role              Description
----------------------------------------------------------------------------------------------
2(2)            100(100)        100(100)       LOCAL          Recovery     ACTIVE            S6500
DeviceA#display switch virtual dual-active aggregateport
Aggregateport dual-active detection enabled: Yes
Aggregateport dual-active interface configured:
  AggregatePort 7: DOWN
    GigabitEthernet 1/0/1: DOWN
    GigabitEthernet 2/0/1: DOWN
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 1/1/1
 port-group 1
!
interface GigabitEthernet 1/1/2
 port-group 1
!
interface GigabitEthernet 2/1/1
 port-group 1
!
interface GigabitEthernet 2/1/2
 port-group 1
!
switch virtual domain 1
dual-active detection aggregateport
dual-active interface aggregatePort 1
!
switch 1
switch 1 priority 100
switch 1 description DeviceA
switch crc errors 10 times 20
!
port-member interface TenGigabitEthernet 0/1
port-member interface TenGigabitEthernet 0/2
!
switch convert mode virtual
!
```

```
end
```

- Device B configuration file

```
hostname DeviceB
!
interface GigabitEthernet 1/1/1
 port-group 1
!
interface GigabitEthernet 1/1/2
 port-group 1
!
interface GigabitEthernet 2/1/1
 port-group 1
!
interface GigabitEthernet 2/1/2
 port-group 1
!
switch virtual domain 1
dual-active detection aggregateport
dual-active interface aggregatePort 1
!
switch 2
switch 2 priority 90
switch 2 description DeviceB
switch crc errors 10 times 20
!
port-member interface TenGigabitEthernet 0/1
port-member interface TenGigabitEthernet 0/2
!
switch convert mode virtual
!

end
```

- Device C configuration file

```
hostname DeviceC
!
interface GigabitEthernet 0/1
 port-group 1
!
interface GigabitEthernet 0/2
 port-group 1
!
interface GigabitEthernet 0/3
 port-group 1
!
interface GigabitEthernet 0/4
```

```
 port-group 1
!
interface AggregatePort 1
 dad relay enable
!

end
```

7. Common Errors

- The AP detection port must be an aggregate interface.
- Either aggregate interface-based or BFD-based DAD is selected.

1.13 Common Misconfigurations

- The VSL port type is not supported.
- Different domain IDs (**domain-id**) are configured.

Interface Configuration

1. Configuring Ethernet Interface
2. Configuring Aggregate Port

Contents

1 Configuring Ethernet Interface 1

1.1 Introduction 1

1.1.1 Interface Classification 1

1.1.2 Interface Numbering Rules 3

1.2 Configuration Task Summary 4

1.3 Configuring Common Features of Ethernet Interfaces 5

1.3.1 Configuration Task Summary 5

1.3.2 Notes 6

1.3.3 Configuring Interfaces in a Range 7

1.3.4 Configuring Interface Description 8

1.3.5 Configuring Layer 3 Ports or Layer 2 Ports 9

1.3.6 Configuring Interface Administrative Status 10

1.3.7 Configuring the Management Status of All Main Interfaces Except Management, VSL, and Filter Interfaces 11

1.3.8 Configuring MTU of an Interface 12

1.3.9 Configuring Interface Bandwidth 13

1.3.10 Configuring Port Rates 15

1.3.11 Configuring the Flow Control Mode 16

1.3.12 Configuring the Duplex Mode of an Interface 17

1.3.13 Configuring the Auto Negotiation Mode 18

1.3.14 Configuring Protected Ports 19

1.3.15 Splitting and Combination of Interfaces 20

1.3.16 Configuring Interface Traffic Statistics 20
1.3.17 Configuring SNMP Features for Interfaces 23
1.3.18 LinkTrap is enabled by reset. Configuring the Syslog Function for Interfaces 23
1.3.19 Configuring Compatibility of Names of Split Subinterfaces with the Names of Common Subinterfaces 24
1.3.20 Configuring Interface Type Change Display for Split Sub-interfaces 24
1.3.21 Configuring the Basic Port Indicator Function 25
1.3.22 Upgrading the PHY Firmware 25
1.3.23 Configuring the Port Module Type 26
1.3.24 Disabling O/E Conversion Module Exception Detection 27
1.3.25 Configuring the Jumbo Frame Length of Traffic Statistics Collection on an Interface 27
1.3.26 Configuring the Delay for Enabling an Interface 28
1.4 Configuring Optical Transceiver Features 28
1.4.1 Configuration Task Summary 28
1.4.2 Notes 29
1.4.3 Configuring Information Collection Intervals of Optical Transceivers 29
1.4.4 Configuring Alarm Detection for Optical Transceivers 30
1.4.5 Configuring Optical Transceiver Antifake Detection 31
1.4.6 Configuring a Center Frequency for an Optical Transceiver 32
1.4.7 Configuring the Function of Triggering an Interface to Go Down in the Case of Optical Transceiver LD 32
1.4.8 Configuring Optical Transceiver Loopback 33
1.4.9 Configuring Transmit Power for Optical Transceivers 34
1.4.10 Configuring Historical Performance Information Collection of Optical Transceivers 35
1.5 Configuring Port Attributes of the Data Link Layer 36

1.5.1 Configuration Task Summary....36
1.5.2 Configuring the MAC Address of an Interface....36
1.5.3 Configuring VLAN Encapsulation Flag on Interfaces....36
1.6 Configuring Port Reliability Attributes....38
1.6.1 Configuration Task Summary....38
1.6.2 Configuring the Carrier-Delay for an Interface....38
1.6.3 Configuring Port Flapping Protection....39
1.6.4 Configuring the FEC Mode of an Interface....40
1.6.5 Configuring Port Errdisable Recovery....41
1.6.6 Configuring CRC Error Frame Detection for an Interface....41
1.6.7 Configuring Bandwidth Usage Alarm Detection on an Interface....42
1.7 Monitoring....43
1.8 Configuration Examples....46
1.8.1 Configuring Interface Attributes....46
1.8.2 Configuring Interconnection Interfaces....49

1 Configuring Ethernet Interface

1.1 Introduction

Interfaces are important in implementing data switching on network devices. The company's devices support two types of interfaces: physical ports and logical interfaces. A physical port is a hardware port on a device, such as the 100M Ethernet interface and Gigabit Ethernet interface. A logical interface, such as the loopback interface and tunnel interface, can be associated with a physical port or independent of any physical port. For network protocols, physical ports and logical interfaces are processed in the same way.

1.1.1 Interface Classification

- Layer 2 port

  Layer 2 port is a physical port on a device and has only Layer 2 switching functions.

- Layer 3 port

  A physical port on a Layer 3 device can be configured as Layer 3 port, which functions as the gateway interface for Layer 3 switching. Layer 3 port cannot implement Layer 2 switching. A Layer 3 port is unrelated with a specific VLAN and only acts as an access port.

- Aggregate interface

  An aggregate interface is a logical link that is formed by binding multiple physical ports together. After the physical ports are bound as an aggregate interface, the bandwidths of the physical ports are added up, which expands the link bandwidth. The interfaces that are bound as an aggregate interface must be of the same type, to be specific, all of them are Layer 3 interfaces or Layer 2 interfaces. Packets that pass the aggregate interface can be distributed to the members of the aggregate interface. If a member link of the aggregate interface becomes invalid, the aggregate interface automatically transfers the traffic on this faulty link to other valid member links, improving reliability of connections.

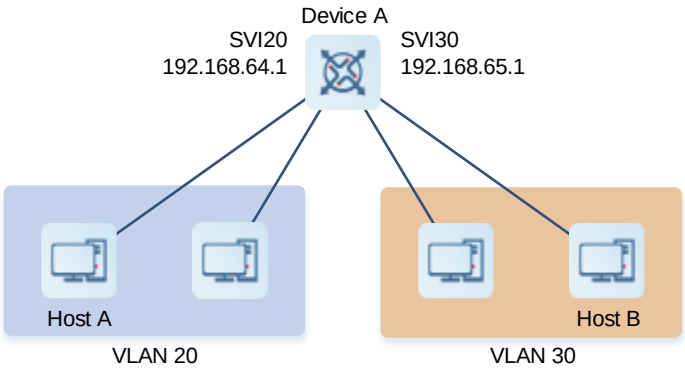
- Switch virtual interface (SVI)

  An SVI acts as a management port of al device, through which the administrator can manage the device. You can also create an SVI as a gateway interface, which is mapped to the virtual interface of a specific virtual local area network (VLAN) to implement routing across VLANs among Layer 3 devices.

  Run the **interface vlan** command to create an SVI and assign an IP address to this interface to set up a route between VLANs.

  As shown in Figure 1-1, the hosts in VLAN 20 can directly communicate with each other without participation of L3 devices. If Host A in VLAN 20 wants to communicate with Host B in VLAN 30, SVI 20 of VLAN 20 and SVI 20 of VLAN 30 must be used.

Figure 1-1 Schematic Diagram of SVI



- Loopback interface

  A loopback interface is a local Layer 3 logical interface simulated by software. The loopback interface is a virtual Ethernet interface that is always in the up state. Packets sent to a loopback interface are processed on the device locally, including route information. The IP address of a loopback interface can be used as the device ID of the Open Shortest Path First (OSPF) routing protocol, the accessed IP address for Telnet server, or the source IP address for Telnet client access.

- Tunnel interface

  A tunnel interface is a virtual interface of a device and implements the tunnel function. Over a tunnel interface, transmission protocols (such as IP) can be used to transmit packets of any protocol. Instead of specifying any transmission protocol or load protocol, a tunnel interface provides a standard point-to-point (P2P) transmission mode. Therefore, a tunnel interface must be configured for every single link.

- Management port

  A management (MGMT) port is a physical port used to manage a device remotely. It is similar to the management method of using a serial port to log in to a device. A management port does not support the communication forwarding function.

  For the internal structure of a device, the management port is separated from the internal forwarding plane of the device. Therefore, the management port does not work for the forwarding plane and the control plane of the device, which means that the communication of the management port is independent from the communication of the services running on the device, thus achieving out-of-band communication. For network communication, a MGMT port equals to any other interface by nature. Therefore, by configuring port attributes, a MGMT port also supports network communication like common ports, but the MGMT port does not support the forwarding function.

  The networks used for management and services can be separated through a MGMT port. Therefore, service network traffic and communication status will not affect device management, thus improving the reliability of device management. Even if the service network fails, the device can be managed through the management network, which is an incomparable advantage over the in-band management method of the service network. In addition, compared with a console port, a MGMT port has a higher bandwidth. In a management network with a log server, a device send logs to the log server through the MGMT port. Therefore, the communication status of the service network will not affect the sending and saving of logs.

1.1.2 Interface Numbering Rules

Table 1-1 shows the numbering rules of physical ports on the devices supporting or not supporting daughter cards in the standalone mode or virtual switch unit (VSU) mode.

Table 1-1 The Numbering Rules of Physical Ports

Daughter Cards	Stand-alone mode	VSU mode
Not Support	A physical port number consists of the slot number and port number on the slot. For example, if the slot number of the port is 2, and the port number on the slot is 3, the port ID is 2/3.	A physical port number consists of the device number, slot number and port number on the slot. For example, if the device number is 1, slot number is 2, and port number on the slot is 3, the port ID is 1/2/3.
Support	A physical port number consists of the slot number, card number and port number on the slot. For example, if the slot number of the port is 2, the card number is 1 and the port number on the slot is 3, the port ID is 2/1/3.	A physical port number consists of the device number, slot number, card number and port number on the slot. For example, if the device number is 1, the slot number of the port is 2, the card number is 1 and the port number on the slot is 3, the port ID is 1/2/1/3.

The numbering rules of the preceding devices, slots, daughter cards, and ports on slots or cards are as follows:

The device ID ranges from 1 to the maximum number of member devices supported.

The slot number rules are as follows: The static slot ID is 0, whereas the ID of a dynamic slot (pluggable module or line card) ranges from 1 to the number of slots. Assume that you are facing the device panel. Dynamic slots are numbered from 1 sequentially from front to rear, from left to right, and from top to bottom.

- Daughter card ID: On a line card supporting daughter cards, the daughter card ID is 0. The daughter card ID of a dynamic slot (pluggable module or line card) ranges from 1 to the number of slots. Dynamic daughter card slots are numbered from 1 sequentially from left to right, and from top to bottom.

The port ID on a slot or a daughter card ranges from 0 to the number of ports on the slot, and is numbered sequentially from left to right.

The ID of a link aggregation ranges from 1 to the number of link aggregations supported by the device.

The ID of an SVI is the VLAN ID corresponding to this SVI.

- The ID of a sub-interface consists of the interface ID followed by “.” and the sub-interface number. The sub-interface number ranges from 1 to the sub-interface configuration range supported by the device.

Note

For the numbering rule of the aggregate interfaces, see the numbering rule of the tunnel interfaces and management ports.

1.2 Configuration Task Summary

The following tasks are optional. Select them as need. Configuring Common Features of Ethernet Interfaces

The following tasks are optional. Select them as need.

- Configuring Interfaces in a Range
- Configuring Interface Description
- Configuring Layer 3 Ports or Layer 2 Ports
- Configuring Interface Administrative Status
- Configuring the Management Status of All Main Interfaces Except Management, VSL, and Filter Interfaces
- Configuring MTU of an Interface
- Configuring Interface Bandwidth
- Configuring Port Rates
- Configuring the Flow Control Mode
- Configuring the Duplex Mode of an Interface
- Configuring the Auto Negotiation Mode
- Configuring Protected Ports
- Splitting and Combination of Interfaces
- Configuring Interface Traffic Statistics
- Configuring SNMP Features for Interfaces
- Configuring the Syslog Function for Interfaces
- Configuring Compatibility of Names of Split Subinterfaces with the Names of Common Subinterfaces
- Configuring Interface Type Change Display for Split Sub-interfaces
- Configuring the Basic Port Indicator Function
- Upgrading the PHY Firmware
- Configuring the Port Module Type
- Disabling O/E Conversion Module Exception Detection
- Configuring the Jumbo Frame Length of Traffic Statistics Collection on an Interface
- Configuring the Delay for Enabling an Interface

- Configuring Optical Transceiver Features

  The following tasks are optional. Select them as need.

  - Configuring Information Collection Intervals of Optical Transceivers
  - Configuring Alarm Detection for Optical Transceivers
  - Configuring Optical Transceiver Antifake Detection
  - Configuring a Center Frequency for an Optical Transceiver
  - Configuring the Function of Triggering an Interface to Go Down in the Case of Optical Transceiver LD
  - Configuring Optical Transceiver Loopback

- Configuring Transmit Power for Optical Transceivers
- Configuring Historical Performance Information Collection of Optical Transceivers

- Configuring Port Attributes of the Data Link Layer

  The following tasks are optional. Select them as need.

  - Configuring the MAC Address of an Interface
  - Configuring VLAN Encapsulation Flag on Interfaces

- Configuring Port Reliability Attributes

  The following tasks are optional. Select them as need.

  - Configuring the Carrier-Delay for an Interface
  - Configuring Port Flapping Protection
  - Configuring the FEC Mode of an Interface
  - Configuring Port Errdisable Recovery
  - Configuring CRC Error Frame Detection for an Interface
  - Configuring Bandwidth Usage Alarm Detection on an Interface

1.3 Configuring Common Features of Ethernet Interfaces

1.3.1 Configuration Task Summary

The following tasks are optional. Select them as required.

- Configuring Interfaces in a Range
- Configuring Interface Description
- Configuring Layer 3 Ports or Layer 2 Ports
- Configuring Interface Administrative Status
- Configuring the Management Status of All Main Interfaces Except Management, VSL, and Filter Interfaces
- Configuring MTU of an Interface
- Configuring Interface Bandwidth
- Configuring Port Rates
- Configuring the Flow Control Mode
- Configuring the Duplex Mode of an Interface
- Configuring the Auto Negotiation Mode
- Configuring Protected Ports
- Splitting and Combination of Interfaces
- Configuring Interface Traffic Statistics
- Configuring SNMP Features for Interfaces
- Configuring the Syslog Function for Interfaces
- Configuring Compatibility of Names of Split Subinterfaces with the Names of Common Subinterfaces
- Configuring Interface Type Change Display for Split Sub-interfaces

- Configuring the Basic Port Indicator Function
- Upgrading the PHY Firmware
- Configuring the Port Module Type
- Disabling O/E Conversion Module Exception Detection
- Configuring the Jumbo Frame Length of Traffic Statistics Collection on an Interface
- Configuring the Delay for Enabling an Interface

1.3.2 Notes

This section describes the precautions for configuring common functions of Ethernet interfaces.

Whether common functions of Ethernet interfaces are supported depends on devices.

Table 1-2 Products Supporting Common Functions of Ethernet Interfaces

Function	Product Supporting the Function
Configuring the Flow Control Mode	S6980-64QC, S6920-4C, S6580-24DC8QC, S6580-48CQ8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ
Configuring the Auto Negotiation Mode	S6980-64QC, S6920-4C, S6580-24DC8QC, S6580-48CQ8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ
Configuring Protected Ports	S6980-64QC, S6920-4C, S6580-24DC8QC, S6580-48CQ8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ
Configuring Interface Type Change Display for Split Sub-interfaces	S6920-4C, S6580-24DC8QC, S6580-48CQ8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ
Upgrading the PHY Firmware	S6510-48VS8CQ, S6250-48XS8CQ
Configuring the Port Module Type	S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ
Disabling O/E Conversion Module Exception Detection	S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ
Configuring the Delay for Enabling an Interface	S6980-64QC, S6920-4C, S6580-24DC8QC, S6580-48CQ8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ

1.3.3 Configuring Interfaces in a Range

1. Overview

For physical ports or logical interfaces created, enter the batch configuration mode of the interfaces within a specified range. For logical interfaces have not been created, the system will create specified logical interfaces in a batch and enter the batch configuration mode of the interfaces within a specified range. When you enter the batch configuration mode to configure multiple interfaces at a time, the attributes configured apply to all interfaces in the selected range.

Run the **interface list** command to specify several range segments, and each range segment can be separated using a comma (,). The types of interfaces within all ranges specified in a command must be the same. The **macro** parameter is used to configure the macro corresponding to a range. Before using the **macro** parameter in the **interface list** command, you must first run the **specify interface-scope** command in the global configuration mode to define these macros. Run the **unset specify interface-scope macro-name** to remove the configuration.

When the **interface list** command is used, the parameter formats vary with different interface types. The common valid interface range formats are as follows:

- **FastEthernet** device/slot/{ start-port } ~ { end-port }
- **GigabitEthernet** device/slot/{ start-port } ~ { end-port }
- **TenGigabitEthernet** device/slot/{ start-port } ~ { end-port }.
- **FortyGigabitEthernet** device/slot/{ start-port } ~ { end-port }
- **AggregatePort** { start-port } - { end-port }. The value range is from 1 to the maximum number of APs supported by the device.
- **vlan** VLAN-start-ID~VLAN-end-ID. The value range of VLAN ID is from 1 to 4,094.
- **Loopback** loopback-start-ID-loopback-end-ID. The value range is from 1 to 2,147,483,647.
- **Tunnel** tunnel-start-ID-tunnel-end-ID. The value range is from 0 to the maximum number of tunnel interfaces supported by the device minus 1.

2. Restrictions and Guidelines

- For logical interfaces, you can run the **unset interface list** command to bacth delete the interfaces in a specified range. But this command cannot batch delete physical ports.
- Run the **reset interface list** command to restore the default configuration of the interfaces within a specified range in the interface configuration mode.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure a macro name for the interfaces to be configured in a batch.

specify interface-scope *macro-name interface-type interface-range-string*

The default macro name is not configured for the interfaces to be configured in a batch.

(4) Enter the batch configuration mode.

interface list { *interface-type interface-range-string* | **macro** *macro-name* }

The default interface range is not configured.

1.3.4 Configuring Interface Description

1. Overview

Configure a name for an interface to identify the interface and help you remember the functions of the interface.

For example, if you want to assign GigabitEthernet 0/1 for exclusive use by user A, describe the interface as "Port for User A."

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 2 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI interface configuration mode.

  interface vlan *interface-number*

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the overlay router interface configuration mode.

  interface overlayrouter *interface-number*

- Enter the virtual PPP interface configuration mode.

  interface virtual-ppp *interface-number*

- Enter the virtual template interface configuration mode.

  interface virtual-template *interface-number*

(4) Configure the description for the interface.

descriptive-name *interface-name*

The default interface name is not configured.

1.3.5 Configuring Layer 3 Ports or Layer 2 Ports

1. Overview

A physical port or a link aggregation on a Layer 3 device can be configured as a Layer 3 port. The Layer 3 port functions as the gateway interface for Layer 3 switching. The Layer 3 port cannot be used for Layer 2 switching. Run the **no switch-port** command to switch a Layer 2 port into a Layer 3 port and assign an IP address to this port to set up a route. Run the **switch-port** command to switch a Layer 3 port into a Layer 2 port.

2. Restrictions and Guidelines

- Running the **unset switch-port** command will delete all Layer 2 features of the interface.
- A Layer 2 link aggregation with member ports cannot be configured as a Layer 3 link aggregation, and vice versa.
- When a port is a member port of a Layer 2 link aggregation or an unauthenticated dot1x port, you cannot change the port type using the **switch-port** command or the **unset switch-port** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

interface aggregateport *interface-number*

(4) Configure the port as a Layer 3 port or a Layer 2 port. Configure at least one of the following tasks.

- Configure the port as a Layer 2 port

  switch-port

- Configure the port as a Layer 3 port.

  unset switch-port

  All ports are Layer 2 ports by default.

1.3.6 Configuring Interface Administrative Status

1. Overview

Enter the interface mode to disable or enable an interface. Two administrative statuses are available for an interface: **Up** and **Down**. The administrative status of an interface is **Down** when the interface is disabled, and the administrative status of an interface is **Up** when the interface is enabled. If you may need to disable an interface. You can directly disable an interface by setting c. If an interface is disabled, it does not receive or send any frames, and the interface loses all its corresponding functions, you can also enable a disabled interface again by setting the management status.

2. Restrictions and Guidelines

The errdisable ports can be recovered using the **shutdown** or **unset shutdown** command. To prevent unwanted link flapping caused by frequently running the **shutdown** or **unset shutdown** command, there should be a certain time interval, which must be greater than the carrier delay of the interface, before you run the **shutdown** or **unset shutdown** command again on an interface.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 2 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregate sub-interface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI interface configuration mode.

  interface vlan *interface-number*

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the overlay router interface configuration mode.

  interface overlayrouter *interface-number*

- Enter the virtual PPP interface configuration mode.

  interface virtual-ppp *interface-number*

- Enter the virtual template interface configuration mode.

  interface virtual-template *interface-number*

(4) Disable the interface.

shutdown

An interface is enabled by default.

1.3.7 Configuring the Management Status of All Main Interfaces Except Management, VSL, and Filter Interfaces

1. Overview

This function is used to disable and enable all main interfaces except management interfaces, VSL Interfaces, and filter interfaces.

2. Restrictions and Guidelines

After the command is executed, the system will perform a shutdown operation on all interfaces except management interfaces, VSL interfaces, and filter interfaces. For interfaces that support sub-interfaces, only the main interface will be shut down, while the sub-interfaces are enabled. When the **display this-config** command is executed in interface configuration mode, the configuration for the shutdown state on a specified interface will be automatically displayed.

After an interface is shut down by this command, you can run the **no shutdown** command in interface configuration mode to enable the interface again, or run the **no shutdown all-interfaces** command in global configuration mode to enable all interfaces that were shut down by the **shutdown all-interfaces** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable all interfaces except management interfaces, VSL interfaces, and filter interfaces.

shutdown whole-interfaces exclude *interface-type interface-range-string* &<1-6>

An interface is enabled by default.

1.3.8 Configuring MTU of an Interface

1. Overview

The MTU is the length of the valid data segment in a frame. It does not include the Ethernet encapsulation overhead. If an interface receives or forwards a frame with a length greater than the MTU, this frame will be discarded.

When a large amount of data is exchanged over an interface, a frame may be greater than a standard Ethernet frame in length. This type of frame is called jumbo frame. If you want a device to receive and handle Jumbo frames, configure the MTU greater than the length of a jumbo frame for the device.

2. Restrictions and Guidelines

- Configuring the system MTU will update the MTU effective values of all the Ethernet interfaces (including the link aggregations) of the system. However, if the interface is configured with an MTU, the MTU configured for the interface will take effect.
- The **mtu forwarding** command works on all the physical ports only and controls only the forwarding of Layer 2 packets. After the command is configured, the protocol MTU is inconsistent with the forwarding MTU, which may cause problems such as flow interruption and protocol exception. For example, in an IPv6 scenario, if the global fwd_mtu configuration is smaller than the default interface value when the default MTU is configured for the interface, the IPv6 packet cannot be sent and received normally. Then the network is interrupted. You are not advised to run the command except for special needs. You are advised to use the **system mtu** command instead of the **mtu forwarding** command.
- You can set an MTU for an interface to control the maximum length of frames that can be received or forwarded by the interface.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an MTU. Configure at least one of the following tasks.

- Configure the system MTU.

  system mtu *mtu-value*

No default system MTU is configured.

- Configure the MTU for a forwarding panel.

  mtu forwarding *mtu-value*

  The default MTU for a forwarding panel is 1500.

a) Enter the interface configuration mode. Perform at least one of the following configuration tasks:

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the SVI interface configuration mode.

  interface vlan *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the overlay router interface configuration mode.

  interface overlayrouter *interface-number*

- Enter the virtual PPP interface configuration mode.

  interface virtual-ppp *interface-number*

b) Configure an MTU for an interface.

mtu *mtu-value*

By default, the MTU of an interface is usually 1500 bytes. The configuration varies with different interface types. The actual product prevail.

The MTU configured for virtual interfaces (including SVIs, tunnel interfaces, and VXLAN routing interfaces) and the MTU configured by the ip mtu command must be smaller than or equal to the MTU configured for the physical interfaces. Otherwise, packets longer than the MTU of the physical interfaces will be discarded.

1.3.9 Configuring Interface Bandwidth

1. Overview

This command can be configured so that some routing protocols (for example, OSPF) can calculate the route metric and the Resource Reservation Protocol (RSVP) can calculate the reserved bandwidth.

2. Restrictions and Guidelines

The **bandwidth** command is a routing parameter, and does not affect the bandwidth of a physical link.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 2 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI interface configuration mode.

  interface vlan *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the overlay router interface configuration mode.

  interface overlayrouter *interface-number*

- Enter the virtual PPP interface configuration mode.

  interface virtual-ppp *interface-number*

- Enter the virtual template interface configuration mode.

  interface virtual-template *interface-number*

(4) Configure the bandwidth for an interface.

bandwidth *bandwidth-value*

No default interface bandwidth is configured.

1.3.10 Configuring Port Rates

1. Overview

Generally, the rate of an Ethernet physical port is determined through auto negotiation with the peer device. The negotiated rate can be any rate within the interface capability. You can also configure any rate within the interface capability for the Ethernet physical port.

2. Restrictions and Guidelines

- If an interface is a member port of a link aggregation, the rate of this interface depends on the rate of the link aggregation. After an interface exits a link aggregation, it will use its own rate.
- The **reset interface** command will not restore the default **port speed**-**mode** configuration of a 25 GE port.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(4) Configure the port rate.

- Configure the rate mode of a non-25 GE interface.

speed [**10** | **100** | **1000** | **10G** | **40G** | **100G** | **200G | 400G** | **auto**]

The interface rate is set to auto by default.

Note

The **10G** keywords is supported by the S6510-4C, or S6510-48VS8CQ only. The **40G** keyword is not supported by the S6980-64QC, and S6580-48CQ8QC, or S6580-24DC8QC only. The **200G** keyword is supported by the S6580-48CQ8QC, or S6580-24DC8QC. The **400G** keyword is supported by the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

- Configure the 25 GE rate mode of the interface.

port speed-mode { **25G** | **10G** }

An interface works at 25 GE by default.

Note

This command is supported by only the S6510-48VS8CQ.

1.3.11 Configuring the Flow Control Mode

Note

Flow control is supported by only the S6980-64QC, S6920-4C, S6580-24DC8QC, S6580-48CQ8QC, S6510-32CQ, S6510-48VS8CQ, and S6250-48XS8CQ.

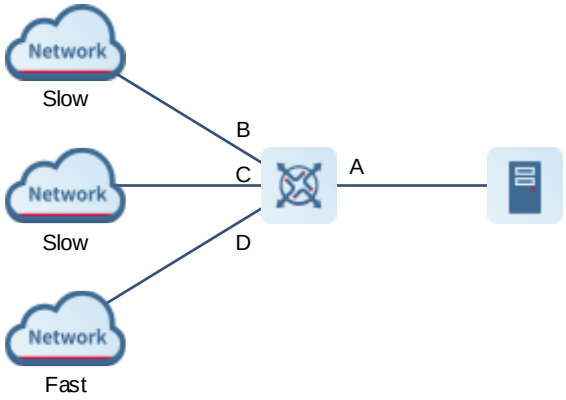
1. Overview

There are two flow control modes of an interface:

- Symmetric flow control mode: Generally, after flow control is enabled on an interface, the interface processes the received flow control frames, and sends the flow control frames when congestion occurs on the interface.
- Asymmetric flow control mode: When an interface works in the asymmetric flow control mode, the interface processes the received flow control frames but does not forward flow control frames.
- The flow control mode of a link aggregation takes effect on all member links.

As shown in Figure 1-2, Port A of the device is an uplink port, and Ports B, C and D are downlink ports. Port B and Port C are connected to different slow networks. Port A is enabled with the functions of sending and receiving flow control frames. As Port B and Port C are connected to different slow networks, if a large amount of data is sent on Port B, Port B and Port C will be congested. Then congestion occurs in the inbound direction of Port A. Therefore, Port A sends flow control frames. When the uplink device responds to the flow control frames, it reduces the data flow sent to Port A, which indirectly slows down the network speed on Port D. In this case, disable the function of sending flow control frames on Port A to ensure the bandwidth usage of the entire network.

Figure 1-2 Interface Flow Control Diagram



2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

(4) Configure the flow control mode of an interface.

flow-control { **auto** | **off** | **on** }

The flow control mode is disabled by default.

1.3.12 Configuring the Duplex Mode of an Interface

1. Overview

The duplex mode of an interface can be configured based on the duplex mode of a peer interface. You can specify one of the following work modes for the interface:

- Full duplex: the interface can receive packets while sending packets.
- Half duplex: the interface can receive or send packets at a time.
- Auto negotiation: the duplex mode of the interface is determined through auto negotiation between the local interface and peer interface.

2. Restrictions and Guidelines

When you configure the duplex mode of a link aggregation, the configuration takes effect on all of its member ports.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(4) Configure the duplex mode.

duplex-mode { **auto** | **full** | **half** }

An interface is in auto-negotiation mode by default.

1.3.13 Configuring the Auto Negotiation Mode

Note

Auto negotiation mode is supported by only the S6980-64QC, S6920-4C, S6580-24DC8QC, S6580-48CQ8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ.

1. Overview

- The auto negotiation mode of an interface can be **on** or **off**. The auto negotiation state of an interface is not completely equivalent to the auto negotiation mode. The auto negotiation state of an interface is jointly determined by the interface rate, duplex mode, flow control mode, and auto negotiation mode.
- When you configure the auto negotiation mode of a link aggregation, the configuration takes effect on all of its member ports. (All these member ports are Ethernet physical ports.)

2. Restrictions and Guidelines

- Generally, if one of the interface rate, duplex mode, and flow control mode is set to **auto**, or the auto negotiation mode of an interface is **on**, the auto negotiation state of the interface is **on**, that is, the auto negotiation function of the interface is enabled. If none of the interface rate, duplex mode, and flow control mode is set to **auto**, and the auto negotiation mode of an interface is **off**, the auto negotiation function of the interface is disabled.
- For a 100 M fiber port, the auto negotiation function is always disabled. For a Gigabit copper port, the auto negotiation function is always enabled, that is, the auto negotiation state is **on**.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

(4) **interface** *ethernet-type interface-number* Configure the auto negotiation mode for the interface.

negotiation mode { **off** | **on** }

The auto negotiation mode of an interface is disabled by default.

1.3.14 Configuring Protected Ports

Note

Protected ports are supported by only the S6980-64QC, S6920-4C, S6580-24DC8QC, S6580-48CQ8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ.

1. Overview

In some application environments, it is required that communication be disabled between some ports on the device. For this purpose, you can configure some ports as protected ports. You can also disable routing between protected ports. After ports are configured as protected ports, the protected ports cannot communicate with each other, but can communicate with non-protected ports. Protected ports work in either of the two modes.

- In the first mode, Layer 2 switching is blocked but routing is allowed between protected ports, which is the default mode.
- In the second mode, Layer 2 switching and routing are both blocked between protected ports.

2. Restrictions and Guidelines

- Two protected ports cannot communicate with each other. But when they are configured as a pair of mirroring ports, frames sent or received by the source port can be mirrored to the destination port.
- Currently, only an Ethernet port or aggregate interface can be configured as a protected port. When a link aggregation is configured as a protected port, all of its member ports are configured as protected ports.
- By default, the Layer 3 routing between protected ports is not blocked. In this case, you can run the **protected-ports route-deny** command to block the routing between protected ports.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Block Layer 3 routing between protected ports.

protected-ports route-deny

Layer 3 routing is not blocked between protected ports by default.

(4) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

(5) Configure a protected port.

switch-port protected

No interface is configured as a protected port by default.

1.3.15 Splitting and Combination of Interfaces

Note

This command is not supported by the S6920-4C.

1. Overview

The 40G/100G Ethernet interface is a high-bandwidth interface. It is mainly used on devices at the convergence layer or core layer to increase the interface bandwidth. 40G/100G interface splitting means splitting one 40G/100G interface into four 10G/25G interfaces. In this case, the 40G/100G interface becomes unavailable, and the four 10G/25G interfaces forward data independently. 40G/100G interface combination means combining four 10G/25G interfaces into one 40G/100G interface. In this case, the four 10G/25G interfaces become unavailable, and only the 40G/100G interface forwards data. Unavailable interfaces cannot forward data. You can flexibly adjust the bandwidth by combining or splitting interfaces.

2. Restrictions and Guidelines

- The S6520-64CQ support splitting a 100GE port into four 25GE ports.
- The S910-3C support splitting a 100GE port into four 25GE ports.
- The S6510-4C support splitting a 100GE port into four 25GE ports.
- The S6250-48XS8CQ and S6510-48VS8CQ support splitting a 100GE port into four 25GE ports.
- The S6510-32CQ supports splitting a 100GE port into four 25GE ports and two 50GE ports.
- The S6580-24DC8QC supports splitting a 200GE port into two 100GE ports.
- The S6980-64QC supports splitting a 400GE port into two 200GE ports. The configuration takes effect without restart.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the splitting and combination of a 40G/100G interface

split interface *interface-type interface-number*

A 40G/100G interface is not split by default.

1.3.16 Configuring Interface Traffic Statistics

1. Overview

By default, virtual interfaces such as sub-interfaces and SVIs do not support interface traffic statistics collection, while the Ethernet port and Ethernet AP support interface traffic statistics collection, but do not support IP traffic statistics collection. When you need to locate a network fault or monitor the network operation status, you can enable the interface traffic statistics collection function and IP traffic statistics collection function. And run he

display interface *interface-type interface-number* command to view the packet statistics of the specified interface and IP packet statistics.

The default statistics collection interval of an Ethernet interface or Ethernet sub-interface is five seconds, that is, interface statistics are updated every five seconds. To meet high requirements for real-time statistics, prolong the collection interval.

The statistics of default sending and receiving rates of a device does not include interframe gaps. The sending and receiving rates of physical layer packets refers to the packet sending and receiving rates including the number of packet bytes and interframe gaps. If you want to check the sending and receiving rates of physical layer packets, enable the device to calculate interframe gaps when it calculates the packet sending and receiving rates. Then the statistics will display the sending and receiving rates of physical layer packets. After the function is enabled, the interframe gap size is fixed at 20 bytes, which consists of 12-byte frame gaps and 8-byte preambles. There is deviation because the device calculates 12-byte frame gaps while the command statistics exclude frame gaps.

2. Restrictions and Guidelines

- Enabling the interface traffic statistics collection function and IP traffic statistics collection function consumes system resources. Enable the functions as required.
- A shorter collection interval consumes more system performance. Therefore, adjust the collection interval as required. You are advised to set the collection interval to more than 10 seconds if there are over 500 physical interfaces.
- After you run the **flow-statistics include-interframe enable** command, the device will clear the statistics of the packet sending and receiving rates of the interface and calculate the rates again.
- After the **statistics** { **enable** | **ip enable** [**brief**] } command is configured, the device clears packet transmitting and receiving rate statistics and recalculates the rates.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the collection interval for Ethernet interface statistics collection.

ethernet-port counter sample-period [*interval*]

The default collection interval of Ethernet interface statistics collection is five seconds.

(4) (Optional) Configure the collection interval for Ethernet sub-interface statistics collection.

ethernet-subport counter route-sample-period [*interval*]

The default collection interval of Ethernet sub-interface statistics collection is five seconds.

(5) Configure that the interface rate statistics include interframe gaps.

flow-statistics include-interframe enable

The interface rate statistics does not include interframe gaps by default.

(6) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 2 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI interface configuration mode.

  interface vlan *interface-number*

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the overlay router interface configuration mode.

  interface overlayrouter *interface-number*

- Enter the VXLAN tunnel interface configuration mode

  interface overlaytunnel *interface-number*

- Enter the virtual PPP interface configuration mode.

  interface virtual-ppp *interface-number*

- Enter the virtual template interface configuration mode.

  interface virtual-template *interface-number*

(7) Enable the traffic statistics collection function on the interface.

statistics { **enable** | **ip enable** [**brief**] }

The IP traffic statistics collection function is disabled on all interfaces by default.

(8) (Optional) Configure the interval of load calculation for the interface.

load-period *interval*

The interval of load calculation for an interface is 10 seconds by default.

1.3.17 Configuring SNMP Features for Interfaces

1. Overview

The Simple Network Management Protocol (SNMP) is used for network monitoring and management. Because many manufactures support SNMP, it becomes a network management standard and is widely used in manufactures' interconnection systems. Through SNMP, a network administrator can check information, configure networks, locate faults and plan capacity for network nodes for network nodes. Our devices support following SNMP features.

- Interface index persistence

  Like the interface name, the interface index also identifies an interface. When an interface is created, the system automatically assigns a unique index to the interface. The index of an interface may change after the device is restarted. Enable the interface index persistence feature to retain the interface index after you restart the device.

- Enhanced name display for interfaces

  After enhanced name display for interfaces is enabled in a standard MIB node, the value of an interface name does not contain spaces. In addition, the ifName node displays the full name of the interface, which does not contain spaces. By default, the value of an interface name in a standard MIB node and the interface name do not contain spaces. In addition, the ifName node displays the short name of the interface.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable interface index persistence for the interface.

snmp-agent if-index persist

The interface index persistence feature is disabled on an interface by default.

(4) Configure enhanced name display for interfaces.

snmp-agent if-name enhance

The function is disabled by default.

After you run this command, save the configuration and restart the device to make the configuration take effect. Then the interface names obtained by SNMP will not contain spaces.

(5) (Optional) Configure the interface to send LinkTrap.

snmp-agent enable linktrap

1.3.18 LinkTrap is enabled by reset. Configuring the Syslog Function for Interfaces

1. Overview

Use the syslog function to determine whether to display information about interface changes or exceptions.

Enable or disable the syslog function as required. By default, this function is enabled. The system displays prompts to alert users when an exception occurs on an interface, for example, the interface status changes, the interface receives error frames, flapping occurs on the interface, an optical transceiver is removed or inserted, or an optical transceiver generates an alarm (or an alarm is cleared).

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the syslog printing function on the interface.

syslog { **error-frame** | **link-dither** | **link-updown** | **res-lack-frame** | **fiber-alarm** | crc-frame | insert-remove }

The syslog printing function is enabled on an interface.

1.3.19 Configuring Compatibility of Names of Split Subinterfaces with the Names of Common Subinterfaces

2. Overview

Configure compatibility of names of split subinterfaces with the names of common subinterfaces so that the NMS and peripheral server management system parse and identify split subinterfaces conveniently.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Configure compatibility of names of split subinterfaces with the names of common subinterfaces.
split compatible mode

The function is disabled by default.

1.3.20 Configuring Interface Type Change Display for Split Sub-interfaces

1. Overview

Configure interface type change display for split sub-interfaces so that the NMS and peripheral server management system parse and identify split sub-interfaces conveniently.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure interface type change display for split sub-interfaces.

split member-type change

By default, this function is disabled.

1.3.21 Configuring the Basic Port Indicator Function

Note

The basic port indicator function is supported by only the S6920-4C, S6580-48CQ8QC, S6580-24DC8QC, S6510-32CQ, S6510-48VS8CQ, and S6250-48XS8CQ.

1. Overview

For ports with the indicator function enabled, their physical LED indicators are steady red, which is convenient for users to distinguish them from other ports when removing or inserting cables.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Enable the port indicator function.

posit-lamp enable

The function is disabled by default.

1.3.22 Upgrading the PHY Firmware

Note

PHY firmware upgrade is supported by only the S6920-4C, S6520-64CQ, S6510-32CQ, S6510-48VS8CQ, and S6250-48XS8CQ.

1. Overview

PHY firmware is the firmware of external PHY chips. The versions of the PHY chip firmware are provided by PHY chip manufacturers. PHY chip manufacturers fix bugs in the firmware of an old version by upgrading the firmware version. So, it is inevitable to upgrade the firmware of PHY chips. Currently, the version of the external PHY firmware is forcibly verified and upgraded when the software version of a switch is upgraded. Upgrading the external firmware takes a long time, and the system startup after software version upgrade is slow. In particular, this method is very inconvenient when the upgrade is performed at the customer site. Therefore, a CLI command for upgrading external PHY firmware is added, and the forcible firmware version verification and upgrade during startup are canceled. You can use the CLI command to upgrade the software and firmware versions as required.

2. Restrictions and Guidelines

- You are advised to isolate services before upgrading the external PHY firmware.
- Do not power off the device during the upgrade. After the upgrade is completed, restart the device to make the new version take effect.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Upgrade the external PHY firmware.

phy firmware upgrade

1.3.23 Configuring the Port Module Type

Note

The port module type configuration is supported by only the S6920-4C, S6510-48VS8CQ, S6520-64CQ, S6510-32CQ, S6510-4C, and S6250-48XS8CQ.

1. Overview

Ports may use different modules and different module types need to be configured for them so that the modules properly send and receive signals and forward traffic.

The port module type does not need to be configured by default. The software automatically adapts to inserted modules, and configures the optical transceiver type for optical transceivers, backlane type for AOC modules, and copper cable type for copper cables.

If a module cannot be identified and "Link" state is not displayed for the port, you can use this function to forcibly set the module type to rectify the fault.

2. Restrictions and Guidelines

- Configure the port module type based on the actually inserted module. If the module is inconsistent with the configuration, a system log is displayed, the quality of port signals deteriorates, a packet cyclic redundancy check (CRC) error or packet loss occurs, and even the port goes down.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure the port module type.

transceiver

1.3.24 Disabling O/E Conversion Module Exception Detection

Note

O/E conversion module exception detection is supported by only the S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ.

1. Overview

This section describes how to disable the link detection function of the 10/100/1000M O/E conversion module.

2. Restrictions and Guidelines

- When a 1G O/E conversion module (supporting 10M/100M/1000M auto-negotiation) is inserted, the "Link" state is displayed only for the interface at one end because the device does not support 1G auto-negotiation. In this case, this function is required for detection.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable the function of detecting exceptions on the 1G O/E conversion module.

gt-transceiver disable

1.3.25 Configuring the Jumbo Frame Length of Traffic Statistics Collection on an Interface

1. Overview

When a large amount of data is exchanged over an interface, a frame may be greater than a standard Ethernet frame in length. This type of frame is called jumbo frame. If the frames are greater than the configured MTU in length, the frames can still be forwarded, but they are not counted as unicast/broadcast/multicast frames. Instead, the frames are classified as jumbo frames.

2. Restrictions and Guidelines

- If the configured value exceeds the MTU, the frames greater than the MTU in length are classified as oversize frames.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure the maximum jumbo frame length.

jumboframe *value*

The maximum jumbo frame length is set to 9,216 by default.

1.3.26 Configuring the Delay for Enabling an Interface

1. Overview

When you reboot the device, replace interface cards, or reset a board, if the interface immediately sends a signal after initialization, data loss may occur due to configurations such as recovery or unsynchronized link switching. In this case, you can configure the delay for enabling the interface to postpone the transmission of signals from the interface.

2. Restrictions and Guidelines

- This function is supported by physical interfaces only and cannot be configured for logical interfaces.
- After you configure the delay for enabling the interface, the configuration will take effect upon next initialization of the interface (interface card reset, interface card power cycle, interface card insertion/removal, board reset, and device reboot).

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the configuration mode of the Layer 2 Ethernet interface.

  interface *ethernet-type interface-number*

- Enter the configuration mode of the Layer 3 Ethernet interface.

  interface *ethernet-type interface-number*

(4) Configure the delay for enabling an interface upon initialization.

link-delay-up on-start *delay-time*

The default delay for enabling the interface is 0 seconds, that is, no delay.

1.4 Configuring Optical Transceiver Features

1.4.1 Configuration Task Summary

The following tasks are optional. Select them as required.

- Configuring Information Collection Intervals of Optical Transceivers
- Configuring Alarm Detection for Optical Transceivers
- Configuring Optical Transceiver Antifake Detection
- Configuring a Center Frequency for an Optical Transceiver
- Configuring the Function of Triggering an Interface to Go Down in the Case of Optical Transceiver LD

- Configuring Optical Transceiver Loopback
- Configuring Transmit Power for Optical Transceivers
- Configuring Historical Performance Information Collection of Optical Transceivers

1.4.2 Notes

This section describes the precautions for configuring optical transceiver functions.

Whether optical transceiver functions are supported depends on devices.

Table 1-3 Products Supporting Optical Transceiver Functions

Function	Product Supporting the Function
Configuring Optical Transceiver Antifake Detection	S6920-4C, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ
Configuring a Center Frequency for an Optical Transceiver	S6980-64QC
Configuring the Function of Triggering an Interface to Go Down in the Case of Optical Transceiver LD	S6980-64QC
Configuring Optical Transceiver Loopback	S6980-64QC, S6580-48CQ8QC, S6580-24DC8QC
Configuring Transmit Power for Optical Transceivers	S6980-64QC, S6580-48CQ8QC, S6580-24DC8QC
Configuring Historical Performance Information Collection of Optical Transceivers	S6980-64QC, S6580-48CQ8QC, S6580-24DC8QC

1.4.3 Configuring Information Collection Intervals of Optical Transceivers

1. Overview

If there is a high requirement for information collection timeliness, you can shorten the information collection interval of optical transceivers, so that the maintenance personnel can obtain information changes of optical transceivers in time.

2. Restrictions and Guidelines

- A shorter collection interval consumes more system performance. Therefore, adjust the collection interval as required.
- Run the **display interfaces transceiver** command to display the information of optical transceivers collected by intervals. Run the **display interface** [*interface-type interface-number*] **transceiver** command

to display the real-time information of the optical transceivers on specified interfaces. Run the **display interfaces transceiver register-dump** command to display register information of optical transceivers.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the information collection interval of optical transceivers on Ethernet interfaces.

ethernet-port fiber-info sample-period [*interval*]

The default information collection interval of optical transceivers on Ethernet interfaces is five minutes.

1.4.4 Configuring Alarm Detection for Optical Transceivers

1. Overview

After this function is configured, you can view the optical transceiver alarm information of an interface by using the management information base (MIB). When events such as insertion or removal of an optical transceiver, optical transceiver exceptions, and optical transceiver exception clearing are detected, related alarm trap messages are sent. After alarm commands are configured, the system reports alarms or alarm clearing information when relevant conditions are met. Exceptions include overhigh/overlow transmit power of optical transceivers, overhigh/overlow receive power of optical transceivers, inter-integrated circuit (IIC) fault of optical transceivers, and unsupported optical transceiver type.

After the alarm detection function is enabled and a corresponding detection period is configured, the SNMP device periodically monitors the optical transceiver status on an interface. When any of the above events is detected, the SNMP device sends a related TRAP message. The SNMP device sends a TRAP message immediately for the optical transceiver insertion or removal, but sends a TRAP message in the next period for optical transceiver exceptions and optical transceiver exception clearing. Only one TRAP message is sent for the same exceptions. The exception will be notified periodically, and the exception clearing event will be notified once only. You can disable the periodic notification function by using the repeated notification mode switch.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the alarm detection function for optical transceivers.

fiber alarm-detect enable [except-shutdown]

The alarm detection function is disabled for optical transceivers by default.

(4) (Optional) Configure the alarm detection period for optical transceivers.

fiber alarm-detect period *minutes*

The default alarm detection period of optical transceiver exceptions is 10 minutes.

(5) (Optional) Configure the periodic repeated notification mode of optical transceiver alarms.

fiber alarm-detect repeat-mode { **on** | **off** }

The repeated notification mode of optical transceiver alarms is enabled by default.

(6) (Optional) Configure the display of optical transceiver alarms.

syslog fiber-alarm

The display function of optical transceiver alarms is enabled by default.

(7) (Optional) Configure enhanced display of optical transceiver alarms.

syslog fiber-enhance

The enhanced display of optical transceiver alarms is disabled by default.

1.4.5 Configuring Optical Transceiver Antifake Detection

Note

Optical transceiver antifake detection is supported by only the S6920-4C, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ.

1. Overview

Configure the optical transceiver antifake detection function to check whether the optical transceiver in use is supplied by our company.

If the optical transceiver is not supplied by our company, the data communication may be affected. If the optical transceiver antifake detection function is enabled, the device can automatically identify an optical transceiver that is not supplied by our company and generate an alarm when such a module is inserted to our company's device.

This function is disabled by default. You can enable this function through configuration.

Each optical transceiver supplied by our company has a unique antifake code. The device can read this antifake code to determine whether the module is supplied by our company. If not, the device will generate syslogs and send trap messages.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure optical transceiver antifake detection.

fiber antifake { **ignore** | **enable** }

The optical transceiver antifake detection function is disabled by default.

1.4.6 Configuring a Center Frequency for an Optical Transceiver

Note

Configuring a center frequency for an optical transceiver is supported by only the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1. Overview

This section describes how to adjust the center frequency of an optical transceiver. You can set different center frequencies for optical transceivers so that the wavelength division multiplexing (WDM) technology is used for signal transmission. For example, two ports on a device transmit signals to ports of the peer device through optical transceivers. You can configure different frequencies for optical transceivers of the ports. After configuration, signals on the two ports are multiplexed by the multiplexer at the transmit end, and then coupled to the same optical fiber for transmission. At the receive end, the demultiplexer demultiplexes and restores the optical carriers of various wavelengths into the two original signals, which are transmitted to the corresponding ports.

2. Restrictions and Guidelines

Specific optical transceivers (such as 400ZR optical transceivers) support this command, which is used to adjust the center frequency of an optical transceiver in WDM scenarios.

When no optical transceiver is inserted into a port, you cannot configure this command nor remove the command configuration. You can run the **reset** form of this command to restore the default configuration. After this command is configured for a port, the configuration is retained if you remove the optical transceiver from the port.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure the center frequency for the optical transceiver.

fiber laser frequency *num*

The default center frequency of an optical transceiver is 193,700,000 MHz.

1.4.7 Configuring the Function of Triggering an Interface to Go Down in the Case of Optical Transceiver LD

Note

Triggering an interface to go down in the case of optical transceiver LD is supported by only the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1. Overview

Specific optical transceivers (such as 400ZR optical transceivers) support this command. When a link is damaged, the bit error rate increases and the link enters the LD state. When the bit error rate reaches a certain threshold, a local fault (LF) is triggered to make the interface go down. You can use this command to configure whether to trigger an LF. No LF is triggered by default.

2. Restrictions and Guidelines

When no optical transceiver is inserted into a port, you cannot configure this command nor remove the command configuration. You can run the **reset** form of this command to restore the default configuration. After this command is configured for a port, the configuration is retained if you remove the optical transceiver from the port.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure the function of triggering an interface to go down in the case of optical transceiver LD.

fiber ld-down enable

The function of triggering an interface to go down in the case of optical transceiver LD is disabled by default.

1.4.8 Configuring Optical Transceiver Loopback

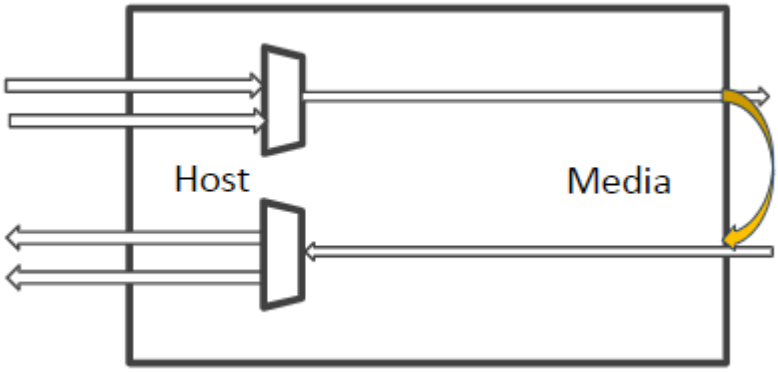
Note

Optical transceiver loopback is supported by only the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1. Overview

Specific optical transceivers (such as 400ZR optical transceivers) support this command. This command can be used to configure loopback detection on the media/host side of an optical transceiver to locate a line fault. After a line fault occurs, you can configure optical transceiver loopback and locate the fault based on the traffic. Suspend service traffic before configuring optical transceiver loopback to ensure that services are not affected.

Figure 1-3 Optical Transceiver Loopback



Media indicates the media (optical fiber) side of an optical transceiver and **Host** indicates the port (device) side of the optical transceiver. When a line fault occurs and communication fails, configure loopback in the outbound direction on the media side, as shown in the preceding figure, and use a tool to generate outgoing traffic. Observe the traffic and locate the fault.

If the traffic loopback is normal, it can be determined that the device and optical transceiver are in good condition. The fault exists on the optical fiber path or peer device. You can configure optical transceiver loopback on the peer device for troubleshooting.

If the traffic loopback is abnormal, it can be determined that the fault exists on the device and optical transceiver. You can configure loopback in the outbound direction on the host side for troubleshooting.

2. Restrictions and Guidelines

Specific optical transceivers (such as 400ZR optical transceivers) support this command. When a line fault occurs, this command can be used to configure loopback detection on the media/host side of an optical transceiver to locate the fault. Suspend service traffic before configuring loopback.

Whether command parameters are supported depends on optical transceivers.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure optical transceiver loopback.

fiber interface *interface-type interface-number* **loopback** { **media | host** } { **input** | **output** | **none** }

No optical transceiver loopback is configured by default.

1.4.9 Configuring Transmit Power for Optical Transceivers

Note

. Configuring transmit power for optical transceivers is supported by only the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1. Overview

This section describes how to adjust the transmit power of optical transceivers. The actual transmit power may differ from the configured transmit power due to environmental factors such as the temperature. You can use this command to adjust the transmit power of optical transceivers as required. Too high or too low optical power affects optical signal transmission. In general, you are not advised to configure this command.

2. Restrictions and Guidelines

Specific optical transceivers (such as 400ZR optical transceivers) support this command. It is used to adjust the transmit power of optical transceivers. In general, you are not advised to configure this command.

If an optical transceiver is not in position, you cannot configure this command nor remove the command configuration on the port. You can run the **reset** form of this command to restore the default configuration. After this command is configured for a port, the configuration is retained if you remove the optical transceiver from the port.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure transmit power for the optical transceiver.

fiber tx-power *num*

The default transmit power of an optical transceiver is –1000, that is, –10.00 dBm.

1.4.10 Configuring Historical Performance Information Collection of Optical Transceivers

1. Overview

After enabling this feature, you can use CLI commands or GRPC to view the historical performance information of optical transceivers on interfaces. This includes media-side bit error rate (BER), host-side BER, dispersion, polarization mode dispersion (PMD).

After historical performance information collection is enabled for optical transceivers and the collection interval is configured, the device will periodically collect the historical performance information for optical transceivers that support this feature on the interfaces. Through CLI commands, the information for up to two collection intervals is displayed.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable historical performance information collection of optical transceivers.

fiber history-performance enable

This feature is disabled by default.

(4) (Optional) Configure the interval of historical performance information collection for optical transceivers.

fiber history-performance period *minutes*

The default interval of historical performance information collection for optical transceivers is 15 minutes.

1.5 Configuring Port Attributes of the Data Link Layer

1.5.1 Configuration Task Summary

The following tasks are optional. Select them as required.

- Configuring the MAC Address of an Interface
- Configuring the Carrier-Delay for an Interface

1.5.2 Configuring the MAC Address of an Interface

1. Overview

By default, each Layer 3 ethernet interface has a globally unique Media Access Control (MAC) address. Our company's device supports modifying the MAC addresses of Ethernet interfaces. However, MAC addresses in the same LAN must be unique.

2. Restrictions and Guidelines

Configuration of MAC addresses may affect internal communication on a LAN. In normal cases, you are advised to modify MAC addresses of Layer 3 Ethernet interfaces.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 3 Ethernet interface configuration mode

interface *ethernet-type interface-number*

(4) Configure the MAC address of an interface.

mac-addr *mac-address*

The system uses factory settings by default.

1.5.3 Configuring VLAN Encapsulation Flag on Interfaces

1. Overview

Virtual local area network (VLAN) is a logical network divided on a physical network and corresponds to the Layer 2 network in the ISO model. In 1999, IEEE released the 802.1Q protocol draft for standardizing the VLAN implementation solution.

The VLAN technology enables the network administrator to divide one physical LAN into multiple broadcast domains (or VLANs). Each VLAN contains a group of workstations with the same requirements and each VLAN has the same attributes as the physical LAN. As VLANs are logically divided, workstations in the same VLAN do not need to be placed in the same physical space, that is, these workstations may belong to different physical LAN network segments. Multicast and unicast traffic in a VLAN will not be forwarded to other VLANs. This helps control traffic, reduces device investment, simplifies network management, and improves the network security.

VLAN is a protocol used to solve Ethernet broadcast and security problems. During packet transmission, a VLAN header is added to Ethernet frames. In addition, VLAN IDs are used to classify users to different work groups to restrict L2 exchange between users in different work groups. Each work group is a VLAN. VLANs can be used to restrict the broadcast scope and form virtual work groups to manage networks dynamically.

To ensure communication with hosts in a VLAN, configure the 802.1Q (VLAN protocol) VLAN encapsulation flag on the Ethernet interface or sub-interface. In this case, when packets are sent over the Ethernet interface, the corresponding VLAN header will be encapsulated. When packets are received, the VLAN header will be deleted from the packet.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an Ethernet sub-interface.

interface *interface-type interface-number[.sub-interface-number]*

No Ethernet sub-interface is configured by default.

(4) Configure the VLAN encapsulation flag for an interface. Configure one of the following tasks.

- Configure the VLAN encapsulation flag for an interface.
  - **subinf-traffic-encapsulation dot1q** { **s-vid** *vlan-id* | **untag** }
  - The Ethernet sub-interface encapsulation protocol is IEEE802.1Q by default, and no VLAN is encapsulated.

Note
This command is not supported by the S6980-64QC.

- Configure a Layer 2 Ethernet sub-interface to encapsulate QinQ.

subinf-traffic-encapsulation qinq s-vid *svlan-id* **c-vid** *cvlan-id*

A Layer 2 Ethernet sub-interface does not encapsulate QinQ by default.

Note
This command is supported by only the S6510-48VS8CQ and S6250-48XS8CQ.

1.6 Configuring Port Reliability Attributes

1.6.1 Configuration Task Summary

The following tasks are optional. Select them as required.

- Configuring the Carrier-Delay for an Interface
- Configuring Port Flapping Protection
- Configuring the FEC Mode of an Interface
- Configuring Port Errdisable Recovery
- Configuring CRC Error Frame Detection for an Interface
- Configuring Bandwidth Usage Alarm Detection on an Interface

1.6.2 Configuring the Carrier-Delay for an Interface

1. Overview

The carrier-delay refers to the delay after which the data carrier detect (DCD) signal changes from **down** to **up** or from **up** to **down**. If the DCD status changes during the delay, the system will ignore this change to avoid negotiation at the upper data link layer. If this parameter is set to a great value, nearly every DCD change is not detected. On the contrary, if the parameter is set to 0, every DCD signal change will be detected, resulting in poor stability.

2. Restrictions and Guidelines

- If the DCD carrier interruption time is long, set a smaller carrier delay so that the routing protocol learns about link state changes quickly, thereby accelerating topology and route convergence.
- If the DCD carrier interruption time is shorter than the network topology or route summarization time, set a larger carrier delay (relative to carrier interruption time) to avoid frequent route flapping.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the SVI interface configuration mode.

  interface vlan *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the virtual PPP interface configuration mode.

  interface virtual-ppp *interface-number*

- Enter the virtual template interface configuration mode

  interface virtual-template *interface-number*

(4) Configure the carrier-delay for an interface.

carrier-delaytime { [**milliseconds**] *delay-interval* | **up** [**milliseconds**] *up-interval* **down** [**milliseconds**] *down-interval* } **[link-effect]**

The default carrier-delay of an interface is two seconds.

1.6.3 Configuring Port Flapping Protection

1. Overview

When flapping occurs on a port, a lot of hardware interruptions occur, consuming a lot of CPU resources. On the other hand, frequent port flapping damages the port. You can configure the flapping protection function to protect ports.

By default, the port flapping protection function is enabled. You can disable this function as required. When flapping occurs on a port, the port detects flapping every 2s or 10s. If flapping occurs six times within 2s on a port, the device displays a prompt. If 10 prompts are displayed continuously, that is, port flapping is detected continuously within 20s, the port is shut down (the violation cause shows **Link Dither**). If flapping occurs 10 times within 10s on a port, the device displays a prompt without shutting down the port.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the port flapping protection function.

physical-port dither protect

The port flapping protection function is enabled by default.

(4) Configure detection parameters for port flapping protection.

physical-port dither period *seconds* **threshold** *number* **period-times** *times*

The default values of **period**, **threshold**, and **period-times** are **2** (in seconds), **6**, and **10** respectively.

(5) Configure detection parameters for port-based port flapping protection.

port-dither period *seconds* [**threshold** *number* **period-times** *times*]

The command is not configured by default and the configuration of the global **physical-port dither** command shall prevail.

1.6.4 Configuring the FEC Mode of an Interface

1. Overview

Forward error correction (FEC) is an error code correction method employing the following working principle: The sender adds a redundancy error-correcting code to the data for sending. The receiver performs error detection on the data based on the error-correcting code. If an error is found, the receiver corrects the error. FEC improves signal quality but also causes signal delay. Users can enable or disable this function according to the actual situation.

There are three FEC modes: rs mode, base-r mode, and auto mode.

2. Restrictions and Guidelines

- When the FEC function is enabled at one end of the link, it must be also enabled at the other end. If the QSFP28-100G-LR4 optical transceiver is used without affecting the negotiation status of the two ends, the FEC function is disabled by default and should not be enabled according to the IEEE standard protocol. If the QSFP28 optical transceivers other than the QSFP28-100G-LR4 are used, the FEC function is enabled by default. In this case, you are not advised to disable this function, which may lead to receiving incorrect packets.
- The FEC function must be enabled for 200G/400G ports, and the none mode cannot be configured to disable FEC. The FEC function can be disabled after the port rate is set to 100 Gbps. If the none mode has been configured after the rate is set to 100 Gbps, the following message will be displayed in the case of rate switching: The speed does not match with the fec mode on interface FHGigabitEthernet 0/1, the fec mode needs to be AUTO or RS.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

(4) Configure the FEC mode.

fec mode { **rs** | **base-r** | **none** | **auto** }

No interface FEC mode is configured by default.

Note

The **base-r**, **none** keywords is not supported by the S6980-64QC. The **base-r** keyword is only supported by the S6250-48XS8CQ.

1.6.5 Configuring Port Errdisable Recovery

1. Overview

Some protocols support the port errdisable recovery function to ensure security and stability of the network. For example, in the port security protocol, when you enable port security and configure the maximum number of security addresses on the port, a port violation event will occur if the number of addresses learned on this port exceeds the maximum number of security addresses. Protocols, such as the Spanning Tree Protocol (STP), 802.1x, Rapid Ethernet Uplink Protection Protocol (REUP) and port flapping protection support the similar functions.

2. Restrictions and Guidelines

- When a port is disabled because it is set to the errdisable state by the REUP link state tracking group function, the port can be restored only by REUP at a scheduled time or by running the REUP errdisable recovery command in the global configuration mode. In other scenarios, run the errdisable recovery command in the global configuration mode to recover all the ports in errdisable state and enable these ports.
- You can manually recover a port, or automatically recover a port at a scheduled time. In addition, you can run the **shutdown** or **unset shutdown** command to recover all the ports in the errdisable state and enable these ports.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enabling the port errdisable recovery function on an interface.

port-violate recovery [**interval** *interval* | **cause** *link-state*]

The port errdisable recovery function is disabled by default.

Note

The **interval** keyword is not supported by the S6980-64QC.

1.6.6 Configuring CRC Error Frame Detection for an Interface

1. Overview

When an interface receives CRC error frames, packet loss or another fault occurs in the link. For an Ethernet interface configured with a backup link, you can configure the function of triggering the interface errdisable state when the interface receives CRC error frames in multiple (configurable) consecutive sampling periods. In this way, services can be switched to the backup link in time, ensuring correct data transmission.

2. Restrictions and Guidelines

- An alarm is generated when an Ethernet interface detects the CRC error frame statistics increase for several consecutive statistics sampling periods, and the increase value is greater than or equal to the CRC error frame threshold. The default alarm mode is to print a system log. After the function of triggering the

errdisable state due to CRC error frames is configured, the interface errdisable state will be triggered. The default number of consecutive periods is **5** (an alarm is generated when the CRC error frame statistics increase of an interface is detected in five consecutive statistics sampling periods), and the default CRC error frame threshold is **1** (that is, an alarm is generated when the CRC error frame statistics increase of an interface in the sampling period is greater than or equal to 1). You can adjust the number of periods and threshold based on the actual environment.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Enable the function of triggering the errdisable state due to CRC error frames on the interface.

port-detect crc-frame port-violate

The function of triggering the errdisable state due to CRC error frames on an interface is disabled by default.

(5) Configure the number of periods for collecting statistics on CRC error frame increase.

port-detect crc-frame period-times

The default number of periods for collecting statistics on CRC error frame increase is **5**, that is, an alarm is generate when the CRC error frame statistics increase of an interface is detected in five consecutive statistics sampling periods. The value range is from 1 to 100.

(6) Configure the CRC error frame threshold.

port-detect crc-frame threshold *num*

The default CRC error frame threshold is **1**, that is, an alarm is generated when the CRC error frame statistics increase of an interface detected within the statistics sampling period is greater than or equal to 1.

1.6.7 Configuring Bandwidth Usage Alarm Detection on an Interface

1. Overview

When the bandwidth resources of an interface are exhausted, the quality of link transmission will be affected. To ensure normal service traffic transmission, enable bandwidth usage alarm detection. Once the alarm threshold and detection interval are set, the system will periodically check bandwidth usage. If the usage exceeds the specified threshold, an alarm message will be generated and the device will send a trap message to the network management system. This enables proactive measures to be taken in advance.

2. Restrictions and Guidelines

- During consecutive detection intervals (using the interface statistics collection interval), if the bandwidth usage exceeds the alarm threshold, the device sends a trap message. If the bandwidth usage falls below the restore threshold, the device also sends a trap message. By default, the maximum number of detection times for triggering an alarm is set to 1 (that is, one statistics collection interval, which is typically 5s). The

default alarm threshold is 90, and the restore threshold is 80 (that is, the bandwidth usage is greater than the alarm threshold and smaller than the restore threshold). The alarm interval and thresholds can be reset based on the actual environment.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Enable the egress and ingress bandwidth usage alarm detection on an interface.

port-detect { **input-rate** | **output-rate** } **enable**

By default, bandwidth usage alarm detection is disabled on an interface.

Configure the maximum number of ingress and egress bandwidth usage detection times for triggering an alarm.

port-detect { **input-rate** | **output-rate** } **period-times** *number*

The default maximum number of ingress and egress bandwidth usage detection times for triggering an alarm is 1. The value range is from 1 to 100.

(5) Configure the alarm threshold and the restore threshold for egress and ingress bandwidth usage alarm detection on an interface.

port-detect { **input-rate** | **output-rate** } **alarm-threshold** *alarm-rate* [**resume-threshold** *resume-rate*]

By default, the alarm threshold for bandwidth usage on an interface is 90 and the restore threshold is 80. Both alarm threshold and restore threshold range from 1 to 100. If the restore threshold is not specified, it is automatically set to the same as the alarm threshold.

1.7 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

Caution

Services may be interrupted due to loss of vital information if you run the **cleanup** command during device operation.

Table 1-4 Ethernet Interface Monitoring

Command	Purpose
display interfaces [*interface-type interface-number*]	Displays all the statuses and configuration information of a specified interface.

Command	Purpose
display interfaces [*interface-type interface-number*] **status**	Displays the interface status.
display interfaces [*interface-type interface-number*] **status detail**	Displays detailed interface status information.
display interfaces [*interface-type interface-number*] **status err-disable**	Displays the interface errdisable status.
display interfaces [*interface-type interface-number*] **link-state-change statistics**	Displays the link status change time and count of a specified interface.
display interfaces [*interface-type interface-number*] **switch-port**	Displays the Layer 2 port status.
display interfaces [*interface-type interface-number*] [**descriptive-name**] [**up | down**]	Displays the configuration and status of an interface.
display interfaces [*interface-type interface-number*] **counters** [**up | down**]	Displays the statistics of a specified port.
display interfaces [*interface-type interface-number*] **counters increment** [**up** | **down**]	Displays the number of packets added in the previous sampling interval.
display interfaces [*interface-type interface-number*] **counters err-num** [**up** | **down**] [**nozero**]	Displays the statistics about error packets.
display interfaces [*interface-type interface-number*] **counters rate** [**up | down**] [**nozero**]	Displays the packet sending/receiving rate of the interface.
display interfaces [*interface-type interface-number*] **counters rate physical-layer** [**up** | **down**] [**nozero**]	Displays the packet sending/receiving rate of an interface at the physical layer.
display interfaces [*interface-type interface-number*] **counters summary** [**up** | **down**] [**nozero**]	Displays a summary of interface packets.
display interfaces [*interface-type interface-number*] **counters drops** [**up | down**]	Displays the statistics of discarded packets of an interface.
display interfaces [*interface-type interface-number*] **line-detect**	Displays the cable status.

Command	Purpose
display interfaces usage [**up | down**] **display interfaces** *interface-type interface-number* **usage**	Displays the bandwidth usage of an interface.
display interfaces [*interface-type interface-number*] **brief** [**up | down**]	Displays brief information of an interface, including the interface status, transmit bandwidth usage, receive bandwidth usage, and numbers of transmitted and received error packets.
display interfaces [interface-type interface-number] **ethernet brief** [**up | down**]	Displays Ethernet information of an interface, including the interface link status, VLAN, auto-negotiation mode, duplex mode, interface rate, bandwidth usage, and description (alias) of the interface.
display interfaces [*interface-type interface-number*] **mtu forwarding**	Displays the information of the forwarding plane MTU.
display vlans	Displays the information of a VLAN sub-interface.
display mgmt virtual	Displays the status and statistics of the member ports of a virtual MGMT port.
display interfaces [*interface-type interface-number*] **transceiver**	Displays the basic information about the optical module on a specified interface.
display interfaces [*interface-type interface-number*] **transceiver alarm**	Displays the current fault alarms of the optical module on a specified interface.
display interfaces [*interface-type interface-number*] **transceiver diagnosis**	Displays the current measurement values of optical module diagnosis parameters of a specified interface.
display interfaces [*interface-type interface-number*] **transceiver state**	Displays information about the optical transceiver status on a specified interface.
display interfaces [*interface-type interface-number*] **transceiver signal-integrity**	Displays information about optical transceiver signal integrity of a specified interface.
display interfaces [*interface-type interface-number*] **transceiver register-dump**	Displays information about the optical transceiver register on a specified interface.
display interfaces [*interface-type interface-number*] **transceiver manuinfo**	Displays the vendor information of the optical transceiver on a specified interface.
display interfaces [*interface-type interface-number*] **transceiver history-performance**	Displays the historical performance information of the optical transceiver.
display split summary	Displays the splitting and combination information of 40G/100G interfaces.

Command	Purpose
cleanup counters [*interface-type interface-number*]	Clears the statistics of a specified interface.
cleanup link-state-change statistics [*interface-type interface-number*]	Clears the statistics of link status changes of an interface.
display interfaces [*interface-type interface-number*] **troubleshooting**	Displays the diagnosis information of an interface.
display interfaces [*interface-type interface-number*] **fault-info**	Displays the fault information of an interface, including the fault cause and occurrence time.

1.8 Configuration Examples

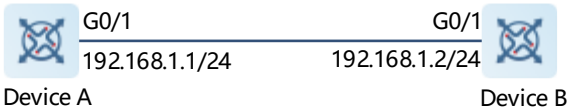
1.8.1 Configuring Interface Attributes

1. Requirements

Interconnect two devices, and configure the basic attributes of the device interfaces.

2. Topology

Figure 1-4 Configuring Interface Attributes



3. Notes

- Connect two devices through the Layer 2 ports.
- Configure an SVI on each of the two devices, and assign IP addresses in the same network segment to the two SVIs.
- Enable interface index persistence on the two devices.
- Enable the link trap function on the two devices.
- Configure the interface administrative status on the two devices.

4. Procedure

Configure as follows on Device A.

```
DeviceA> privilege
DeviceA # config-system terminal
DeviceA(cfg)# snmp-agent if-index persist
DeviceA(cfg)# interface vlan 1
DeviceA(cfg-if-VLAN 1)# ip address 192.168.1.1 255.255.255.0
```

```
DeviceA(cfg-if-VLAN 1)# exit
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# snmp trap link-status
DeviceA(cfg-if-GigabitEthernet 0/1)# shutdown
```

Configure as follows on Device B.

```
DeviceB# config-system terminal
DeviceB(cfg)# snmp-agent if-index persist
DeviceB(cfg)# interface vlan 1
DeviceB(cfg-if-VLAN 1)# ip address 192.168.1.2 255.255.255.0
DeviceB(cfg-if-VLAN 1)# exit
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# snmp trap link-status
DeviceB(cfg-if-GigabitEthernet 0/1)# shutdown
```

5. Verification

Perform verification on Device A and Device B respectively as follows:

- Run the **shutdown** command on the interface GigabitEthernet 0/1, and check whether GigabitEthernet 0/1 and SVI 1 are in **Down** state.
- Run the **shutdown** command on the interface GigabitEthernet 0/1, and check whether a trap message indicating the **Down** state of this interface is sent.
- Restart the device, and check whether the interface index of GigabitEthernet 0/1 is the same as that before the restart.

Device A:

```
DeviceA# display interfaces gigabitEthernet 0/1
Index(dec):1 (hex):1
GigabitEthernet 0/1 is administratively down    , line protocol is DOWN
Hardware is GigabitEthernet, address is 00d0.f865.de9b (bia 00d0.f865.de9b)
Interface address is: unset ip address
  MTU 1500 bytes, BW 1000000 Kbit
  Encapsulation protocol is Bridge, loopback not set
  Keepalive interval is 10 sec , set
  Carrier delay is 2 sec
  Rxload is 1/255, Txload is 1/255
  Queue      Transmitted packets      Transmitted bytes      Dropped packets      Dropped bytes
      0                        0                      0                    0                  0
      1                        0                      0                    0                  0
      2                        0                      0                    0                  0
      3                        0                      0                    0                  0
      4                        0                      0                    0                  0
      5                        0                      0                    0                  0
      6                        0                      0                    0                  0
      7                        4                    440                    0                  0
  Switchport attributes:
    interface's description:""
```

```
    lastchange time:0 Day:20 Hour:15 Minute:22 Second
    Priority is 0
    admin medium-type is Copper, oper medium-type is Copper    admin duplex mode is AUTO, oper duplex is Unknown
    admin speed is AUTO, oper speed is Unknown
    flow control admin status is OFF, flow control oper status is Unknown
    admin negotiation mode is OFF, oper negotiation state is ON
    Storm Control: Broadcast is OFF, Multicast is OFF, Unicast is OFF
  Port-type: access
    Vlan id: 1
   10 seconds input rate 0 bits/sec, 0 packets/sec
   10 seconds output rate 0 bits/sec, 0 packets/sec
    4 packets input, 408 bytes, 0 unset buffer, 0 dropped
    Received 0 broadcasts, 0 runts, 0 giants
    0 input errors, 0 CRC, 0 frame, 0 overrun, 0 abort
    4 packets output, 408 bytes, 0 underruns , 0 dropped
0 output errors, 0 collisions, 0 interface resets
A# display interfaces vlan 1
Index(dec):4097 (hex):1001
VLAN 1 is UP    , line protocol is DOWN
Hardware is   VLAN, address is 00d0.f822.33af (bia 00d0.f822.33af)
Interface address is: 192.168.1.1/24
ARP type: ARPA, ARP Timeout: 3600 seconds
  MTU 1500 bytes, BW 1000000 Kbit
  Encapsulation protocol is Ethernet-II, loopback not set
  Keepalive interval is 10 sec , set
  Carrier delay is 2 sec
  Rxload is 0/255, Txload is 0/255
```

Device B:

```
DeviceB# display interfaces gigabitEthernet 0/1
Index(dec):1 (hex):1
GigabitEthernet 0/1 is administratively down    , line protocol is DOWN
Hardware is GigabitEthernet
Interface address is: unset ip address, address is 00d0.f865.de9b (bia 00d0.f865.de9b)
  MTU 1500 bytes, BW 1000000 Kbit
  Encapsulation protocol is Bridge, loopback not set
  Keepalive interval is 10 sec , set
  Carrier delay is 2 sec
  Rxload is 1/255, Txload is 1/255
  Queue      Transmitted packets      Transmitted bytes      Dropped packets      Dropped bytes
      0                        0                      0                    0                  0
      1                        0                      0                    0                  0
      2                        0                      0                    0                  0
      3                        0                      0                    0                  0
      4                        0                      0                    0                  0
      5                        0                      0                    0                  0
```

```
      6                0                0                0                0
      7                4                440              0                0
  Switchport attributes:
    interface's description:""
    lastchange time:0 Day:20 Hour:15 Minute:22 Second
    Priority is 0
    admin medium-type is Copper, oper medium-type is Copper
    admin duplex mode is AUTO, oper duplex is Unknown
    admin speed is AUTO, oper speed is Unknown
    flow control admin status is OFF, flow control oper status is Unknown
    admin negotiation mode is OFF, oper negotiation state is ON
    Storm Control: Broadcast is OFF, Multicast is OFF, Unicast is OFF
  Port-type: access
    Vlan id: 1
   10 seconds input rate 0 bits/sec, 0 packets/sec
   10 seconds output rate 0 bits/sec, 0 packets/sec
    4 packets input, 408 bytes, 0 unset buffer, 0 dropped
    Received 0 broadcasts, 0 runts, 0 giants
    0 input errors, 0 CRC, 0 frame, 0 overrun, 0 abort
    4 packets output, 408 bytes, 0 underruns , 0 dropped
0 output errors, 0 collisions, 0 interface resets
B# display interfaces vlan 1
Index(dec):4097 (hex):1001
VLAN 1 is UP    , line protocol is DOWN
Hardware is    VLAN, address is 00d0.f822.33af (bia 00d0.f822.33af)
Interface address is: 192.168.1.2/24
ARP type: ARPA, ARP Timeout: 3600 seconds
  MTU 1500 bytes, BW 1000000 Kbit
  Encapsulation protocol is Ethernet-II, loopback not set
  Keepalive interval is 10 sec , set
  Carrier delay is 2 sec
  Rxload is 0/255, Txload is 0/255
```

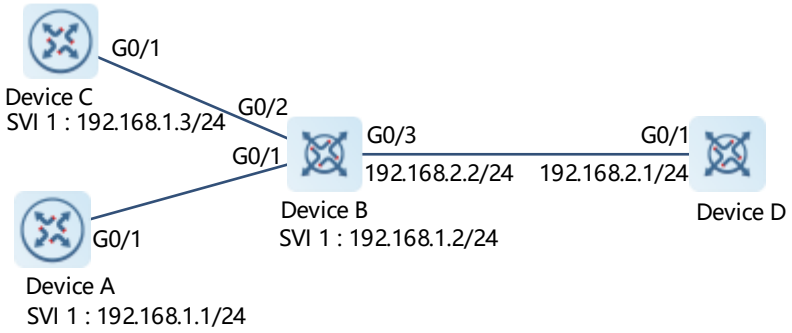
1.8.2 Configuring Interconnection Interfaces

1. Requirements

Interconnect two devices, and configure attributes of the device interfaces.

2. Topology

Figure 1-5 Configuring Interconnection Interfaces



3. Notes

- On Device A, configure GigabitEthernet 0/1 as a Layer 2 port in access mode, with the default VLAN ID as 1. Configure SVI 1, assign an IP address to SVI 1, and set up a route to Switch D.
- On Device B, configure GigabitEthernet 0/1 and GigabitEthernet 0/2 as Layer 2 ports in trunk mode, with the native VLAN ID as 1. Configure SVI 1, and assign an IP address to SVI 1. Configure GigabitEthernet 0/3 as a Layer 3 port, and assign an IP address in another network segment to this port.
- On Device C, configure GigabitEthernet 0/1 as a Layer 2 port in access mode, with the default VLAN ID as 1. Configure SVI 1, and assign an IP address to SVI 1.
- On Device D, configure GigabitEthernet 0/1 as a Layer 3 port, assign an IP address to this port, and set up a route to Switch A.

4. Procedure

Configure as follows on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port mode access
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port access vlan 1
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface vlan 1
DeviceA(cfg-if-VLAN 1)# ip address 192.168.1.1 255.255.255.0
DeviceA(cfg-if-VLAN 1)# exit
DeviceA(cfg)# ip route-config 192.168.2.0 255.255.255.0 VLAN 1 192.168.1.2
```

Configure as follows on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk-port
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
DeviceB(cfg)# interface GigabitEthernet 0/2
```

```
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port mode trunk-port
DeviceB(cfg-if-GigabitEthernet 0/2)# exit
DeviceB(cfg)# interface vlan 1
DeviceB(cfg-if-VLAN 1)# ip address 192.168.1.2 255.255.255.0
DeviceB(cfg-if-VLAN 1)# exit
DeviceB(cfg)# interface GigabitEthernet 0/3
DeviceB(cfg-if-GigabitEthernet 0/3)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/3)# ip address 192.168.2.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/3)# exit
```

Configure as follows on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/1
DeviceC(cfg-if-GigabitEthernet 0/1)# port-group 1
DeviceC(cfg-if-GigabitEthernet 0/1)# exit
DeviceC(cfg)# interface aggregateport 1
DeviceC(cfg-if-AggregatePort 1)# switch-port mode access
DeviceC(cfg-if-AggregatePort 1)# switch-port access vlan 1
DeviceC(cfg-if-AggregatePort 1)# exit
DeviceC(cfg)# interface vlan 1
DeviceC(cfg-if-VLAN 1)# ip address 192.168.1.3 255.255.255.0
DeviceC(cfg-if-VLAN 1)# exit
```

Configure as follows on Device D.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# interface GigabitEthernet 0/1
DeviceD(cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.1 255.255.255.0
DeviceD(cfg-if-GigabitEthernet 0/1)# exit
DeviceD(cfg)# ip route-config 192.168.1.0 255.255.255.0    GigabitEthernet 0/1 192.168.2.2
```

5. Verification

Perform verification on Device A, Device B, Device C, and Device D respectively as follows:

- On Switch A, ping the IP addresses of interfaces of the other three devices. Verify that you can access the other two on each device.
- Verify that Device B and Device D can be pinged mutually.
- Verify that the interface status is correct.

Check the status of GigabitEthernet 0/1 on Device A.

```
DeviceA# display interfaces gigabitEthernet 0/1
Index(dec):1 (hex):1
GigabitEthernet 0/1 is UP    , line protocol is UP
Hardware is GigabitEthernet, address is 00d0.f865.de90 (bia 00d0.f865.de90)
Interface address is: unset ip address
  MTU 1500 bytes, BW 100000 Kbit
```

```
  Encapsulation protocol is Ethernet-II, loopback not set
  Keepalive interval is 10 sec , set
  Carrier delay is 2 sec
  Ethernet attributes:
    Last link state change time: 2012-12-22 14:00:48
    Time duration since last link state change: 3 days, 2 hours, 50 minutes, 50 seconds
    Priority is 0
    Admin medium-type is Copper, oper medium-type is Copper
    Admin duplex mode is AUTO, oper duplex is Full
    Admin speed is AUTO, oper speed is 100M
    Flow control admin status is OFF, flow control oper status is OFF
    Admin negotiation mode is OFF, oper negotiation state is ON
Storm Control: Broadcast is OFF, Multicast is OFF, Unicast is OFF
  Bridge attributes:
Port-type: access
Vlan id: 1
  Rxload is 1/255, Txload is 1/255
  10 seconds input rate 0 bits/sec, 0 packets/sec
  10 seconds output rate 67 bits/sec, 0 packets/sec
    362 packets input, 87760 bytes, 0 unset buffer, 0 dropped
    Received 0 broadcasts, 0 runts, 0 giants
    0 input errors, 0 CRC, 0 frame, 0 overrun, 0 abort
    363 packets output, 82260 bytes, 0 underruns , 0 dropped
    0 output errors, 0 collisions, 0 interface resets
```

Check the status of GigabitEthernet 0/1 on Device B.

```
DeviceB# display interfaces gigabitEthernet 0/1
Index(dec):1 (hex):1
GigabitEthernet 0/1 is UP    , line protocol is UP
Hardware is GigabitEthernet, address is 00d0.f865.de91 (bia 00d0.f865.de91)
Interface address is: unset ip address
  MTU 1500 bytes, BW 100000 Kbit
  Encapsulation protocol is Ethernet-II, loopback not set
  Keepalive interval is 10 sec , set
  Carrier delay is 2 sec
  Ethernet attributes:
    Last link state change time: 2012-12-22 14:00:48
    Time duration since last link state change: 3 days, 2 hours, 50 minutes, 50 seconds
    Priority is 0
    Admin medium-type is Copper, oper medium-type is Copper
    Admin duplex mode is AUTO, oper duplex is Full
    Admin speed is AUTO, oper speed is 100M
    Flow control admin status is OFF, flow control oper status is OFF
    Admin negotiation mode is OFF, oper negotiation state is ON
Storm Control: Broadcast is OFF, Multicast is OFF, Unicast is OFF
  Bridge attributes:
    Port-type: trunk-port
```

```
    Native vlan: 1
    Allowed vlan lists: 1-4094
Active vlan lists: 1
  Rxload is 1/255, Txload is 1/255
  10 seconds input rate 0 bits/sec, 0 packets/sec
  10 seconds output rate 67 bits/sec, 0 packets/sec
362 packets input, 87760 bytes, 0 unset buffer, 0 dropped
Received 0 broadcasts, 0 runts, 0 giants
0 input errors, 0 CRC, 0 frame, 0 overrun, 0 abort
363 packets output, 82260 bytes, 0 underruns , 0 dropped
0 output errors, 0 collisions, 0 interface resets
```

Check the status of GigabitEthernet 0/1 on Device C.

```
DeviceC# display interfaces gigabitEthernet 0/1
Index(dec):1 (hex):1
GigabitEthernet 0/1 is UP    , line protocol is UP
Hardware is GigabitEthernet, address is 00d0.f865.de92 (bia 00d0.f865.de92)
Interface address is: unset ip address
  MTU 1500 bytes, BW 100000 Kbit
  Encapsulation protocol is Ethernet-II, loopback not set
  Keepalive interval is 10 sec , set
  Carrier delay is 2 sec
  Ethernet attributes:
Last link state change time: 2012-12-22 14:00:48
Time duration since last link state change: 3 days, 2 hours, 50 minutes, 50 seconds
Priority is 0
Admin medium-type is Copper, oper medium-type is Copper
Admin duplex mode is AUTO, oper duplex is Full
Admin speed is AUTO, oper speed is 100M
Flow control admin status is OFF, flow control oper status is OFF
Admin negotiation mode is OFF, oper negotiation state is ON
Storm Control: Broadcast is OFF, Multicast is OFF, Unicast is OFF
  Rxload is 1/255, Txload is 1/255
  10 seconds input rate 0 bits/sec, 0 packets/sec
  10 seconds output rate 67 bits/sec, 0 packets/sec
362 packets input, 87760 bytes, 0 unset buffer, 0 dropped
Received 0 broadcasts, 0 runts, 0 giants
0 input errors, 0 CRC, 0 frame, 0 overrun, 0 abort
363 packets output, 82260 bytes, 0 underruns , 0 dropped
0 output errors, 0 collisions, 0 interface resets
```

Check the status of GigabitEthernet 0/1 on Device D.

```
DeviceD# display interfaces gigabitEthernet 0/1
Index(dec):1 (hex):1
GigabitEthernet 0/1 is UP    , line protocol is UP
Hardware is GigabitEthernet, address is 00d0.f865.de93 (bia 00d0.f865.de93)
Interface address is: 192.168.2.1/24
```

```
  MTU 1500 bytes, BW 100000 Kbit
  Encapsulation protocol is Ethernet-II, loopback not set
  Keepalive interval is 10 sec , set
  Carrier delay is 2 sec
  Ethernet attributes:
Last link state change time: 2012-12-22 14:00:48
Time duration since last link state change: 3 days, 2 hours, 50 minutes, 50 seconds
Priority is 0
Admin medium-type is Copper, oper medium-type is Copper
Admin duplex mode is AUTO, oper duplex is Full
Admin speed is AUTO, oper speed is 100M
Flow control admin status is OFF, flow control oper status is OFF
Admin negotiation mode is OFF, oper negotiation state is ON
Storm Control: Broadcast is OFF, Multicast is OFF, Unicast is OFF
  Rxload is 1/255, Txload is 1/255
  10 seconds input rate 0 bits/sec, 0 packets/sec
  10 seconds output rate 67 bits/sec, 0 packets/sec
362 packets input, 87760 bytes, 0 unset buffer, 0 dropped
Received 0 broadcasts, 0 runts, 0 giants
0 input errors, 0 CRC, 0 frame, 0 overrun, 0 abort
363 packets output, 82260 bytes, 0 underruns , 0 dropped
0 output errors, 0 collisions, 0 interface resets
```

Contents

1 Configuring Aggregate Interface 1

1.1 Introduction 1

1.1.1 Basic Concepts 1

1.1.2 Load Balancing 4

1.1.3 Member Port BFD Detection 7

1.1.4 Protocols and Standards 7

1.2 Configuration Task Summary 7

1.3 Notes 7

1.4 Configuring Static Aggregate Interfaces 8

1.4.1 Overview 8

1.4.2 Restrictions and Guidelines 8

1.4.3 Procedure 9

1.5 Configuring LACP Interfaces 10

1.5.1 Overview 10

1.5.2 Restrictions and Guidelines 10

1.5.3 Procedure 10

1.6 Enabling LinkTrap 11

1.6.1 Overview 11

1.6.2 Procedure 12

1.7 Configuring a Load Balancing Mode 12

1.7.1 Overview 12

1.7.2 Restrictions and Guidelines 12

1.7.3 Procedure....12
1.8 Configuring a Capacity Mode....15
1.8.1 Overview....15
1.8.2 Restrictions and Guidelines....15
1.8.3 Procedure....15
1.9 Enabling BFD for Member Ports....16
1.9.1 Overview....16
1.9.2 Restrictions and Guidelines....16
1.9.3 Procedure....16
1.10 Configuring the Minimum Number of Member Ports....16
1.10.1 Overview....16
1.10.2 Restrictions and Guidelines....17
1.10.3 Procedure....17
1.11 Enabling the LACP Independent Port Function....17
1.11.1 Overview....17
1.11.2 Restrictions and Guidelines....18
1.11.3 Procedure....18
1.12 Monitoring....18
1.13 Configuration Examples....19
1.13.1 Configuring Static Aggregate Interfaces....19
1.13.2 Configuring LACP Interfaces....21
1.13.3 Enabling LinkTrap....23
1.13.4 Configuring a Load Balancing Mode....25
1.13.5 Configuring Hash-based Load Balancing Control....27

1.13.6 Configuring a Capacity Mode 29

1.13.7 Enabling BFD for Member Ports 30

1.13.8 Configuring the Minimum Number of Member Ports (When the Number of Member Ports Is Less Than the Minimum Number) 33

1.13.9 Configuring the Minimum Number of Member Ports (When the Number of Member Ports Is Not Less Than the Minimum Number) 35

1.13.10 Enabling the LACP Independent Port Function 37

1 Configuring Aggregate Interface

1.1 Introduction

An aggregate interface is used to bundle multiple physical links into one logical link to increase the link bandwidth and improve connection reliability.

1.1.1 Basic Concepts

An aggregate interface supports load balancing, that is, distributes load evenly among member links. Besides, an aggregate interface realizes link backup. When a member link of the aggregate interface is disconnected, the load carried by the link is automatically allocated to other functional member links. A member link does not forward broadcast or multicast packets to other member links.

- If the link between two devices supports a maximum bandwidth of 1,000 Mbps (assuming that the devices each have a 1,000 Mbps port), when the service traffic carried by the link exceeds 1,000 Mbps, the traffic in excess will be discarded. Port aggregation can be used to solve the problem. For example, you can connect the two devices with network cables and combine multiple links to form a logical link capable of multiples of 1,000 Mbps.
- If the two devices are connected by one network cable, when the link between the two ports of the devices is disconnected, the services carried by the link will be interrupted. After the connected ports are aggregated, the services will not be affected as long as one link remains connected.

There are two kinds of link aggregation. One is the static aggregate interface mode, and the other is dynamic aggregation through Link Aggregation Control Protocol (LACP).

1. Static Aggregate Interface

The static aggregate interface mode means manually configuring physical ports and adding them to a link aggregation group. An aggregate interface in static aggregate interface mode is called a static aggregate interface, and its member ports are called member ports of a static aggregate interface. The static aggregate interface configuration is simple. You only need to run a command to add the specified physical port to the aggregate interface. After joining the aggregation group, a member port can receive and transmit data and participate in load balancing within the group.

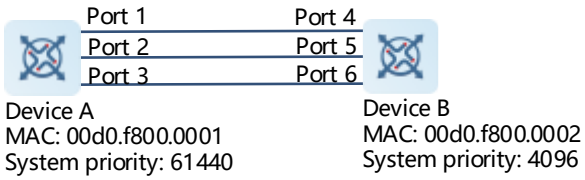
2. Dynamic Aggregate Interface (LACP)

LACP is a protocol about dynamic link aggregation. It exchanges information with the connected device through link aggregation control protocol data units (LACPDU). An aggregate interface in LACP mode is called a dynamic aggregate interface, and its member ports are called member ports of a dynamic aggregate interface.

An LACP-enabled port sends LACPDUs to advertise its system priority, system MAC address, port priority, port number, and operation key. When receiving the LACPDU from the peer end, the device compares the system priorities of both ends based on the system ID in the packet. The end with a higher system ID priority sets the ports in the aggregation group to **Bundle** state based on the port ID priorities in a descending order, and sends an updated LACPDU. When receiving the LACPDU, the peer end sets corresponding ports to **Bundle** state so that both ends maintain consistency when a port exits or joins the aggregation group. The physical link can forward packets only after the ports at both ends are bundled dynamically.

After link aggregation, the LACP member ports periodically exchange LACPDUs. When a port does not receive an LACPDU in the specified time, a timeout occurs and the links are unbundled. In this case, the member ports cannot forward packets. There are two timeout modes: long timeout and short timeout. In long timeout mode, a port sends a packet every 30s. If it does not receive a packet from the peer end in 90s, a timeout occurs. In short timeout mode, a port sends a packet every 1s. If it does not receive a packet from the peer end in 3s, a timeout occurs.

Figure 1-1 LACP Negotiation



As shown in Figure 1-1, Device A is connected to Device B through three ports. Set the system priority of Device A to 61440 and that of Device B to 4096. Enables LACP link aggregation on the three directly connected ports of Device A and Device B, set the aggregation mode of the three ports to the active mode, and set the port priority to the default value 32768.

When receiving an LACPDU from Device A, Device B finds that it has a higher system ID priority than Device A (a smaller value indicates a higher priority). Therefore, Device B sets Port 4, Port 5, and Port 6 to **Bundle** state based on the order of port ID priorities (or in an ascending order of port numbers if the port priorities are the same). When receiving an updated LACPDU from Device B, Device A finds that Device B has a higher system ID priority and has set Port 4, Port 5, and Port 6 to **Bundle** state. Then Device A also sets Port 1, Port 2, and Port 3 to **Bundle** state.

3. Member Port Mode

There are three aggregation modes available, namely, active, passive, and static.

Member ports in active mode initiate LACP negotiation. Member ports in passive mode only respond to received LACPDUs. Member ports in static mode do not send LACPDUs for negotiation. The following table lists the requirements for peer port mode.

Table 1-1 Requirements for Peer Port Aggregation in Aggregation Mode

Port Mode	Peer Port Mode
Active mode	Active or passive mode
Passive mode	Active mode
Static Mode	Static Mode

4. Member Port State

Static member ports can be in **Up** or **Down** state.

- When a member port is **Down**, the port cannot forward packets. The **Down** state is displayed.

- When a member port is **Up** and the link protocol is ready, the port can forward packets. The **Up** state is displayed.

LACP member ports has the following three states:

- When the link of a port is **Down**, the port cannot forward packets. The **Down** state is displayed.
- When the link of a port is **Up** and the port is added to an aggregation group (as a member of the aggregation group, the port participates in data packet forwarding of the aggregation group) after LACP negotiation, the **bndl** state is displayed.
- When the link of a port is **Up** but the port is suspended because the peer end is not enabled with LACP or the attributes of the ports are inconsistent with those of the master port, the **susp** state is displayed. (The port in **susp** state does not forward packets.)

Note

- Only full-duplex ports are capable of LACP aggregation.
- LACP aggregation can be implemented only when the rates, flow control approaches, medium types, and L2/L3 attributes of member ports are consistent.
- If you modify the preceding attributes of a member port in the aggregation group, LACP aggregation will fail.
- The ports which are prohibited from joining or exiting an aggregate interface cannot be added to or removed from a static aggregate interface or an LACP aggregate interface.

5. Capacity Mode

The maximum number of member ports is fixed, which is equal to the maximum number of aggregate interfaces multiplied by the maximum number of member ports supported by a single aggregate interface. If you want to increase the maximum number of aggregate interfaces, the maximum number of member ports supported by a single aggregate interface must be reduced, and vice versa. This concerns the aggregate interface capacity mode concept. Some devices support the configuration of the aggregate interface capacity mode. For example, if the system supports 16,384 member ports, you can select 1024 x 16, 512 x 32, and other aggregate interface capacity modes (Maximum number of aggregate interfaces multiplied by the maximum number of member ports supported by a single aggregate interface).

6. LACP System ID

By default, all the LACP interfaces on a device belong to the same LACP aggregation group.

The system is identified by a system ID and each system has a priority, which is a configurable value. The system ID consists of the LACP system priority and MAC address of the device. A lower system priority indicates a higher priority of the system ID. If the system priorities are the same, a smaller MAC address of the device indicates a higher priority of the system ID. The system with an ID of a higher priority determines the port state. The port state of a system with an ID of a lower priority keeps consistent with that of a higher priority.

The LACP system ID can be configured when LACP interfaces of multiple (a maximum of four) independent devices need to negotiate with the LACP interface of a specific device (for example, LACP interfaces of two independent ASWs need to negotiate with the LACP interface of the NC). You can set the system IDs of the LACP interfaces of independent devices to the same MAC address and configure different device IDs to implement normal negotiation.

7. LACP Device ID

The LACP device ID can be configured when LACP interfaces of multiple independent devices need to negotiate with the LACP interface of a specific device. It must be configured together with the system ID.

8. LACP Port ID

Each port has an independent LACP interface priority, which is a configurable value. The port ID consists of the LACP interface priority and port number. A smaller port priority indicates a higher priority of the port ID. If the port priorities are the same, a smaller port number indicates a higher priority of the port ID.

9. LACP Master Port

When dynamic member ports are **Up**, LACP selects one of those ports to be the master port based on the rates and duplex modes, ID priorities of the ports in the aggregation group, and the bundling state of the member ports in the **Up** state. Only the ports that have the same attributes as the master port are in **Bundle** state and participate in data forwarding. When the attributes of ports are changed, LACP reselects a master port.

10. Minimum Number of Member Ports

An aggregate interface can be configured with a minimum number of member ports. When a member port exits the link aggregation group, causing the number of member ports to be smaller than the minimum number, the other member ports in the group are unbundled, and the aggregate interface is **Down**. When the member port rejoins the group, causing the number of member ports to be greater than the minimum number, the member ports in the group are automatically bundled, and the aggregate interface is **Up** again.

11. LACP Independent Ports

In normal cases, LACP independent ports are used for interworking between access switches and servers with two NICs. If the operating system (OS) is not pre-installed when a server with two NICs starts, the OS needs to be installed via the remote preboot execution environment (PXE) OS installation device. Before the OS is installed, the server with two NICs cannot perform LACP negotiation with the access device, and only one NIC can work. In this case, the port on the access device must be able to automatically change to a common Ethernet physical port to ensure normal communication between the server and the remote PXE OS installation device. After the OS is installed and both NICs can run the LACP, the port on the access device must be able to enable the LACP again for negotiation.



Caution

- LACP independent ports can work only at layer 2. After an LACP independent port is enabled through configuration, if the LACP independent port does not receive LACP packets, it automatically changes to a common Ethernet port, which automatically copies the rate, duplex mode, flow control, and VLAN configuration from the aggregate interface to ensure port forwarding capabilities.
- An LACP independent port automatically changes to a common Ethernet port only if it does not receive LACP packets within the set timeout period. After the port receives LACP packets, it changes to an LACP member port again.

1.1.2 Load Balancing

Aggregated ports segregate packet flows by using load balancing algorithms based on packet features, such as the source and destination MAC addresses, source and destination IP addresses, and L4 source and destination port numbers. The packet flow with the consistent feature is transmitted by one member link, and

different packet flows are evenly distributed to member links. For example, in source MAC address-based load balancing, packets are distributed to the member links based on the source MAC addresses of the packets. Packets with different source MAC addresses are evenly distributed to member links. Packets with the identical source MAC address are forwarded by one member link.

Currently, there are several link aggregation load balancing modes as follows:

- Source MAC address or destination MAC address
- Source MAC address + destination MAC address
- Source IP address or destination IP address
- Source IP address + destination IP address
- L4 source port number or L4 destination port number
- L4 source port number + L4 destination port number
- Source IP address + L4 source port number
- Source IP address + L4 destination port number
- Destination IP address + L4 source port number
- Destination IP address + L4 destination port number
- Source IP address + L4 source port number + L4 destination port number
- Destination IP address + L4 source port number + L4 destination port number
- Source IP address + destination IP address + L4 source port number
- Source IP address + destination IP address + L4 destination port number
- Source IP address + destination IP address + L4 source port number + L4 destination port number
- Panel port for incoming packets
- Labels of Multiprotocol Label Switching (MPLS) packets
- Aggregation member port polling
- Enhanced load balancing mode

Note

- Load balancing based on IP addresses or port numbers is applicable only to L3 packets. When a device enabled with this load balancing method receives L2 packets, it automatically switches to the default load balancing method.
- All the load balancing methods use a load algorithm (hash algorithm) to calculate the member links based on the input parameters of the methods. The input parameters include the source MAC address, destination MAC address, source MAC address + destination MAC address, source IP address, destination IP address, source IP address + destination IP addresses, source IP address + destination IP address + L4 port number and so on. The algorithm ensures that packets with different input parameters are evenly distributed to member links. It does not indicate that these packets are always distributed to different member links. For example, in IP address-based load balancing, two packets with different source and destination IP addresses may be distributed to the same member link through calculation.

1. Enhanced Load Balancing

Enhanced load balancing allows the combination of multiple fields in different types of packets. These fields include **src-mac**, **dst-mac**, **l2-protocol**, **vlan**, **src-port**, **src-dst-mac**, and **dst-port** in L2 packets, **src-ip**,

dst-ip, **protocol**, **l4-src-port**, **l4-dst-port**, **vlan**, **src-port**, **dst-port**, **l2-etype**, **src-mac**, **src-dst-ip**, **src-dst-l4port**, and **dst-mac** in IPv4 packets, **src-ip**, **dst-ip**, **protocol**, **l4-src-port**, **l4-dst-port**, **vlan**, **src-port**, **dst-port**, **l2-etype**, **src-mac**, **flow-label**, **src-dst-ip**, **src-dst-l4port,** and **dst-mac** in IPv6 packets; **top-label**, **2nd-label**, **3rd-label**, **src-ip**, **dst-ip**, **vlan**, **src-port**, **dst-port**, **src-mac**, **dst-mac**, **protocol**, **l4-src-port**, **l4-dst-port**, and **l2-etype** in MPLS packets; **vlan**, **src-port**, **src-mac**, **src-ip**, **protocol**, **l4-src-port**, **l4-dst-port**, **l2-etype**, **ing-nick**, **egr-nick**, **dst-port**, **dst-mac**, and **dst-ip** in TRILL packets, and **vlan**, **src-port**, **src-id**, **rx-id**, **ox-id**, **fabric-id**, **dst-port**, and **dst-id** in FCoE packets.

A device enabled with enhanced load balancing first determines the types of packets to be transmitted and performs load balancing based on the specified fields in the packets. For example, the aggregate interface performs source IP-based load balancing on the packets containing an ever-changing source IPv4 address.

Note

- All the load balancing methods are applicable to L2 and L3 aggregate interfaces. You need to configure proper load distribution methods based on different network environments to fully utilize network bandwidth.
- Perform enhanced load balancing based on the **src-mac**, **dst-mac**, and **vlan** fields in L2 packets, and the **src-ip** field in IPv4 packets. If the incoming packet is an IPv4 packet with an ever-changing source MAC address, the enhanced balancing algorithm does not take effect, because the device performs load balancing only based on the **src-ip** field in the IPv4 packet after finding that it is an IPv4 packet.
- In enhanced load balancing, the MPLS balancing algorithm takes effect only for MPLS L3 VPN packets, but does not take effect for MPLS L2 VPN packets.

2. Hash Load Balancing Control

Hash load balancing enables users to flexibly control load balancing in different scenarios. Currently, Ruijie adopts the following hash load balancing control functions:

Hash disturbance factor: Traffic over aggregate interfaces is hashed for balancing. For two devices of the same type, the same path is calculated for load balancing for the same stream. When the equal-cost multi-path (ECMP) is deployed, the same stream of the two devices may be balanced to the same destination device, resulting in hash polarization. The hash disturbance factor is used to affect the load balancing algorithm. Different disturbance factors are configured for different devices to ensure that different paths are provided for the same stream.

Symmetric hashing: To ensure network security, a firewall cluster is deployed between the internal and external networks for traffic cleaning. This requires that both the uplink and downlink traffic of a session is transmitted to the same device in the firewall cluster for processing. The source and destination IP addresses contained in the uplink and downlink streams of a session are reversed. The uplink and downlink streams will be directed to different firewalls in the firewall cluster based on the traditional hash algorithm. Symmetric hashing ensures that uplink and downlink streams of a session be transmitted over the same path.

Flexible hash: In normal cases, the load balancing cluster interworks with multiple ToRs via the ECMP, and the ToRs are required to balance all packets of a session to the same scheduling server. However, when a server is faulty, all session streams will be ordered and balanced again based on the traditional load balancing mechanism of ToRs. In this case, packets of the same session may be balanced to two servers. The flexible hash function ensures that when one link is faulty, traffic on other links is not affected and only the traffic of the faulty link is balanced to other active links. In this case, traffic of a session is balanced to the same server.

Hash algorithm mode: Apply the most appropriate hash algorithm mode to different traffic models, so that when traffic changes, balance can still be kept. For example, if the source and destination MAC addresses of a stream increase by 1 at the same time, configuring the algorithm based on the source and destination MAC addresses cannot keep stream balance. In this case, a suitable hash algorithm mode needs to be applied.

Hash factor acquisition mode: There are an inner layer and an outer layer in the header in each of VXLAN packets, GRE packets, and other tunnel packets. It can be specified to obtain the hash factor from the inner layer or outer layer, to achieve a better balancing effect. For example, in some scenarios, tunnel packets share the same outer IP address but have different inner IP addresses. In this case, the inner IP address can be specified as the hash factor to achieve an optimal balancing effect.

Note

The performance of hash algorithm mode may vary with different devices.

1.1.3 Member Port BFD Detection

Bidirectional Forwarding Detection (BFD) is a protocol that delivers fast detection of path failures. According to RFC7130, LACP takes 3s to detect link failures even in short timeout mode. All the packets sent to the faulty link during the 3-second period will be dropped. BFD correlation delivers faster failure detection. You can configure BFD on member ports to detect link failure and switch load to other member links in case of a link failure. BFD is classified into IPv4 detection and IPv6 detection, which detect IPv4 and IPv6 paths respectively. When BFD detects that the path on a member port fails, the packets are not distributed to the member port. After BFD is enabled on an aggregate interface, BFD sessions are set up on its member ports in forwarding state independently.

1.1.4 Protocols and Standards

- IEEE 802.3ad

1.2 Configuration Task Summary

The aggregate interface configuration includes the following tasks:

(1) Configuring Static Aggregate Interfaces

(2) (Optional) Configuring LACP Interfaces

(3) (Optional) Enabling LinkTrap

(4) (Optional) Configuring a Load Balancing Mode

(5) (Optional) Configuring a Capacity Mode

(6) (Optional) Enabling BFD for Member Ports

(7) (Optional) Configuring the Minimum Number of Member Ports

(8) (Optional) Enabling the LACP Independent Port Function

1.3 Notes

This section describes the precautions for configuring the link aggregate interface function.

Whether the link aggregate interface function is supported depends on devices.

Table 1-2 Products Supporting Link Aggregate Interface

Function	Product Supporting the Function
Configuring a Capacity Mode	S6980-64QC, S6920-4C, S6580-48CQ8QC, S6580-24DC8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ

1.4 Configuring Static Aggregate Interfaces

1.4.1 Overview

- Configure multiple physical ports as member ports to realize link aggregation.
- The bandwidth of the aggregation link is equal to the sum of the member link bandwidths.
- When a member link of the aggregate interface is disconnected, the load carried by the link is automatically allocated to other functional member links.

1.4.2 Restrictions and Guidelines

- Only physical ports can be added to an aggregate interface.
- Before adding an interface to a static aggregate interface, run the **shutdown** command in the configuration mode of the interface to prevent flapping. After adding the interface and its peer interface to the static aggregate interface, remove the **shutdown** command configuration.
- To prevent a transient loop when a member port exits, you need to set the member port to **shutdown** state before it exits the aggregate interface.
- The ports of different medium types or port modes cannot be added to the same aggregate interface.
- L2 ports can be added to only a L2 aggregate interface, and L3 ports can be added to only a L3 aggregate interface. The L2/L3 attributes of an aggregate interface that contains member ports cannot be modified.
- After a port is added to an aggregate interface, the attributes of the port are replaced by those of the aggregate interface.
- After a port is removed from an aggregate interface, the attributes of the port before the adding to the aggregate interface are restored.
- There is continuous traffic at the aggregate interface. If a member port changes from **Link Up** to **Link Down**, an error may occur in traffic statistics of the aggregate interface.
- When the aggregate interface includes a member port that supports **port speed-mode** configuration, the aggregate interface does not support speed configuration, and the original speed configuration on the aggregate interface is automatically cleared. The configured rate of the aggregate interface is switched to that of **port speed-mode** for the member port.
- You can run the **port-group** command to add a physical port to a static aggregate interface in the interface configuration mode of the specified Ethernet port. If the aggregate interface does not exist, it will be created automatically.
- You can run the **port-group mode** command to add a physical port to an LACP aggregate interface in the interface configuration mode of the specified physical port. If the aggregate interface does not exist, it will be

created automatically.

- The aggregate interface feature must be configured on the devices at both ends of a link and the aggregate interface mode must be the same (static aggregate interface or LACP aggregate interface).
- After a member port exits the aggregate interface, the default configurations of the member port may be restored. The original configurations of aggregate interface members are processed in different ways for different functions. You are advised to check and confirm the port configurations after a member port exits an aggregate interface.
- After a member port exits an aggregate interface, the port is shut down to avoid loops. After confirming that the topology is normal, you can run **unset shutdown** in interface configuration mode to enable the port again.
- If the VLAN resource is in a critical state (the number of configured VLANs and the number of routed ports reach the VLAN capacity values) when a member port of L3 aggregate interface exits the aggregate interface, the port may be switched to a L2 port.
- The aggregate interface created on a L3 device that does not support L2 features is a L3 aggregate interface; otherwise the aggregate interface is a L2 aggregate interface.

Warning

- After a port is added to an aggregate interface, the attributes of the port are consistent with those of the aggregate interface. Therefore, it is generally not allowed to perform configuration on the member ports or apply configurations to a specific member port. However, some commands or functions (such as the **shutdown** and **unset shutdown** commands) can be configured on the member ports. When using the member ports, you need to check whether the function that you want to configure can take effect on a specific member port, and perform the configuration properly.
- The error detection function is unavailable to static aggregation. Member ports are bound if linked up. Therefore, you need to ensure topology accuracy and negotiated attribute consistency among the member ports.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an Ethernet aggregate interface.

interface aggregateport *ap-number*

Perform this configuration on an aggregate interface-enabled device. When an Ethernet port needs the aggregation function, create the corresponding Ethernet aggregate interface

No aggregate interface is configured by default.

(4) Enter the interface configuration mode.

interface *interface-type interface-number*

(5) Configure member ports of a static aggregate interface.

port-group *ap-number*

Perform this configuration on aggregate interface-enabled devices. To use the static aggregation function, configure the corresponding member ports of a static aggregate interface.

By default, the Ethernet physical port does not belong to any static aggregate interface or LACP aggregate interface.

(6) (Optional) Convert L2 aggregate interfaces to L3 aggregate interfaces.

unset switch-port

The aggregate interface is a L2 port by default.

(7) (Optional) Create a sub-interface of an Ethernet aggregate interface.

interface aggregateport *sub-ap-number*

No sub-interface of an Ethernet aggregate interface is configured by default.

1.5 Configuring LACP Interfaces

1.5.1 Overview

- Connected devices perform auto negotiation through LACP to realize dynamic link aggregation.
- The bandwidth of the aggregation link is equal to the sum of the member link bandwidths.
- When a member link of the aggregate interface is disconnected, the load carried by the link is automatically allocated to other functional member links.
- It takes 90s for LACP to detect a link failure in long timeout mode and 3s in short timeout mode.

1.5.2 Restrictions and Guidelines

- After a port exits an LACP aggregate interface, the default configurations of the port may be restored. The original configurations of LACP members are processed in different ways for different functions. You are advised to check and confirm the port configurations after a member port exits an LACP aggregate interface.
- To prevent a transient loop when a member port exits, you need to set the member port to **shutdown** state before it exits the aggregate interface.
- Changing the LACP system priority may cause LACP member ports to be disaggregated and aggregated again.
- Changing the priority of an LACP member port may cause the other member ports to be disaggregated and aggregated again.
- There is continuous traffic at the aggregate interface. If a member port changes from **Link Up** to **Link Down**, an error may occur in traffic statistics of the aggregate interface.
- When the aggregate interface includes a member port that supports **port speed-mode** configuration, the aggregate interface does not support speed configuration, and the original speed configuration on the aggregate interface is automatically cleared. The configured rate of the aggregate interface is switched to that of **port speed-mode** for the member port.
- The LACP member port configurations at both ends of a link must be consistent.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure LACP member ports.

port-group *key-number* **mode** { **active** | **passive** }

By default, the Ethernet physical port does not belong to any LACP aggregate interface.

(5) (Optional) Configure the LACP system ID.

lacp system-id *system-id*

Configure the LACP system ID when LACP interfaces of multiple (a maximum of four) independent devices need to negotiate with the LACP interface of a specific device.

The LACP system ID is the system ID (MAC address) of the device by default.

(6) (Optional) Configure the LACP device ID.

lacp device *number*

Configure the LACP device ID when LACP interfaces of multiple (a maximum of four) independent devices need to negotiate with the LACP interface of a specific device.

The device ID is 0 by default.

(7) (Optional) Configure the LACP system priority.

lacp system-priority *system-priority*

Perform this configuration when you need to adjust the system ID priority. A smaller value indicates a higher system ID priority. The device with a higher system ID priority takes precedence in selecting an aggregate interface.

The LACP system priority of a port is 32768 by default.

(8) Configure the priority of an LACP member port.

lacp port-priority *port-priority*

Perform this configuration when you need to specify the port ID priority. A smaller value indicates a higher port ID priority. The port with the highest port ID priority will be selected as the master port.

The priority of the LACP member port is 32768 by default.

(9) Configure the timeout mode of LACP member ports.

lacp short-timeout

When you need to implement real-time link failure detection, configure the short timeout mode. A timeout occurs if a port does not receive a packet within 3 seconds in short timeout mode and 90 seconds in long timeout mode.

The timeout mode of an LACP member port is long timeout by default.

1.6 Enabling LinkTrap

1.6.1 Overview

Enable the system with LinkTrap to send LinkTrap messages when aggregation links are changed.

1.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the aggregate interface configuration mode.

interface Aggregateport *ap-interface-number*

(4) (Optional) Configure the LinkTrap function of an aggregate interface.

snmp trap link-status

The LinkTrap notification function of an aggregate interface is enabled by default.

(5) (Optional) Configure the LinkTrap function of member ports.

aggregateport member linktrap

The LinkTrap notification function of member ports is disabled by default.

1.7 Configuring a Load Balancing Mode

1.7.1 Overview

- The system distributes incoming packets among member links by using the specified load balancing algorithm. The packet flow with the consistent feature is transmitted by one member link, whereas different packet flows are evenly distributed to various links. A device enabled with enhanced load balancing first determines the types of packets to be transmitted and performs load balancing based on the specified fields in the packets. For example, the aggregate interface performs source IP-based load balancing on the packets containing an ever-changing source IPv4 address.
- In enhanced load balancing mode, configure the hash disturbance factor to ensure that same packets from two devices of the same type is balanced to different links.
- In enhanced load balancing mode, enable symmetric hashing to ensure that uplink and downlink packets of the same type are transmitted over the same link. For example, in load balancing based on the source and destination IP addresses, enable symmetric hashing for IPv4 packets to ensure that the uplink and downlink IPv4 packets are transmitted over the same path.
- Enable flexible hash to balance traffic of a faulty link over a member port to other active links without affecting original data streams on these active links.

1.7.2 Restrictions and Guidelines

Different disturbance factors may lead to the same disturbance effect.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the global load balancing algorithm of an aggregate interface.

aggregateport load-balance { **dst-ip** | **dst-ip-dst-l4port** | **dst-ip-src-dst-l4port** | **dst-ip-src-l4port** | **dst-l4port** | **dst-mac** | **enhanced profile** *profile-name* | **mpls-label** | **src-dst-ip** | **src-dst-ip-dst-l4port** | **src-dst-ip-l4port** | **src-dst-ip-src-l4port** | **src-dst-l4port** | **src-dst-mac** | **src-ip** | **src-ip-dst-l4port** | **src-ip-src-dst-l4port** | **src-ip-src-l4port** | **src-l4port** | **src-mac** | **src-port** }

By default, for a device on which an aggregate interface is created as a L2 aggregate interface by default, load is distributed according to the source MAC addresses and destination MAC addresses of the incoming packets. For a device on which an aggregate interface is created as a L3 aggregate interface by default, load is distributed according to the source IP addresses and destination IP addresses of the incoming packets.

Note

The S6980-64QC, S6920-4C support only the **enhanced profile** keyword.

(4) (Optional) Rename the enhanced load balancing profile.

load-balance-profile *profile-name*

By default, if a device supports enhanced load balancing, the system creates a profile named **reset** for enhanced load balancing. You can use this function to rename the profile or restore the default configurations.

(5) (Optional) Configure the load balancing mode of L2 packets.

l2 field { **dst-mac** | **dst-port** | **l2-protocol** | **src-dst-mac** | **src-mac** | **src-port** | **vlan** } *

When the enhanced load balancing profile configuration is used as the flow balancing mode of aggregate interface, you can perform this configuration according to the traffic feature configuration of the network to implement better load balance.

The load balancing mode of L2 packets is a combination of **src-mac**, **dst-mac**, and **vlan** by default.

(6) (Optional) Configure the load balancing mode of IPv4 packets.

ipv4 field { **dst-ip** | **dst-mac** | **dst-port** | **l2-etype** | **l4-dst-port** | **l4-src-port** | **protocol** | **src-ip** | **src-mac** | **src-port** | **vlan** | **inner-src-ip** | **inner-dst-ip** | **inner-l4-sport** | **inner-l4-dport** | **src-dst-ip** | **src-dst-l4port** } *

Perform this configuration to specify the load balancing mode of IPv4 packets.

The load balancing mode of IPv4 packets is a combination of **src-ip** and **dst-ip** by default.

(7) (Optional) Configure the load balancing mode of IPv6 packets.

ipv6 field { **dst-ip** | **dst-mac** | **dst-port** | **l2-etype** | **l4-dst-port** | **l4-src-port** | **protocol** | **src-ip** | **src-mac** | **src-port** | **vlan** | **inner-src-ip** | **inner-dst-ip** | **inner-l4-sport** | **inner-l4-dport** | **src-dst-ip** | **src-dst-l4port** } *

Perform this configuration to specify the load balancing mode of IPv6 packets.

The load balancing mode of IPv6 packets is a combination of **src-ip** and **dst-ip** by default.

(8) (Optional) Configure the load balancing mode of MPLS packets.

mpls field { [**top-label**] [**2nd-label**] [**3rd-label**] [**src-ip**] [**dst-ip**] [**vlan**] [**src-port**] [**dst-port**] [**src-mac**] [**dst-mac**] [**protocol**] [**l4-src-port**] [**l4-dst-port**] [**l2-etype**] }

Perform this configuration to specify the load balancing mode of MPLS packets.

The load balancing mode of MPLS packets is a combination of **top-label** and **2nd-label** by default.

Note

This command is not supported by the S6920-4C, and S6980-64QC.

(9) (Optional) Configure the load balancing mode for VXLAN packets.

vxlan field { **src-ip** | dst-ip | **l4-src-**port | l**4-**dst-port | **inner-**src-**ip** | **inner-dst-ip** | **inner-l4-sport** | **inner-l4-dport** | **offset** *offset-value* } *

Note

The S6980-64QC supports only the **inner-src-ip** and **inner-dst-ip** keywords.

No load balancing mode is configured for VXLAN packets in the enhanced load balancing profile by default.

To change the load balancing mode of VXLAN packets, perform this configuration in the enhanced profile configuration mode.

Note

This command is supported by only the S6980-64QC.

(10) (Optional) Configure a hash disturbance factor.

hash-disturb { *string* | { [**seed** *seed-value*] [**offset** *offset*] } }

Perform this operation to balance packets of the same type over the aggregate interface for devices of the same type. It is not guaranteed that either configuration affects the balancing effect. If there is no expected effect, the other values can be configured.

The load balancing hash disturbance function is disabled by default.

(11) (Optional) Enable or disable symmetric hashing.

hash-symmetrical { **ipv4** | **ipv6** | **fcoe** }

Perform this operation to ensure that uplink and downlink streams of the same packet type are transmitted over the same path.

Symmetric hashing is disabled by default.

(12) (Optional) Enable or disable flexible hash.

aggregateport hash-elasticity enable

Perform this operation to balance the load of a faulty link over a member port to other active links without affecting original flow on these active links.

The flexible hash balancing function is disabled by default.

Note

This command is supported by only the Only S6520-64CQ, S6510-4C, S6510-32CQ, S6510-48VS8CQ and S6250-48XS8CQ support.

(13) (Optional) Configure the global load balance algorithm mode on an aggregate interface.

aggregateport algorithm mode *algorithm-number*

No load balancing algorithm mode is configured for all link aggregate interfaces by default.

The algorithm mode varies with the product version by default. The actual product prevails.

(14) (Optional) Configure the balancing factor acquisition mode for a tunnel packet.

aggregateport hash-header { **inner** | **outer** }

When configuring load balancing, you can run this command to specify the balancing factor acquisition mode for a specific tunnel packet, achieving the optimal load balancing effect.

No balancing factor acquisition mode is configured for tunnel packets by default.

Note

The S6580-48CQ8QC, S6580-24DC8QC do not support the command.

1.8 Configuring a Capacity Mode

Note

Aggregate interface capacity mode is supported by only the S6980-64QC, S6920-4C, S6580-48CQ8QC, S6580-24DC8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ.

1.8.1 Overview

- Change the maximum number of configurable aggregate interfaces and the maximum number of member ports in each aggregate interface.

1.8.2 Restrictions and Guidelines

- The system has a default aggregate interface capacity mode. You can run **display aggregateport capacity** to display the current capacity mode.
- If the current configuration (maximum number of aggregate interfaces or the number of member ports in each aggregate interface) exceeds the capacity to be configured, the capacity mode configuration will fail.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure an aggregate interface capacity mode.

aggregateport capacity mode *capacity-mode*

Perform this configuration to change the aggregate interface capacity, so as to meet the requirement imposed by changes in the number of aggregate interfaces in network deployment or the number of member ports that can be bound to each aggregate interface.

The capacity mode of an aggregate interface varies with the product version by default. The actual product prevails.

1.9 Enabling BFD for Member Ports

1.9.1 Overview

- Enable BFD for all the member ports of a specified aggregate interface.
- After BFD is enabled for an aggregate interface, each member port performs BFD to determine whether the packets should be distributed to the member port for load balancing. When BFD detects a member port in **Down** state, the packets are not distributed to the port. When BFD detects that the member port is restored to **Up**, the packets are distributed to the port.

1.9.2 Restrictions and Guidelines

- After BFD is enabled for an aggregate interface, BFD sessions are set up. To make the sessions take effect, you need to configure BFD parameters. For details, see *Configuring BFD*.
- Enabling or disabling BFD for a single member port is not supported. You must enable or disable BFD for the entire aggregate interface group.
- Only member ports in forwarding state are enabled with BFD. If a member port is not in forwarding state because the link or LACP is **Down**, the BFD session on the member port is automatically deleted.
- If there is only one member port in forwarding state, BFD is inactive automatically. When the number of member ports reaches 2 or greater, BFD is active again.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure BFD for member ports.

aggregate bfd-detect { **ipv4** *source-ipv4-address destination-ipv4-address* | **ipv6** *source-ipv6-address destination-ipv6-address* }

Enable BFD when you need to detect path failure on member ports in milliseconds. The flow on the faulty member port will be switched to other member port in case of link failure.

The BFD function of a member port is disabled by default.

1.10 Configuring the Minimum Number of Member Ports

1.10.1 Overview

- After the minimum number of member ports is configured, the aggregation group takes effect only when the number of active member ports is greater than the minimum number.
- You can perform this operation to ensure that the link bandwidth provided by an aggregation group is greater than or equal to the sum of bandwidth provided by member ports of aggregate interfaces.
- After an action for the minimum number of member ports is configured, for example, shutting down the aggregate interface, the action is triggered when the number of member ports in **Up** state is less than the minimum number of the member ports.

1.10.2 Restrictions and Guidelines

- If the number of member ports for an LACP interface is less than the minimum number of member ports of the aggregate interface configured, all member ports are in the unbinding state, and the peer end is also automatically unbound.
- After the minimum number of member ports of a static aggregate interface is configured, if the number of member ports of a static aggregate interface in **Up** state is less than the minimum number, the member ports of a static aggregate interface in **Up** state cannot forward data and the corresponding aggregate interface is **Down**. However, the state of the peer end is not affected. Therefore, the corresponding functions need to be configured on the peer end.
- After an action for the minimum number of member ports is configured, the action is triggered only when the number of member ports in **Up** state decreases to an insufficient number. For example, if the minimum number is set to 3, the action is triggered only when the number of member ports decreases from 3 or more to less than 3.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the aggregate interface configuration mode.

interface aggregateport *ap-number*

(4) (Optional) Configure the minimum number of member ports.

aggregateport member minimum { *port-number* | **action shutdown** }

Perform this configuration to specify the minimum number of the member ports of a link aggregation.

The minimum number of member ports of an aggregate interface is 1 by default.

(5) (Optional) Configure the action triggered when the number of member ports is smaller than the minimum value.

aggregateport member minimum { *port-number* | **action shutdown** }

Perform this configuration when some actions need to be triggered when the number of member ports in **Up** state is less than the minimum number of member ports.

By default, no action is configured for the minimum number of member ports in an aggregate interface.

1.11 Enabling the LACP Independent Port Function

1.11.1 Overview

- After the LACP independent port function is enabled, an LACP member port automatically changes to a common physical port if the LACP member port does not receive LACP packets within the set timeout period. The LACP member port state is changed to individual and the LACP member port can forward packets properly.
- After the LACP member port receives LACP packets, it changes to an LACP independent port again to

perform LACP packet negotiation.

- The timeout period of an independent port can be adjusted by configuration.

1.11.2 Restrictions and Guidelines

- After the LACP independent port function is enabled, an LACP member port does not immediately change to a common physical port. Instead, the LACP member port changes to an independent port (a common physical port) only if it does not receive LACP packets within the set timeout period.
- The timeout period configuration of the LACP independent port only affects LACP member ports that have not changed into independent ports. After the timeout period is configured, the period calculation restarts.
- In long timeout mode, the LACP packet is sent every 30s. The timeout period should be longer than 30s so as not to affect the normal LACP negotiation. You are advised to configure the timeout period at least twice the period of LACP packet sending. In short timeout mode, the timeout period is not limited.

1.11.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) (Optional) Enable the LACP independent port function.

lacp individual-port enable

Perform this operation so that an LACP member port can forward packets normally when the LACP member port cannot perform LACP negotiation.

The LACP independent port function is disabled by default.

(5) (Optional) Configure the timeout period of an LACP independent port.

lacp individual-timeout period *time*

Run this command to adjust the timeout period of an LACP independent port.

The timeout period of an LACP independent port is 90s by default.

1.12 Monitoring

This section describes the **display** commands used for checking the running status of a configured function to verify the configuration effect.

Run the **debugging** command to debug the configuration.

⚠ Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging function immediately after use.

You can run the **cleanup** commands to clear information.

Caution

- Running the **cleanup** command during operation of the device may lose vital information and interrupt services.

Table 1-3 Aggregate Interface Monitoring

Command	Purpose
cleanup lacp counters [*key-number* | *interface-type interface-number*]	Clears the statistics of LACP packets on an LACP member port.
display load-balance-profile [*profile-name*]	Displays the configuration of the enhanced load balancing profile.
display lacp summary [*key-number*]	Displays the state of an LACP interface. You can view the information of a specified LACP interface by specifying **key-number**.
display lacp counters [*key-number*]	Displays the statistics of LACP packets on LACP member ports. You can view the information of a specified LACP interface by specifying **key-number**.
display aggregateport [*ap-number*] **summary**	Displays the summary of an aggregate interface.
display aggregateport load-balance	Displays the load balancing information of an aggregate interface.
display aggregateport capacity	Displays the capacity mode and usage of an aggregate interface.
debugging lsm ap	Debugs an aggregate interface.
debugging lacp { **cache-database** | **cli** | **ef-packet** | **thread** | **netconf** | **pkt-agent** | **pkt-statis** | **pkt-thread** | **packet** | **event** | **ha** | **realtime** | **stm** | **timer** | **all** }	Debugs LACP.

1.13 Configuration Examples

1.13.1 Configuring Static Aggregate Interfaces

1. Requirements

Create aggregate interface interconnection for two devices on the IPv4 network, and configure static aggregate interfaces.

2. Topology

Figure 1-2 Configuring Static Aggregate Interfaces



3. Notes

- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device A to static aggregate interface 3.
- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device B to static aggregate interface 3.

4. Procedure

Create aggregate interface 3 on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface list GigabitEthernet 0/1-2
DeviceA(cfg-if-list)# port-group 3
```

Create aggregate interface 3 on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface list GigabitEthernet 0/1-2
DeviceB(cfg-if-list)# port-group 3
```

5. Verification

Run **display aggregateport summary** to check whether aggregate interface 3 contains member ports GigabitEthernet 0/1 and GigabitEthernet 0/2.

```
DeviceA# display aggregateport summary
AggregatePort MaxPorts Switch-Port Mode     Ports
------------- -------- ---------- ------ -----------------------------------
Ag3           8        Enabled    ACCESS Gi0/1,Gi0/2
DeviceB# display aggregateport summary
AggregatePort MaxPorts Switch-Port Mode     Ports
------------- -------- ---------- ------ -----------------------------------
Ag3           8        Enabled    ACCESS Gi0/1,Gi0/2
```

6. Configuration Files

- Device A configuration file

```
!
interface GigabitEthernet 0/1
 port-group 3
!
interface GigabitEthernet 0/2
 port-group 3
```

```
!
interface AggregatePort 3
!
```

- Device B configuration file

```
!
interface GigabitEthernet 0/1
 port-group 3
!
interface GigabitEthernet 0/2
 port-group 3
!
interface AggregatePort 3
!
```

1.13.2 Configuring LACP Interfaces

1. Requirements

Create aggregate interface interconnection for two devices on the IPv4 network, and configure dynamic aggregate interfaces.

2. Topology

Figure 1-3 Configuring LACP Ports



3. Notes

- On Device A, set the LACP system priority to 4096.
- Enable dynamic link aggregation on the GigabitEthernet0/1 and GigabitEthernet0/2 ports on Device A and add the ports to LACP interface 3.
- On Device B, set the LACP system priority to 61440.
- Enable dynamic link aggregation on the GigabitEthernet0/1 and GigabitEthernet0/2 ports on Device B and add the ports to LACP interface 3.

4. Procedure

Configure LACP Ports on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# lacp system-priority 4096
DeviceA(cfg)# interface list GigabitEthernet 0/1-2
DeviceA(cfg-if-list)# port-group 3 mode active
```

```
DeviceA(cfg-if-list)# end
```

Configure LACP Ports on Device B.

```
DevicB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# lacp system-priority 61440
DeviceB(cfg)# interface list GigabitEthernet 0/1-2
DeviceB(cfg-if-list)# port-group 3 mode active
```

5. Verification

Run **display lacp summary 3** to check whether LACP interface 3 contains member ports GigabitEthernet0/1 and GigabitEthernet0/2.

```
DeviceA# display LACP summary 3
System Id:32768,00d0.f8fb.0001
Flags:  S - Device is requesting Slow LACPDUs
  F - Device is requesting Fast LACPDUs.
A - Device is in active mode.          P - Device is in passive mode.
Aggregate port 3:
Local information:
LACP port          Oper      Port        Port
Port       Flags      State       Priority          Key       Number      State
---------------------------------------------------------------------------
Gi0/1      SA         bndl        32768             0x3       0x1         0x3d
Gi0/2      SA         bndl        32768             0x3       0x2         0x3d
Partner information:
                        LACP port                       Oper      Port      Port
Port       Flags      Priority       Dev ID          Key       Number    State
---------------------------------------------------------------------------
Gi0/1      SA         32768       00d0.f800.0002    0x3       0x1       0x3d
Gi0/2      SA         32768       00d0.f800.0002    0x3       0x2       0x3d
DeviceB # display LACP summary 3
SystemId:32768,00d0.f8fb.0002
Flags:  S - Device is requesting Slow LACPDUs
  F - Device is requesting Fast LACPDUs.
A - Device is in active mode.          P - Device is in passive mode.
Aggregate port 3:
Local information:
LACP port          Oper      Port        Port
Port       Flags      State       Priority          Key       Number      State
---------------------------------------------------------------------------
Gi0/1      SA         bndl        32768             0x3       0x1         0x3d
Gi0/2      SA         bndl        32768             0x3       0x2         0x3d
Partner information:
                        LACP port                       Oper      Port      Port
Port       Flags      Priority       Dev ID          Key       Number    State
---------------------------------------------------------------------------
```

Gi0/1	SA	32768	00d0.f800.0001	0x3	0x1	0x3d
Gi0/2	SA	32768	00d0.f800.0001	0x3	0x2	0x3d

6. Configuration Files

- Device A configuration file

```
!
lacp system-priority 4096
!
interface GigabitEthernet 0/1
 port-group 3 mode active
!
interface GigabitEthernet 0/2
 port-group 3 mode active
!
interface AggregatePort 3
!
```

- Device B configuration file

```
!
lacp system-priority 61440
!
interface GigabitEthernet 0/1
 port-group 3 mode active
!
interface GigabitEthernet 0/2
 port-group 3 mode active
!
interface AggregatePort 3
!
```

1.13.3 Enabling LinkTrap

1. Requirements

On the IPv4 network, create aggregate interface interconnection for two devices, configure static aggregate interfaces, and enable LinkTrap.

2. Topology

Figure 1-4 Enabling LinkTrap



3. Notes

- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device A to static aggregate interface 3.

- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device B to static aggregate interface 3.
- On Device A, disable LinkTrap for aggregate interface 3 and enable LinkTrap for its member ports.
- On Device B, disable LinkTrap for aggregate interface 3 and enable LinkTrap for its member ports.

4. Procedure

Perform the following configuration on Device A.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list GigabitEthernet 0/1-2
DeviceA(cfg-if-list)# port-group 3
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# aggregateport member linktrap
DeviceA(cfg)# interface Aggregateport 3
DeviceA(cfg-if-AggregatePort 3)# unset snmp trap link-status
```

Perform the following configuration on Device B:

```
DeviceB# config-system terminal
DeviceB(cfg)# interface list GigabitEthernet 0/1-2
DeviceB(cfg-if-list)# port-group 3
DeviceB(cfg-if-list)# exit
DeviceB(cfg)# aggregateport member linktrap
DeviceB(cfg)# interface Aggregateport 3
DeviceB(cfg-if-AggregatePort 3)# unset snmp trap link-status
```

5. Verification

Run **display current-config** on device A to check whether LinkTrap is enabled for aggregate interface 3 and its member ports.

```
DeviceA# display current-config | include AggregatePort 3
Building configuration...
Current configuration: 54 bytes
interface AggregatePort 3
 unset snmp trap link-status
DeviceA# display current-config | include AggregatePort
aggregateport member linktrap
```

Run **display current-config** on device B to check whether LinkTrap is enabled for aggregate interface 3 and its member ports.

```
DeviceB# display current-config | include AggregatePort 3
Building configuration...
Current configuration: 54 bytes
interface AggregatePort 3
 unset snmp trap link-status
DeviceB# display current-config | include AggregatePort
aggregateport member linktrap
```

6. Configuration Files

- Device A configuration file

```
!
interface GigabitEthernet 0/1
 port-group 3
!
interface GigabitEthernet 0/2
 port-group 3
!
interface AggregatePort 3
 unset snmp trap link-status
!
aggregateport member linktrap
!
```

- Device B configuration file

```
!
interface GigabitEthernet 0/1
 port-group 3 mode active
!
interface GigabitEthernet 0/2
 port-group 3 mode active
!
interface AggregatePort 3
 unset snmp trap link-status
!
aggregateport member linktrap
!
```

1.13.4 Configuring a Load Balancing Mode

1. Requirements

On the IPv4 network, create aggregate interface interconnection for two devices, configure static aggregate interfaces, and enable load balancing.

2. Topology

Figure 1-5 Configuring a Load Balancing Mode



3. Notes

- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device A to static aggregate interface 3.
- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device B to static aggregate interface 3.
- On Device A, configure source MAC address-based load balancing for aggregate interface 3 in global configuration mode.

- On Device B, configure destination MAC address-based load balancing for aggregate interface 3 in global configuration mode.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list GigabitEthernet 0/1-2
DeviceA(cfg-if-list)# port-group 3
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# aggregateport load-balance src-mac
```

Perform the following configuration on Device B:

```
DeviceB# config-system terminal
DeviceB(cfg)# interface list GigabitEthernet 0/1-2
DeviceB(cfg-if-list)# port-group 3
DeviceB(cfg-if-list)# exit
DeviceB(cfg)# aggregateport load-balance dst-mac
```

5. Verification

Run **display aggregateport load-balance** on device A to check the load balancing algorithm configuration.

```
DeviceA      # display aggregatePort load-balance
Load-balance     : Source MAC
```

Run **display aggregateport load-balance** on device B to check the load balancing algorithm configuration.

```
DeviceB      # display aggregatePort load-balance
Load-balance     : Destination MAC
```

6. Configuration Files

- Device A configuration file

```
!
interface GigabitEthernet 0/1
 port-group 3
!
interface GigabitEthernet 0/2
 port-group 3
!
interface AggregatePort 3
!
aggregateport load-balance src-mac
!
```

- Device B configuration file

```
!
interface GigabitEthernet 0/1
 port-group 3
!
interface GigabitEthernet 0/2
```

```
 port-group 3
!
interface AggregatePort 3
!
aggregateport load-balance dst-mac
!
```

7. Common Errors

A user configures the factors **ipv4**, **ipv6**, **fcoe**, and **on** for enabling symmetric hashing. However, no configuration is displayed when the user runs **display current-config**. This is because symmetric hashing is enabled for some products by default. The configuration will be displayed after the user disables the function.

1.13.5 Configuring Hash-based Load Balancing Control

1. Requirements

Create aggregate interface interconnection for two devices on the IPv4 network.

2. Topology

Figure 1-6 Configuring Hash-based Load Balancing Control



3. Notes

- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device A to static aggregate interface 3.
- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device B to static aggregate interface 3.
- On Device A, disable symmetric hashing for FCoE packets.
- On Device B, disable symmetric hashing for FCoE packets.
- On Device A, configure the hash disturbance factor A.
- On Device B, configure the hash disturbance factor B.
- On Device A, enable flexible hash.
- On Device B, enable flexible hash.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list GigabitEthernet 0/1-2
DeviceA(cfg-if-list)# port-group 3
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# load-balance-profile reset
DeviceA(cfg-load-balance-profile)# unset hash-symmetrical fcoe
DeviceA(cfg-load-balance-profile)# hash-disturb A
```

```
DeviceA(cfg-load-balance-profile)# exit
DeviceA(cfg)# aggregateport hash-elasticity enable
```

Perform the following configuration on Device B:

```
DeviceB# config-system terminal
DeviceB(cfg)# interface list GigabitEthernet 0/1-2
DeviceB(cfg-if-list)# port-group 3
DeviceB(cfg-if-list)# exit
DeviceB(cfg)# load-balance-profile reset
DeviceB(cfg-load-balance-profile)# unset hash-symmetrical fcoe
DeviceA(cfg-load-balance-profile)# hash-disturb B
DeviceB(cfg-load-balance-profile)# exit
DeviceB(cfg)# aggregateport hash-elasticity enable
```

5. Verification

Run **display current-config** to check whether the configuration is correct.

6. Configuration Files

- Device A configuration file

```
!
load-balance-profile reset
 unset hash-symmetrical fcoe
 hash-disturb A
!
interface GigabitEthernet 0/1
 port-group 3
!
interface GigabitEthernet 0/2
 port-group 3
!
interface AggregatePort 3
!
aggregateport hash-elasticity enable
!
```

- Device B configuration file

```
!
load-balance-profile reset
 unset hash-symmetrical fcoe
 hash-disturb B
!
interface GigabitEthernet 0/1
 port-group 3
!
interface GigabitEthernet 0/2
 port-group 3
!
```

```
interface AggregatePort 3
!
aggregateport hash-elasticity enable
!
```

7. Common Errors

A user configures the factors **ipv4**, **ipv6**, **fcoe**, and **on** for enabling symmetric hashing. However, no configuration is displayed when the user runs **display current-config**. This is because symmetric hashing is enabled for some products by default. The configuration will be displayed after the user disables the function.

1.13.6 Configuring a Capacity Mode

1. Requirements

Create aggregate interface interconnection for two devices on the IPv4 network.

2. Topology

Figure 1-7 Configuring an Aggregate Interface Capacity Mode



3. Notes

- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device A to static aggregate interface 3.
- Add the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device B to static aggregate interface 3.
- On Device A, configure the 128 x 128 capacity mode.
- On Device B, configure the 256 x 64 capacity mode.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list GigabitEthernet 0/1-2
DeviceA(cfg-if-list)# port-group 3
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# aggregateport capacity mode 128*128
```

Perform the following configuration on Device B:

```
DeviceB# config-system terminal
DeviceB(cfg)# interface list GigabitEthernet 0/1-2
DeviceB(cfg-if-list)# port-group 3
DeviceB(cfg-if-list)# exit
DeviceB(cfg)# aggregateport capacity mode 256*64
```

5. Verification

Run **display aggregateport capacity** to check the aggregate interface capacity mode configuration.

```
DeviceA# display aggregatePort capacity
AggregatePort Capacity Information:
Configuration Capacity Mode: 128*128.
Effective Capacity Mode       : 128*128.
Available Capacity Mode       : 128*128.
Total Number : 128, Used: 1, Available: 127.
DeviceB       DeviceB# display aggregatePort capacity
AggregatePort Capacity Information:
Configuration Capacity Mode: 256*64.
Effective Capacity Mode       : 256*64.
Available Capacity Mode       : 256*64.
Total Number : 256, Used: 1, Available: 255.
```

6. Configuration Files

- Device A configuration file

```
!
aggregateport capacity mode 128*128
!
interface GigabitEthernet 0/1
 port-group 3
!
interface GigabitEthernet 0/2
 port-group 3
!
interface AggregatePort 3
!
aggregateport hash-elasticity enable
!
```

- Device B configuration file

```
!
aggregateport capacity mode 256*64
!
interface GigabitEthernet 0/1
 port-group 3
!
interface GigabitEthernet 0/2
 port-group 3
!
interface AggregatePort 3
!
```

1.13.7 Enabling BFD for Member Ports

1. Requirements

Create aggregate interface interconnection for two devices on the IPv4 network.

2. Topology

Figure 1-8 Enabling BFD for Member Ports



3. Notes

- Enable LACP for the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device A and add the ports to LACP interface 3.
- Enable LACP for the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device B and add the ports to LACP interface 3.
- Configure IP address 1.0.0.1 for aggregate interface 3 on Device A and enable IPv4 BFD.
- Configure IP address 1.0.0.2 for aggregate interface 3 on Device B and enable IPv4 BFD.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list GigabitEthernet 0/1-2
DeviceA(cfg-if-list)# unset switch-port
DeviceA(cfg-if-list)# port-group 3 mode active
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# interface aggregateport 3
DeviceA(cfg-if-Aggregateport 3)# ip address 1.0.0.1 255.255.255.0
DeviceA(cfg-if-Aggregateport 3)# aggregate bfd-detect ipv4 1.0.0.1 1.0.0.2
DeviceA(cfg-if-Aggregateport 3)# bfd interval 50 min_rx 50 multiplier 3
```

Perform the following configuration on Device B:

```
DeviceB# config-system terminal
DeviceB(cfg)# interface list GigabitEthernet 0/1-2
DeviceB(cfg-if-list)# unset switch-port
DeviceB(cfg-if-list)# port-group 3 mode active
DeviceB(cfg-if-list)# exit
DeviceB(cfg)# interface aggregateport 3
DeviceB(cfg-if-Aggregateport 3)# ip address 1.0.0.2 255.255.255.0
DeviceB(cfg-if-Aggregateport 3)# aggregate bfd-detect ipv4 1.0.0.2 1.0.0.1
DeviceB(cfg-if-Aggregateport 3)# bfd interval 50 min_rx 50 multiplier 3
```

5. Verification

Run **display current-config** to check whether the configuration is correct.

Run **display interface aggregateport** to view the BFD state of each member port.

```
DeviceA# display current-config | include AggregatePort 3
Building configuration…
```

```
Current configuration: 54 bytes
interface AggregatePort 3
 unset switch-port
 ip address 1.0.0.1 255.255.255.0
 aggregate bfd-detect ipv4 1.0.0.1 1.0.0.2
 bfd interval 50 min_rx 50 multiplier 3
DeviceA# display interface aggregateport 3
…
Aggregate Port Informations:
        Aggregate Number: 3
        Nam“: “AggregatePor” 3"
        Members: (count=2)
        GigabitEthernet 0/1     Link Status: Up     Lacp Status: bndl     BFD Status: UP
        GigabitEthernet 0/2     Link Status: Up     Lacp Status: bndl     BFD Status: UP
```

```
DeviceB# display current-config | include AggregatePort 3
Building configuration...
Current configuration: 54 bytes
interface AggregatePort 3
 unset switch-port
 ip address 1.0.0.2 255.255.255.0
 aggregate bfd-detect ipv4 1.0.0.2 1.0.0.1
 bfd interval 50 min_rx 50 multiplier 3
DeviceB# display interface aggregateport 3
…
Aggregate Port Informations:
        Aggregate Number: 3
        Nam“: "AggregatePor” 3"
        Members: (count=2)
        GigabitEthernet 0/1     Link Status: Up     Lacp Status: bndl     BFD Status: UP
        GigabitEthernet 0/2     Link Status: Up     Lacp Status: bndl     BFD Status: UP
```

6. Configuration Files

- Device A configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 port-group 3 mode active
!
interface GigabitEthernet 0/2
 unset switch-port
 port-group 3 mode active
!
interface AggregatePort 3
 unset switch-port
```

```
 aggregate bfd-detect ipv4 1.0.0.1 1.0.0.2
 ip address 1.0.0.1 255.255.255.0
 bfd interval 50 min_rx 50 multiplier 3
!
```

- Device B configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 port-group 3 mode active
!
interface GigabitEthernet 0/2
 unset switch-port
 port-group 3 mode active
!
interface AggregatePort 3
 unset switch-port
 aggregate bfd-detect ipv4 1.0.0.2 1.0.0.1
 ip address 1.0.0.2 255.255.255.0
 bfd interval 50 min_rx 50 multiplier 3
!
```

7. Common Errors

- If BFD is enabled for an aggregate interface without BFD parameters, BFD does not take effect.
- After BFD is enabled for an aggregate interface, the BFD neighbor must be a directly connected aggregate interface enabled with BFD.

1.13.8 Configuring the Minimum Number of Member Ports (When the Number of Member Ports Is Less Than the Minimum Number)

1. Requirements

On the IPv4 network, create aggregate interface interconnection for two devices, and configure the number of member ports less than the minimum number of member ports.

2. Topology

Figure 1-9 Topology for Configuring the Minimum Number of Member Ports (When the Number of Member Port is Less Than the Minimum Number)



3. Notes

- Enable LACP for the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on Device A and add the ports to LACP interface 3.

- Enable LACP for the GigabitEthernet 0/1 and GigabitEthernet0/2 ports on Device B and add the ports to LACP interface 3.
- On Device A, set the minimum number of the member ports of aggregate interface 3 to 3.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list GigabitEthernet 0/1-2
DeviceA(cfg-if-list)# unset switch-port
DeviceA(cfg-if-list)# port-group 3 mode active
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# interface aggregateport 3
DeviceA(cfg-if-Aggregateport 3)# aggregateport member minimum 3
```

Perform the following configuration on Device B:

```
DeviceB# config-system terminal
DeviceB(cfg)# interface list GigabitEthernet 0/1-2
DeviceB(cfg-if-list)# unset switch-port
DeviceB(cfg-if-list)# port-group 3 mode active
DeviceB(cfg-if-list)# exit
DeviceB(cfg)# interface aggregateport 3
DeviceB(cfg-if-Aggregateport 3)# aggregateport member minimum 3
```

5. Verification

Run **display current-config** to check whether the configuration is correct.

Run **display lacp summery** to view the aggregation state of each member port.

```
DeviceA# display LACP summary 3
System Id:32768, 00d0.f8fb.0001
Flags: –S - Device is requesting Slow LACPDUs
–F - Device is requesting Fast LACPDUs.–A - Device is in active mode.        –P - Device is in passive mode.
Aggregate port 3:
Local information:
LACP port          Oper      Port        Port
Port       Flags      State      Priority          Key       Number      Sta--
Gi0/1      SA           susp          32768          0x3       0x1         0x3d
Gi0/2      SA           susp          32768          0x3       0x2         0x3d
Partner information:
                    LACP port                    Oper      Port      Port
Port       Flags      Priority      Dev ID          Key       Number    Sta--
Gi0/1      SA           32768       00d0.f800.0002  0x3       0x1       0x3d
Gi0/2      SA           32768       00d0.f800.0002  0x3       0x2       0x3d
```

6. Configuration Files

- Configuration file of device A

```
!
```

```
interface GigabitEthernet 0/1
 unset switch-port
 port-group 3 mode active
!
interface GigabitEthernet 0/2
 unset switch-port
 port-group 3 mode active
!
interface AggregatePort 3
 unset switch-port
 aggregateport member minimum 3
!
```

- Configuration file of Device B

```
!
interface GigabitEthernet 0/1
 unset switch-port
 port-group 3 mode active
!
interface GigabitEthernet 0/2
 unset switch-port
 port-group 3 mode active
!
interface AggregatePort 3
 unset switch-port
 aggregateport member minimum 3
!
```

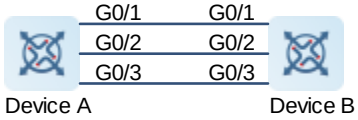
1.13.9 Configuring the Minimum Number of Member Ports (When the Number of Member Ports Is Not Less Than the Minimum Number)

1. Requirements

On the IPv4 network, create aggregate interface interconnection for two devices, and configure the number of member ports not less than the minimum number of member ports.

2. Topology

Figure 1-10 Configuring the Minimum Number of Member Ports (When the Number of Member Port is Not Less Than the Minimum Number)



3. Notes

- Enable LACP for the GigabitEthernet 0/1, GigabitEthernet 0/2, and GigabitEthernet 0/3 ports on Device A and add the ports to LACP interface 3.

- Enable LACP for the GigabitEthernet 0/1, GigabitEthernet 0/2, and GigabitEthernet 0/3 ports on Device B and add the ports to LACP interface 3.
- On Device A, set the minimum number of member ports of aggregate interface 3 to 2.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list GigabitEthernet 0/1-3
DeviceA(cfg-if-list)# unset switch-port
DeviceA(cfg-if-list)# port-group 3 mode active
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# interface aggregateport 3
DeviceA(cfg-if-Aggregateport 3)# aggregateport member minimum 2
```

Perform the following configuration on Device B:

```
DeviceB# config-system terminal
DeviceB(cfg)# interface list GigabitEthernet 0/1-3
DeviceB(cfg-if-list)# unset switch-port
DeviceB(cfg-if-list)# port-group 3 mode active
DeviceB(cfg-if-list)# exit
DeviceB(cfg)# interface aggregateport 3
DeviceB(cfg-if-Aggregateport 3)# aggregateport member minimum 2
```

5. Verification

Run **display current-config** to check whether the configuration is correct.

Run **display lacp summery** to view the aggregation state of each member port.

```
DeviceA# display LACP summary 3
System Id:32768,00d0.f8fb.0001
Flags: –S - Device is requesting Slow LACPDUs
–F - Device is requesting Fast LACPDUs.–A - Device is in active mode.        –P - Device is in passive mode.
Aggregate port 3:
Local information:
LACP port          Oper      Port        Port
Port       Flags      State      Priority          Key       Number      Sta--
Gi0/1      SA           bndl          32768           0x3       0x1         0x3d
Gi0/2      SA           bndl          32768           0x3       0x2         0x3d
Gi0/3      SA           bndl          32768           0x3       0x3         0x3d

Partner information:
                    LACP port                        Oper      Port      Port
Port       Flags      Priority       Dev ID          Key       Number    Sta--
Gi0/1      SA           32768        00d0.f800.0002  0x3       0x1       0x3d
Gi0/2      SA           32768        00d0.f800.0002  0x3       0x2       0x3d
Gi0/3      SA           32768        00d0.f800.0002  0x3       0x3       0x3d
```

6. Configuration Files

- Configuration file of device A

```
!
interface GigabitEthernet 0/1
 unset switch-port
 port-group 3 mode active
!
interface GigabitEthernet 0/2
 unset switch-port
 port-group 3 mode active
!
interface GigabitEthernet 0/3
 unset switch-port
 port-group 3 mode active
!
interface AggregatePort 3
 unset switch-port
 aggregateport member minimum 2
!
```

- Configuration file of device A

```
!
interface GigabitEthernet 0/1
 unset switch-port
 port-group 3 mode active
!
interface GigabitEthernet 0/2
 unset switch-port
 port-group 3 mode active
!
interface GigabitEthernet 0/3
 unset switch-port
 port-group 3 mode active
!
interface AggregatePort 3
 unset switch-port
 aggregateport member minimum 2
!
```

1.13.10 Enabling the LACP Independent Port Function

1. Requirements

The server uses NIC 1 and NIC 2 as the communication ports to get access to the GigabitEthernet0/1 and GigabitEthernet0/2 of the access device. GigabitEthernet0/1 and GigabitEthernet0/2 are added to the LACP interface, for example, aggregate interface 3. A specific VLAN, for example, VLAN 10 is allocated. The LACP independent port function is enabled for GigabitEthernet0/1 and GigabitEthernet0/2. When the OS is not

installed on the server, LACP negotiation between the server and the access device fails. In this case, GigabitEthernet0/1 and GigabitEthernet0/2 of the access device change to common physical ports and are allocated to VLAN 10 automatically. The server uses NIC 1 or NIC 2 to communicate with the remote OS installation device. After the OS is installed, the server connects to the access device in LACP mode.

2. Topology

Figure 1-11 Enabling the LACP Independent Port Function



3. Notes

- Enable LACP for the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on the access device and add the ports to LACP interface 3.
- Enable the LACP independent port function for the GigabitEthernet 0/1 and GigabitEthernet 0/2 ports on the access device.
- Allocate aggregate interface 3 on the access device to VLAN 10.

4. Procedure

Create an aggregate interface on Device A.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list GigabitEthernet 0/1-2
DeviceA(cfg-if-list)# port-group 3 mode active
DeviceA(cfg-if-list)# lacp individual-port enable
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# interface aggregateport 3
DeviceA(cfg-if-Aggregateport 3)# switch access vlan 10
```

5. Verification

Run **display current-config** to check whether the configuration is correct.

Run **display lacp summery** to view the aggregation state of each member port.

```
DeviceA# display LACP summary 3
System Id:32768,00d0.f8fb.0001
Flags: –S - Device is requesting Slow LACPDUs
–F - Device is requesting Fast LACPDUs.–A - Device is in active mode.        –P - Device is in passive mode.
Aggregate port 3:
Local information:
LACP port          Oper      Port        Port
Port      Flags    State     Priority    Key      Number    Sta--
Gi0/1     SA       individual  32768       0x3      0x1       0x3d
Gi0/2     SA       individual  32768       0x3      0x2       0x3d
Partner information:
```

Port	Flags	LACP port Priority	Dev ID	Oper Key	Port Number	Port Sta--
Gi0/1	SA	32768	00d0.f800.0002	0x3	0x1	0x3d
Gi0/2	SA	32768	00d0.f800.0002	0x3	0x2	0x3d

6. Configuration Files

- Device A configuration file

```
!
interface GigabitEthernet 0/1
 port-group 3 mode active
 lacp individual-port enable
!
interface GigabitEthernet 0/2
 port-group 3 mode active
 lacp individual-port enable
!
interface AggregatePort 3
 switch-port access vlan 10
!
```

Ethernet Switching Configuration

1. Configuring MAC Address
2. Configuring MAC Loopback
3. Configuring VLAN
4. Configuring MAC VLAN
5. Configuring Protocol VLAN
6. Configuring Private VLAN
7. Configuring Super VLAN
8. Configuring GVRP
9. Configuring QinQ
10. Configuring MSTP
11. Configuring LLDP

Contents

1 Configuring MAC Address....1
1.1 Introduction....1
1.1.1 MAC Address Table....1
1.1.2 Classification of MAC Address Entries....1
1.1.3 Generation of MAC Address Entries....1
1.1.4 Update and Aging of MAC Address Entries....3
1.1.5 Protocols and Standards....3
1.2 Restrictions and Guidelines....4
1.3 Configuration Task Summary....4
1.4 Configuring Static MAC Address Entries....4
1.4.1 Overview....4
1.4.2 Restrictions and Guidelines....4
1.4.3 Procedure....4
1.5 Configuring Filtering MAC Address Entries....5
1.5.1 Overview....5
1.5.2 Procedure....5
1.6 Disabling Dynamic MAC Address Learning....5
1.6.1 Overview....5
1.6.2 Disabling MAC Address Learning in Global Configuration Mode....5
1.6.3 Disabling MAC Address Learning for an Interface....6
1.7 Configuring Upper Limit of Learned MAC Addresses....6
1.7.1 Overview....6

1.7.2 Configuring Upper Limit of MAC Addresses Learned from a VLAN....7
1.7.3 Configuring Upper Limit of MAC Addresses Learned from an Interface....7
1.8 Configuring Aging Time of Dynamic MAC Address Entries....8
1.9 Enabling MAC Address Flapping Detection and Protection....8
1.9.1 Overview....8
1.9.2 Procedure....9
1.10 Configuring Reporting Interval of MAC Address Table Usage Alarms and MAC Address Table Usage Threshold....9
1.10.1 Overview....9
1.10.2 Procedure....9
1.11 Configuring the Function of MAC Address Entry Change Notification....10
1.11.1 Overview....10
1.11.2 Restrictions and Guidelines....11
1.11.3 Procedure....11
1.12 Monitoring....12
1.13 Configuration Examples....13
1.13.1 Configuring MAC Address Flapping Detection....13
1.13.2 Configuring the Function of MAC Address Change Notification....14

1 Configuring MAC Address

1.1 Introduction

A Media Access Control (MAC) address is used to identify the position of a device in a network. Generally, a MAC address consists of 12 hexadecimals and contains a total of 48 bits (six bytes). The first 24 bits are applied for by vendors from The Internet Engineering Task Force (IETF) and used to identify a network device manufacturer. The last 24 bits are assigned by the vendors and used to identify network cards produced by the manufacturers.

1.1.1 MAC Address Table

A MAC address table records mappings of MAC addresses and interfaces to virtual local area networks (VLANs). A device queries the MAC address table based on the destination MAC address in a received packet. If the device finds an entry that is consistent with the destination MAC Address in the packet, the device forwards the packet through the interface corresponding to the entry in unicast mode. If the device does not find such an entry, it forwards the packet through all interfaces other than the receiving interface in broadcast mode.

Note

This section describes static, dynamic, and filtering MAC address entries. For more information about multicast MAC address entries and specific configurations, see *Configuring IGMP*.

1.1.2 Classification of MAC Address Entries

MAC address entries are classified into the following types:

- Static MAC address entries: Manually configured by the administrator. Packets whose destination MAC address matches the one in such an entry are forwarded through the correct interface. This type of entries does not age.
- Dynamic MAC address entries: Automatically generated by devices. Packets whose destination MAC address matches the one in such an entry are forwarded through the correct interface. This type of entries ages.
- Filtering MAC address entries: Manually configured by the administrator. Packets whose source or destination MAC address matches the one in such an entry are discarded. This type of entries does not age.

The priorities of these types of MAC address entries are ranked in descending order as follows: Filtering MAC address entries > Static MAC address entries > Dynamic MAC address entries. If two entries have the same MAC address and VLAN, the higher-priority entry can overwrite the lower-priority one, and entries with the same priority can overwrite each other.

1.1.3 Generation of MAC Address Entries

A MAC address entry can be automatically learned or manually configured.

1. Automatic learning

A device automatically generates a MAC address entry based on the source MAC address in a received packet. The learning procedure is as follows:

(1) Initially, the MAC address table of Device A is empty. To send a packet to User C, User A sends the packet to the interface GigabitEthernet 0/1 of Device A. Upon receiving the packet, Device A generates a MAC address entry, with the MAC address being MAC A and the interface being GigabitEthernet 0/1.

(2) As Device A does not have an entry corresponding to MAC A, it forwards the packet through all interfaces other than GigabitEthernet 0/1 in broadcast mode. In this case, other users also receive the packet that User A sends to User C.

(3) User C sends a reply packet destined for User A to Device A through the interface GigabitEthernet 0/3. Upon receiving the packet, Device A generates a MAC address entry, with the MAC address being MAC C and the interface being GigabitEthernet 0/3. Meanwhile, Device A finds an entry corresponding to MAC A and then forwards the packet to User A through GigabitEthernet 0/1 in unicast mode.

(4) After the communication between User A and User C is complete, Device A learns two MAC address entries. Later, packets are forwarded between User A and User C in unicast mode and other users do not receive packets destined for these two users.

Figure 1-1 Automatically Learning MAC Address Entry I

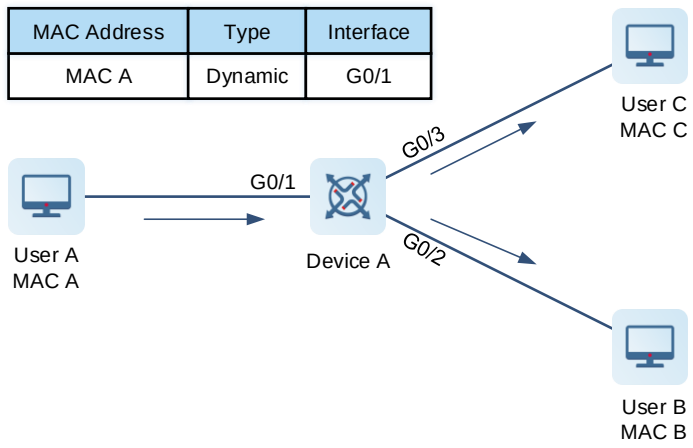
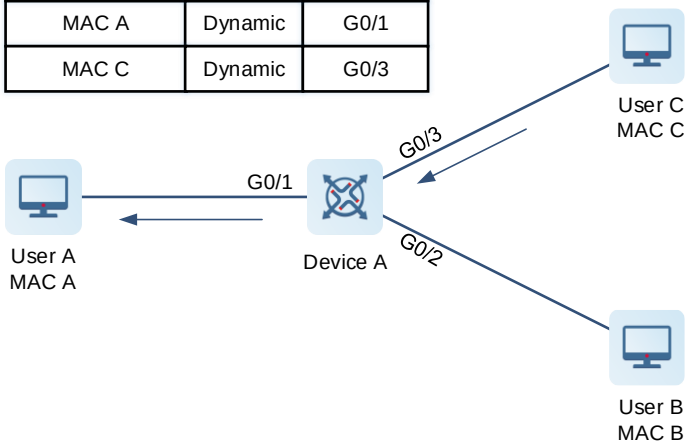


Figure 1-2 Automatically Learning MAC Address Entry II

MAC Address	Type	Interface
MAC A	Dynamic	G0/1
MAC C	Dynamic	G0/3



2. Manual configuration

If the device forwards packets based on automatically generated MAC address entries, security risks exist. For example, if an attacker forges a packet by using a MAC address of a valid user and sends the packet to another interface of the device, the device automatically updates the corresponding MAC address entry and sends subsequent packets destined for the user to the attacker. To improve network security, the network administrator can add a specified MAC address entry to bind the MAC address of the valid user to a device interface. The static MAC address entry configured in this way is not updated due to packet forwarding. If a packet whose source MAC address is the MAC address of the valid user is received from another user, the device directly discards this packet.

1.1.4 Update and Aging of MAC Address Entries

Static and filtering MAC address entries do not automatically update or age. Dynamic MAC address entries are updated and age in the following scenarios:

(1) When the device learns a new dynamic MAC address entry, the device sets the age flag bit of this entry to 1.

(2) If the source MAC address in a received packet already exists in the MAC address table, the device updates the entry and sets the age flag bit to 1.

(3) The device checks the age flag bit of a dynamic MAC address entry at the interval of the aging time. For entries whose flag bit is 1, the device sets the flag bit to 0. For entries whose flag bit is 0, the device deletes the entries.

Therefore, the actual aging time of a dynamic MAC address entry ranges between the specified age value and two times the age value.

1.1.5 Protocols and Standards

- IEEE 802.3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications
- IEEE 802.1Q: Virtual Bridged Local Area Networks

1.2 Restrictions and Guidelines

If a Layer 2 interface has subinterfaces, running the **reset interface** command in global configuration mode does not affect configurations related to the MAC address learning function, the upper limit of learned MAC addresses, the forwarding rule to be used after the upper limit of learned MAC addresses is reached, and the MAC address flapping protection policy.

1.3 Configuration Task Summary

All the following configuration tasks are optional and may be selected as needed.

- Configuring Static MAC Address Entries
- Configuring Filtering MAC Address Entries
- Disabling Dynamic MAC Address Learning
  - Disabling MAC Address Learning in Global Configuration Mode
  - Disabling MAC Address Learning for an Interface
- Configuring Upper Limit of Learned MAC Addresses
  - Configuring Upper Limit of MAC Addresses Learned from a VLAN
  - Configuring Upper Limit of MAC Addresses Learned from an Interface
- Configuring Aging Time of Dynamic MAC Address Entries
- Enabling MAC Address Flapping Detection and Protection
- Configuring Reporting Interval of MAC Address Table Usage Alarms and MAC Address Table Usage Threshold
- Configuring the Function of MAC Address Entry Change Notification

1.4 Configuring Static MAC Address Entries

1.4.1 Overview

You can manually configure static MAC address entries to bind MAC addresses to device interfaces.

1.4.2 Restrictions and Guidelines

- Interfaces in the MAC address entries can be specified as Layer 2 Ethernet interfaces or Layer 2 aggregation interfaces.
- If a MAC address is bound to different interfaces in the same VLAN, the last configuration prevails.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static MAC address.

mac-addr-table static *mac-address* **vlan** *vlan-id* **interface** *interface-type interface-number*

No static MAC address entry exists by default.

1.5 Configuring Filtering MAC Address Entries

1.5.1 Overview

To prohibit a user from sending and receiving packets in certain scenarios, you can add the MAC address of the user to a filtering MAC address entry. After the entry is configured, packets whose source or destination MAC address matches the MAC address in the filtering MAC address entry are directly discarded.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a filtering MAC address.

mac-addr-table filtering *mac-address* **vlan-id** *vlan-id*

No filtering MAC address entry exists by default.

1.6 Disabling Dynamic MAC Address Learning

1.6.1 Overview

MAC address learning is enabled by default. A device automatically generates a dynamic MAC address entry based on the source MAC address in a received packet. The device queries the MAC address table to perform unicast forwarding, which reduces broadcast frequency and therefore reduces the network traffic. The device learns or updates the MAC address entry each time it receives a packet, which results in low security. An attacker may forge and send a large number of packets with different source MAC addresses to the device to use up the MAC address entry resources. Consequently, the device cannot update the MAC address entries in time. You can disable dynamic MAC address learning if required.

1.6.2 Disabling MAC Address Learning in Global Configuration Mode

1. Overview

If you disable MAC address learning in global configuration mode, the device does not learn MAC addresses.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable MAC address learning in global configuration mode.

mac-address-learning disable

MAC address learning is enabled in global configuration mode by default.

1.6.3 Disabling MAC Address Learning for an Interface

1. Overview

If MAC address learning is enabled in global configuration mode, you can disable this function for a specific interface.

2. Restrictions and Guidelines

- If 802.1X, IP Source Guard, or the port security function is configured on an interface, MAC address learning must be disabled for this interface. For more information about 802.1X, IP Source Guard, or the port security function, see *Configuring IEEE 802.1X*, *Configuring IP Source Guard*, or *Configuring Port Security*.
- This function can be enabled on Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces only.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(1) Enter the global configuration mode.

config-system terminal

(2) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

(3) Disable MAC address learning for an interface.

unset mac-address-learning

No MAC address learning is enabled for an interface by default.

1.7 Configuring Upper Limit of Learned MAC Addresses

1.7.1 Overview

The number of MAC address entries on the device is limited. Packets can be forwarded only in broadcast mode in order to prevent the device from learning a large number of dynamic MAC address entries from an interface or a VLAN (for example, when it is attacked by packets with different source addresses) and prevent MAC address learning failures on other interfaces and VLANs. In this case, you can configure the upper limit of MAC addresses learned from each interface or VLAN based on actual network plans. When the number of MAC addresses reaches the upper limit, the device stops learning MAC addresses from the corresponding interface or VLAN. You can also configure a forwarding rule to be used after the number of learned MAC addresses reaches the upper limit, to determine whether to forward packets whose source MAC address is not in the MAC address table. With this function, you can save MAC address entry resources and enhance network security.

This function is available only to dynamic MAC address entries.

1.7.2 Configuring Upper Limit of MAC Addresses Learned from a VLAN

1. Overview

You can configure the upper limit of MAC addresses learned from a VLAN and the packet forwarding rule to be used after the number of learned MAC addresses reaches the upper limit.

2. Restrictions and Guidelines

- If the number of MAC addresses learned from a VLAN is greater than the upper limit, the device stops learning MAC addresses from the VLAN, and starts learning again only after the number drops below the upper limit as aged address entries are deleted.
- This function does not limit replication of MAC address entries to a VLAN whose number of learned MAC address entries reaches the upper limit.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a VLAN and enter the VLAN configuration mode.

vlan *vlan-id*

(4) Configure the upper limit of dynamic MAC addresses learned from a VLAN.

max-dynamic-mac-count *count*

No upper limit of learned MAC addresses is configured for a VLAN by default.

(5) (Optional) Configure a forwarding rule to be used after the number of dynamic MAC addresses learned from a VLAN reaches the upper limit.

max-dynamic-mac-count exceed-action { **forward** | **discard** }

By default, after the number of MAC addresses learned from a VLAN reaches the upper limit, the device continues forwarding packets whose source MAC address is not learned.

1.7.3 Configuring Upper Limit of MAC Addresses Learned from an Interface

1. Overview

You can configure the upper limit of MAC addresses learned from an interface and the packet forwarding rule to be used after the number of learned MAC addresses reaches the upper limit.

2. Restrictions and Guidelines

If the number of MAC addresses learned from an interface is greater than the upper limit, the device stops learning MAC addresses from the interface, and starts learning again only after the number drops below the upper limit as aged MAC address entries are deleted.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

(4) Configure the upper limit of dynamic MAC addresses learned from an interface.

max-dynamic-mac-count *count*

No upper limit of learned MAC addresses is configured for an interface by default.

(5) (Optional) Configure a forwarding rule to be used after the number of dynamic MAC addresses learned from an interface reaches the upper limit.

max-dynamic-mac-count exceed-action { **forward** | **discard** }

By default, after the number of MAC addresses learned from an interface reaches the upper limit, the device continues forwarding packets whose source MAC address is not learned.

1.8 Configuring Aging Time of Dynamic MAC Address Entries

1. Overview

The device deletes useless dynamic MAC address entries based on the aging time to save entry resources on the device. An overly long aging time may lead to untimely deletion of useless entries, whereas an overly short aging time may lead to deletion of some valid entries and repeated learning of MAC addresses by the device, which increases the packet broadcast frequency. Therefore, you are advised to configure a proper aging time of dynamic MAC address entries as required to save device resources without affecting network stability.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the aging time of dynamic MAC addresses.

mac-addr-table aging-time *time*

The aging time of dynamic MAC addresses is 300s by default.

1.9 Enabling MAC Address Flapping Detection and Protection

1.9.1 Overview

If a MAC address is learned from different interfaces in the same VLAN, MAC address flapping occurs. If MAC address flapping occurs frequently between two specific interfaces, a network loop is formed. In this case, packets with the same source MAC address are received from different interfaces, and the MAC address entry

is updated frequently. By enabling MAC address flapping detection, you can effectively monitor MAC address flapping events on the Layer 2 network. Each time a MAC address flapping event occurs, the device reports a log alarm message. You can configure the MAC address flapping protection policy on the interfaces. When the device detects that MAC address flapping occurs between two interfaces with different priorities, the device reports an alarm message and disables the lower-priority interface.

1.9.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure MAC address flapping detection.

mac-addr-table flapping-logging

MAC address flapping detection is disabled by default.

(4) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

(5) (Optional) Configure a MAC address flapping protection policy.

mac-addr-table flapping action { **error-down** | **priority** *priority* }

No MAC address flapping protection policy is configured by default.

The MAC address flapping protection policy is available only for an interface with MAC address flapping detection enabled.

1.10 Configuring Reporting Interval of MAC Address Table Usage Alarms and MAC Address Table Usage Threshold

1.10.1 Overview

When the usage of a MAC address table of a device exceeds the upper or lower limit, the device reports an alarm message. With this function, you can adjust the reporting interval of MAC address table usage alarms and MAC address table usage threshold (in percentage).

When the usage of a MAC address table is higher than the upper limit, the device reports an alarm message at the specified interval. After the usage drops below the upper limit, the device reports a restoration message and stops reporting the alarm message. When the usage of a MAC address table is lower than the lower limit, the device reports an alarm message at the specified interval. After the usage rises above the lower limit, the device reports a restoration message and stops reporting the alarm message.

1.10.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a reporting interval of the MAC address table usage alarms.

mac-addr-table warning-interval *interval*

The reporting interval of the MAC address table usage alarms is 3600s by default.

If you set the interval to 0s, MAC address table usage alarm reporting is disabled.

(4) Configure the upper and lower limits of the MAC address table usage.

mac-addr-table warning-threshold upper-limit *upper-limit-threshold* **lower-limit** *lower-limit-threshold*

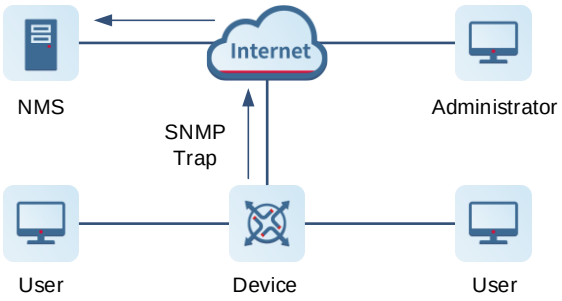
By default, the upper and lower limits for reporting MAC address table usage alarms are 80% and 70%, respectively.

1.11 Configuring the Function of MAC Address Entry Change Notification

1.11.1 Overview

After the function of MAC address entry change notification is configured on a device, the device generates a notification message for a MAC address entry change when the device learns a new MAC address or has aged a learned MAC address. In addition, the device sends the notification message in SNMP Trap mode to a specified Network Management Station (NMS). This function helps the network administrator monitor user change of network devices.

Figure 1-3 Application Scenario of the Function of MAC Address Entry Change Notification



As shown in Figure 1-3, when the device generates a notification for an added MAC address, a new user identified by the MAC address starts to use the network. When the device generates a notification for a deleted MAC address, the corresponding user does not send packets in the address aging time. In this case, it is considered that the user stops using the network. When the device is accessed by a large number of users, the device may generate a large number of MAC address entry change notifications in a short time, and the traffic increases.

To reduce traffic, you can configure an interval for sending MAC address entry change notifications. When the configured interval arrives, the system sends all notifications in a period of time at a time. This helps reduce the network traffic. Each notification message contains information about multiple MAC address changes.

A generated notification is stored in a history table for MAC address entry change notifications. The administrator may know recent MAC address changes by checking the history table even without NMS.

1.11.2 Restrictions and Guidelines

A MAC address entry change notification is generated only for a dynamic MAC address entry.

1.11.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an NMS host address for receiving MAC address entry change notifications.

snmp-agent host { *ipv4-address* | **ipv6** *ipv6-address* } **traps** [**version** { **1** | **2c** | **3** [**auth** | **noauth** | **priv**] }] *community-string*

No NMS host address is configured for receiving MAC address entry change notifications by default.

(4) Enable sending of MAC address entry change notifications in Trap messages.

snmp-agent enable traps

Sending Trap messages to the NMS host is forbidden by default.

(5) Enable the function of MAC address entry change notification in global configuration mode.

mac-addr-table notification

The function of MAC address entry change notification is disabled in global configuration mode by default.

(6) (Optional) Configure an interval for sending MAC address entry change notifications.

mac-addr-table notification interval *interval*

The device sends MAC address entry change notifications at an interval of 1s by default.

(7) (Optional) Configure the size of a history table for MAC address entry change notifications.

mac-addr-table notification history-size *size*

The size of the history table for MAC address entry change notifications is 50 by default.

(8) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

(9) Enable the function of MAC address entry change notification for an interface.

snmp-agent trap-enable mac-notification { **added** | **removed** }

The function of MAC address entry change notification is disabled for an interface by default.

1.12 Monitoring

This section describes the **display** commands used for checking the running status of a configured function to verify the configuration effect.

You can run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup mac-addr-table dynamic [**address** *mac-address*] [**interface** *interface-type interface-number*] [**vlan-id** *vlan-id*]	Clears the dynamic MAC address entries.
cleanup mac-addr-table flapping record	Clears the MAC address flapping records.
display mac-addr-table { **dynamic** | **static** | **filter** } [**address** *mac-address*] [**vlan-id** *vlan-id*] [**interface** *interface-type interface-number*]	Displays the MAC address table.
display mac-addr-table { **evpn** | **mlag** }	Displays the EVPN or MLAG address table.
display mac-addr-table address *mac-address* [**vsi** *vsi-id*] [**vni** *vni-id*]	Displays the specific EVPN or MAC address table.
display mac-addr-table [**vsi** *vsi-id*] [**vni** *vni-id*]	Displays the specific VNI or VSI address table.
display mac-addr-table all	Displays all types of addresses.
display mac-addr-table aging-time	Displays the aging time for dynamic MAC addresses.
display mac-addr-table max-dynamic-mac-count	Displays the maximum number of dynamic MAC addresses.
display mac-addr-table notification [**interface** [*interface-type interface-number*] **| history**]	Displays the configurations and history table for MAC address entry change notifications.
display mac-addr-table flapping record	Displays the dynamic MAC address flapping records.

1.13 Configuration Examples

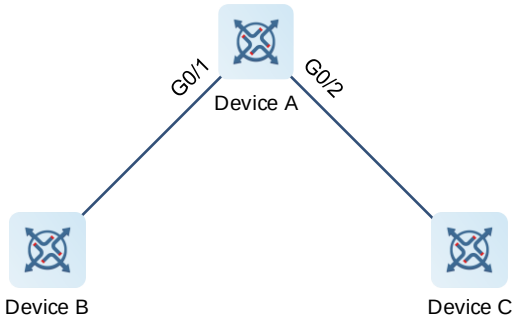
1.13.1 Configuring MAC Address Flapping Detection

1. Requirements

As shown in Figure 1-4, a Layer 2 loop may occur between Device B and Device C due to misoperation. After you enable MAC address flapping detection and configure a protection policy, the device can report alarms upon loops, and disable the faulty interface based on the protection policy.

2. Topology

Figure 1-4 Configuring MAC Address Flapping Detection



3. Notes

- Enable MAC address flapping detection on Device A.
- Configure a MAC address flapping protection policy on the interfaces of Device A.
- Set the priority of GigabitEthernet 0/1 to be higher than that of GigabitEthernet 0/2 on Device A.

4. Procedure

Enable MAC address flapping detection.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# mac-addr-table flapping-logging
```

Configure a protection policy to disable GigabitEthernet 0/1 upon MAC address flapping and set the priority of this interface to **5**.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg)# (cfg-if-GigabitEthernet 0/1)# mac-addr-table flapping action error-down
DeviceA(cfg)# (cfg-if-GigabitEthernet 0/1)# mac-addr-table flapping action priority 5
DeviceA(cfg)# (cfg-if-GigabitEthernet 0/1)# exit
```

Configure a protection policy to disable GigabitEthernet 0/2 upon MAC address flapping and set the priority of this interface to **1**.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg)# (cfg-if-GigabitEthernet 0/2)# mac-addr-table flapping action error-down
DeviceA(cfg)# (cfg-if-GigabitEthernet 0/2)# mac-addr-table flapping action priority 1
```

5. Verification

After the preceding configuration, the device generates an alarm message when MAC address flapping from GigabitEthernet 0/1 to GigabitEthernet 0/2 occurs. Run the **display mac-addr-table flapping record** command to display the dynamic MAC address flapping records.

View the records on Device A after MAC address flapping is detected.

```
DeviceA# display mac-addr-table flapping record
 Mac address flapping detect status           : on
 Mac address flapping detect interval         : 1s
 Mac address flapping syslog supress time : 1800s
 Mac address flapping record max count        : 300
 Mac address flapping record total count    : 2

 Move-Time               VLAN   MAC-Address      Original-Port    Move-Ports     Status
 ------------------- ----- -------------- ---------------- ------------ --------
 2020.11.14 12:10:46 1      0001.1111.1111 ge0/1            ge0/2          Normal
 2020.11.14 12:10:58 1      0001.1111.1111 ge0/2            ge0/1          ERR-DOWN
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
mac-addr-table flapping-logging
!
!
interface gigabitEthernet 0/1
mac-addr-table flapping action error-down
mac-addr-table flapping action priority 5
!
!
interface gigabitEthernet 0/2
mac-addr-table flapping action error-down
mac-addr-table flapping action priority 1
!
```

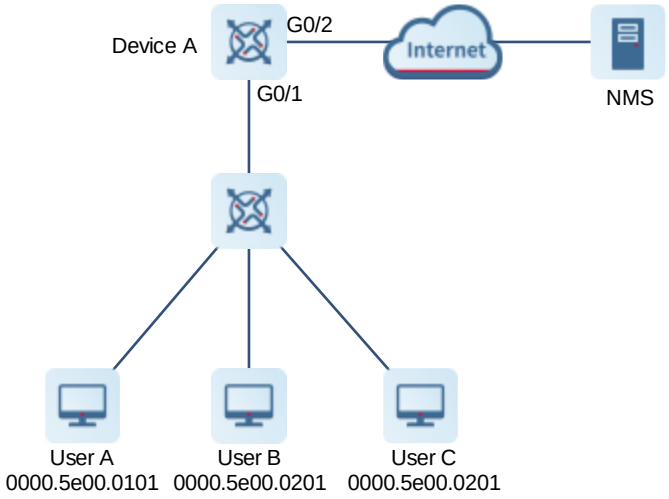
1.13.2 Configuring MAC Address Change Notification

1. Requirements

As shown in Figure 1-5, Device A is connected to devices of User A, User B, and User C, so that Device A records new MAC address entries learned from GigabitEthernet 0/1 and aged MAC address entries, sends MAC address entry change notifications to the specified NMS in SNMP Trap messages, and does not generate massive MAC address change information in a short time to prevent overusing network resources.

2. Topology

Figure 1-5 Configuring MAC Address Change Notification



3. Notes

- Enable the function of MAC address entry change notification in global configuration mode on Device A and configure an interval for sending MAC address Trap messages.
- Enable the function of MAC address entry change notification on GigabitEthernet 0/1 of Device A.
- Specify an SNMP host address and enable Trap notification of MAC address events.
- Ensure that the route between Device A and the NMS is reachable.

4. Procedure

Enable the function of MAC address entry change notification in global configuration mode on Device A and set the interval for sending MAC address Trap messages to 300s.

```
DeviceA# config-system terminal
DeviceA(cfg)# mac-addr-table notification
DeviceA(cfg)# mac-addr-table notification interval 300
```

Enable the function of MAC address entry change notification on GigabitEthernet 0/1 and enable the function of sending Trap messages upon MAC address addition or deletion.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/2)# snmp-agent trap-enable mac-notification added
DeviceA(cfg-if-GigabitEthernet 0/2)# snmp-agent trap-enable mac-notification removed
```

Set the SNMP host address to 192.168.1.1, SNMP version to SNMP v1 and community string to **123** to actively send MAC address Trap messages.

```
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
DeviceA(cfg)# snmp-agent host 192.168.1.10 traps version 2c 123 mac-notification
```

Enable Trap notification of MAC address events.

```
DeviceA(cfg)# snmp-agent enable traps
```

5. Verification

After the preceding configuration, run the **display mac-addr-table notification** command to display the configuration status of the function of MAC address entry change notification in global configuration mode. Run the **display mac-addr-table notification history** command to display the history table for MAC address entry change notifications.

Display the configuration status of the function of MAC address entry change notification in global configuration mode.

```
DeviceA# display mac-addr-table notification
MAC Notification Feature : Enabled
Interval(Sec): 300
Maximum History Size : 50
Current History Size : 0
```

Display the history table for MAC address entry change notifications after the MAC address changes.

```
DeviceA# display mac-addr-table notification history
History Index : 0
Entry Timestamp: 221683
MAC Changed Message :
Operation:ADD Vlan:1 MAC Addr: 0000.5e00.0201 GigabitEthernet 0/1
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
mac-addr-table notification
mac-addr-table notification interval 300
!
!
interface GigabitEthernet 0/1
 snmp-agent trap-enable mac-notification added
 snmp-agent trap-enable mac-notification removed
!
!
snmp-agent host 192.168.1.10 traps version 2c 7 $10$135$2c+O$ mac-notification
snmp-agent enable traps
!
```

Contents

1 Configuring MAC Loopback 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuring Basic Features 1
1.2.1 Overview 1
1.2.2 Procedure 1
1.3 Monitoring 2

1 Configuring MAC Loopback

1.1 Introduction

1.1.1 Overview

Media Access Control (MAC) loopback refers to a case in which packets sent from a local end are transferred internally in the system and returned to the local end for interface self-loop.

MAC loopback can be used to determine whether an Ethernet interface is faulty. For an interface with MAC loopback enabled, if the interface runs normally, the physical status of the interface is always up even without network cable connection; and if the interface runs abnormally, the physical status of the interface is down.

MAC loopback can be applied to multi-port mirroring scenarios. One-to-many mirroring can be implemented by configuring an interface as a MAC loopback interface, and adding the mirroring destination port and expected outbound mirroring port to a remote VLAN. To remove the inner VLAN tag from the inbound direction of a loopback interface, run the following command: **mac-loopback remove-itag**.

1.1.2 Principles

After a MAC loopback interface is configured on a device, the chip of the device sends test packets and then receives these packets through the loopback created internally. By this, the loopback test of the interface is completed.

Caution

With MAC loopback enabled on an interface, data forwarding on this interface is abnormal. After the test, disable MAC loopback.

1.2 Restrictions and Guidelines

- MAC loopback can only be enabled on a physical port instead of a logical interface.
- The function supports working on a Layer 2 or Layer 3 Ethernet interface.

1.3 Configuring Basic Features

1.3.1 Overview

This section describes how to enable MAC loopback on an interface for interface self-loop.

1.3.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

(4) Configure an interface as a loopback interface, and enable MAC loopback on the interface. If the inner VLAN tag needs to be removed from the inbound direction of a loopback interface, add the **remove-itag** parameter.

mac-loopback [**remove-itag**]

MAC loopback is disabled for an interface by default.

Note

The **remove-itag** keyword is not supported by the S6980-64QC, S6920-4C, S6580-48CQ8QC, or S6580-24DC8QC.

(5) Disable MAC loopback after the test.

unset mac-loopback

MAC loopback is disabled for an interface by default.

1.4 Monitoring

This section describes the **display** commands used for checking the running status of a configured function to verify the configuration effect.

Table 1-1 MAC Loopback Monitoring

Command	Purpose
display mac-loopback	Displays the interfaces enabled with MAC loopback.

Contents

1 Configuring VLAN....1

1.1 Introduction....1

1.1.1 Overview....1

1.1.2 Background and Functions....1

1.1.3 Frame Format....2

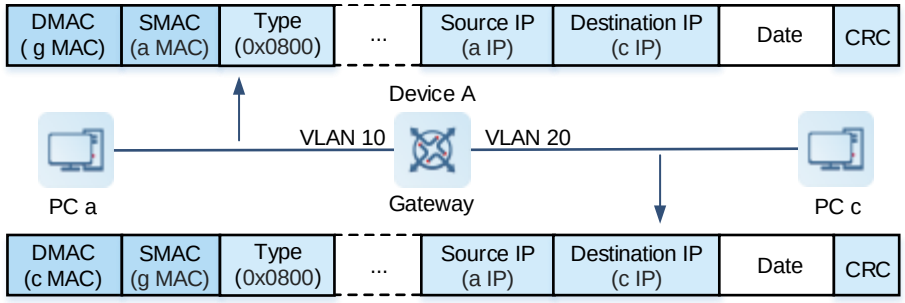
1.1.4 Port Types and Link Types....2

1.1.5 VLAN Frame Processing Rules....4

1.1.6 Intra-VLAN Communication on One Device....4

1.1.7 Intra-VLAN Communication Across Devices....8

1.1.8 Inter-VLAN Communication....9

1.1.9 

VLAN

Assignment Methods....14

1.1.10 Protocols and Standards....15

1.2 Restrictions and Guidelines....15

1.3 Configuration Task Summary....15

1.4 Configuring VLANs....16

1.4.1 Overview....16

1.4.2 Configuration Tasks....16

1.4.3 Creating and Configuring VLANs....16

1.4.4 Configuring VLAN Names....17
1.4.5 Configuring SVIs....17
1.5 Configuring Access Ports....18
1.5.1 Overview....18
1.5.2 Restrictions and Guidelines....18
1.5.3 Procedure (Interface Configuration Mode)....18
1.5.4 Procedure (VLAN Configuration Mode)....19
1.6 Configuring Trunk Ports....20
1.6.1 Overview....20
1.6.2 Restrictions and Guidelines....20
1.6.3 Procedure....20
1.7 Configuring Uplink Ports....21
1.7.1 Overview....21
1.7.2 Restrictions and Guidelines....21
1.7.3 Procedure....21
1.8 Configuring Hybrid Ports....22
1.8.1 Overview....22
1.8.2 Restrictions and Guidelines....22
1.8.3 Procedure....23
1.9 Monitoring....23
1.10 Configuration Examples....24
1.10.1 Configuring Layer 2 Isolation and Layer 3 Interconnection of VLANs....24

1 Configuring VLAN

1.1 Introduction

1.1.1 Overview

A virtual local area network (VLAN) is a logical network created by dividing a physical network. Every VLAN has an independent broadcast domain, and different VLANs are isolated in Layer 2. Devices on different VLANs can implement communication through Layer 3 devices or Layer 3 interfaces.

1.1.2 Background and Functions

1. Background

The early local area network (LAN) adopts a bus structure, and STAs on the bus belong to the same contention domain. Bridges and Layer 2 switches can isolate STAs onto different ports and control packet forwarding to isolate the contention domain. All ports of Layer 2 switches belong to the same broadcast domain. When a switch receives a frame with an unknown destination address, the switch broadcasts this frame to all ports in the broadcast domain other than the source port. After receiving a reply from a port that matches the destination address, the switch learns the new MAC address based on the source address in the reply packet. This is the basic working mode of the switch. The working mode will lead to two problems:

- If the network scale is enlarged, the increasing broadcast packets waste bandwidth.
- As packets in a LAN are broadcast, each STA in the broadcast domain can receive broadcast packets in the entire LAN. If users obtain these packets using a packet capture tool, the content in the packets may be intercepted, which affects information security.

Use routers to solve the preceding two problems. Routers communicate based on Layer 3 IP addresses, and different subnets belong to different broadcast domains. This dampens the forwarding of broadcast packets. However, high forwarding performance of routers depends on central processing units (CPUs), and requires a high cost. Therefore, routers are usually used as egresses between LANs and wide area networks (WANs). In contrast, a VLAN is a solution with lower cost and solves the preceding problems by dividing the physical LAN into multiple logical LANs.

2. Functions

- Layer 2 isolation

  You can divide a physical LAN into multiple VLANs by using different division methods so as to isolate broadcast domains, which saves bandwidth and enhances LAN security. A VLAN shares the same attributes with a common physical LAN except the restriction on physical location. Every VLAN has an independent broadcast domain. If no Layer 3 interface interconnected with VLANs is configured, the VLANs are mutually isolated on Layer 2. Users in different VLANs cannot communicate with each other. Layer 2 unicast, broadcast and multicast frames are forwarded and transmitted in the same VLAN without entering other VLANs.

- Layer 3 interconnection

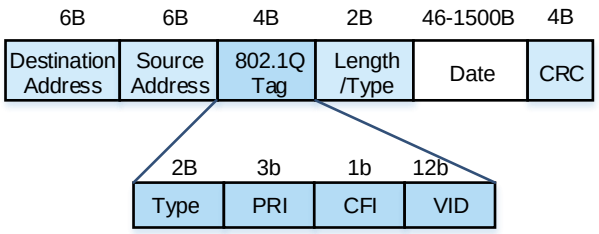
To realize Layer 3 communication between VLANs, a virtual interface is developed for an Layer 3 switch, based on the concept of router subinterfaces. The virtual interface is referred to as Switch Virtual Interface (SVI). In a Layer 3 switch, you can create an SVI and an SVI IP address for each VLAN. The SVI functions as a VLAN gateway to implement Layer 3 communication between VLANs. Therefore, an SVI is also referred to as a VLAN interface.

1.1.3 Frame Format

Based on IEEE 802.1Q, a 4-byte 802.1Q tag is inserted between the source MAC address field and protocol type field of an Ethernet frame, as shown in Figure 1-1. The VLAN ID (VID) field in the tag is used to represent the VLAN where the frame comes from. Not all devices identify tagged 802.1Q frames. Therefore, Ethernet frames can be tagged or untagged in the VLAN.

- Type: two bytes. The value **0X8100** indicates 802.1Q frames. Devices not supporting 802.1Q discard these frames upon receiving them.
- PRI: three bits. It indicates the Layer 2 priority and corresponds to the Class of Service (CoS) priority of the Quality of Service (QoS). The value range is from 0 to 7. A larger value indicates a higher priority.
- CFI: one bit. It differentiates among Ethernet frames, Fiber Distributed Data Interface (FDDI) frames, and token ring network frames. The value **0** indicates Ethernet frames.
- VID: 12 bits. It indicates the ID of a VLAN to which packets belong. The value range of a VLAN ID is from 1 to 4094. The value **0** and **4095** are reserved and unusable.

Figure 1-1 802.1Q Frame Format



1.1.4 Port Types and Link Types

Interfaces can be classified into Layer 2 interfaces and Layer 3 interfaces. Only Layer 2 interfaces can be added to VLANs. Table 1-1 lists basic types of Layer 2 interfaces. You can configure the Layer 2 interface type of an interface to determine the allowed VLANs of the interface, and whether the packets forwarded by the interface are tagged.

Table 1-1 Basic Types of Layer 2 Interfaces

Port Type	Function
Access	An access port carries traffic for only one VLAN. Only frames from this VLAN can pass the port. This VLAN is referred to as access VLAN, which owns attributes of a native VLAN and an allowed VLAN. The frames sent from the access port do not carry tags. When the access port receives an untagged frame from a peer device, the local device determines that the frame comes from the access VLAN and adds the access VLAN ID to the frame.
Trunk	A trunk port carries traffic for one native VLAN and multiple allowed VLANs. The frames forwarded by the trunk port from the native VLAN do not carry tags and the frames forwarded by the trunk port from allowed VLANs carry tags. Note that the trunk ports of a link must be configured with the same native VLAN.
Uplink	An uplink port carries traffic for one native VLAN and multiple allowed VLANs. Frames sent by the uplink port are tagged. In a QinQ scenario, frames sent by a Provider Edge (PE) device port connected to a public network must carry tags. Therefore, the port must be configured as an uplink port. A Customer Edge (CE) device port connected to the PE device can also be configured as an uplink port. For more information, see *Configuring QinQ*.
Hybrid	A hybrid port carries traffic for one native VLAN and multiple allowed VLANs. Allowed VLANs are classified into tagged VLANs and untagged VLANs. The frames forwarded by the hybrid port from a tagged VLAN carry tags, and the frames forwarded by the hybrid port from an untagged VLAN do not carry tags. The frames forwarded by the hybrid port from the native VLAN must not carry tags, because the native VLAN is an untagged VLAN. The MAC VLAN function applies to only untagged VLANs on the hybrid port. Therefore, the hybrid port can be used as an interface to connect to downlink clients in MAC VLAN scenarios. For more information, see *Configuring MAC VLAN*.

VLAN links include access links and trunk links.

- An access link connects STAs to a switch interface. Although some STAs can identify frames that carry VLAN tags, most STAs cannot do so. Generally, a port transmits untagged frames in the access link to ensure the compatibility of different STAs.
- A trunk link can bear data of different VLANs. You must ensure that the devices connected by the trunk link can identify the VLAN of frames transmitted on the trunk link. Untagged frames and tagged frames are transmitted based on different types of ports. Ensure that the native VLAN configuration on the peer device is consistent with that on the local device so that the trunk port transmits untagged frames from the native VLAN and tagged frames from other allowed VLANs,

Access ports can be connected by an access link only. Hybrid ports can be connected by both an access link and a trunk link. Trunk ports and uplink ports can be connected by a trunk link only.

1.1.5 VLAN Frame Processing Rules

- Frame receiving rule

  When detecting an incoming untagged frame, the port regards that the frame comes from the native VLAN of the port. If the native VLAN is an allowed VLAN of the port, the port receives the packet, and identifies the frame as one from the native VLAN. Otherwise, the port discards the packet.

  When detecting a tagged frame, the port receives the packet if the port allows the VLAN identified by the tag. Otherwise, the port discards the packet.

- Frame forwarding rule

  A Layer 2 switch can forward packets to only ports that match allowed VLANs (other than the source port).

  An Layer 3 switch can determine an egress VLAN by querying the routing table in software and forward packets to a port that allows the egress VLAN.

- Frame sending rule

  After packets are forwarded in the switch, the switch determines whether to add tags to the packets based on the VLAN type (tagged VLAN or untagged VLAN).

The following example displays actual application of the above rules. Table 1-2 shows the interfaces and configurations in the example.

Table 1-2 Interface Examples

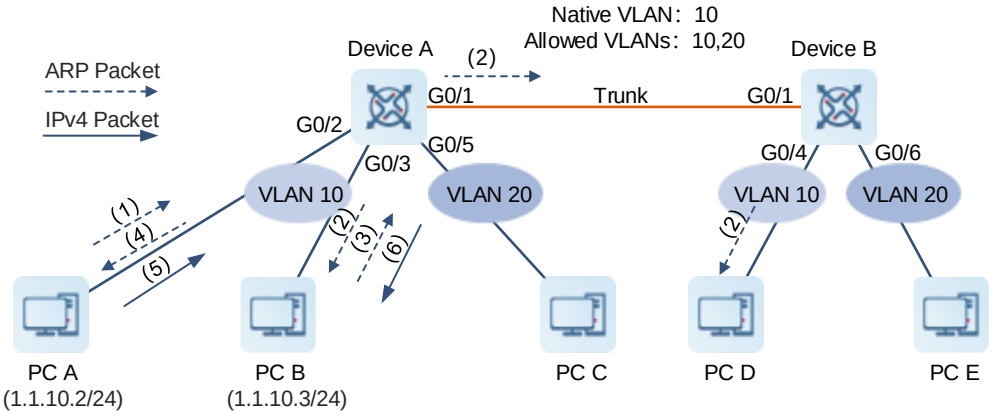
Interface	Mode	Native VLAN	Allowed VLAN
GigabitEthernet 0/1	Access	10	10 (Untagged)
GigabitEthernet 0/2	Trunk	10	10 (Untagged), 20 (Tagged), and 30 (Tagged)
GigabitEthernetGi 0/3	Hybrid	20	20 (Untagged), 10 (Tagged), and 30 (Tagged)
GigabitEthernetGi 0/4	Uplink	30	30 (Tagged) and 20 (Tagged)

When GigabitEthernet 0/1 receives an untagged frame or a frame with a VLAN 10 tag, it forwards the frame to GigabitEthernet 0/2 and GigabitEthernet 0/3 that allow VLAN 10. The VLAN 10 frame sent to GigabitEthernet 0/2 is untagged, and the VLAN 10 frame sent to GigabitEthernet 0/3 is tagged. When GigabitEthernet 0/4 receives an untagged frame or a frame with a VLAN 30 tag, it forwards the frame to GigabitEthernet 0/2 and GigabitEthernet 0/3 that allow VLAN 30. The VLAN 30 frame sent to GigabitEthernet 0/2 is tagged, and the VLAN 30 frame sent to GigabitEthernet 0/3 is untagged.

1.1.6 Intra-VLAN Communication on One Device

As shown in the topology in Figure 1-2, Device A and Device B are connected through the trunk port GigabitEthernet 0/1. PC A and PC B are connected to VLAN 10 through device A. The communication between PC A and PC B is intra-VLAN communication on one device.

Figure 1-2 An Example of Intra-VLAN Communication Within One Device



Through IP addresses, PC A (1.1.10.2/24) first determines that PC B (1.1.10.3/24) is in the same subnet as the device. The device uses the MAC address of PC B as the destination MAC address for communication. The communication steps are as follows:

If PC A does not find the MAC address of PC b, the communication goes from step (1); if PC A finds the MAC address of PC B, the communication goes from step (5).

(1) PC A does not find the MAC address of PC B in the Address Resolution Protocol (ARP) table. PC A sends a broadcast ARP request packet to request the MAC address of PC b. Device A learns the MAC address of PC A through the ARP request packet.

(2) Device A forwards the broadcast ARP request packet. PC B learns the MAC address of PC A through the ARP request packet.

(3) PC B sends a unicast ARP reply packet to PC A. Device A learns the MAC address of PC B through the ARP reply packet.

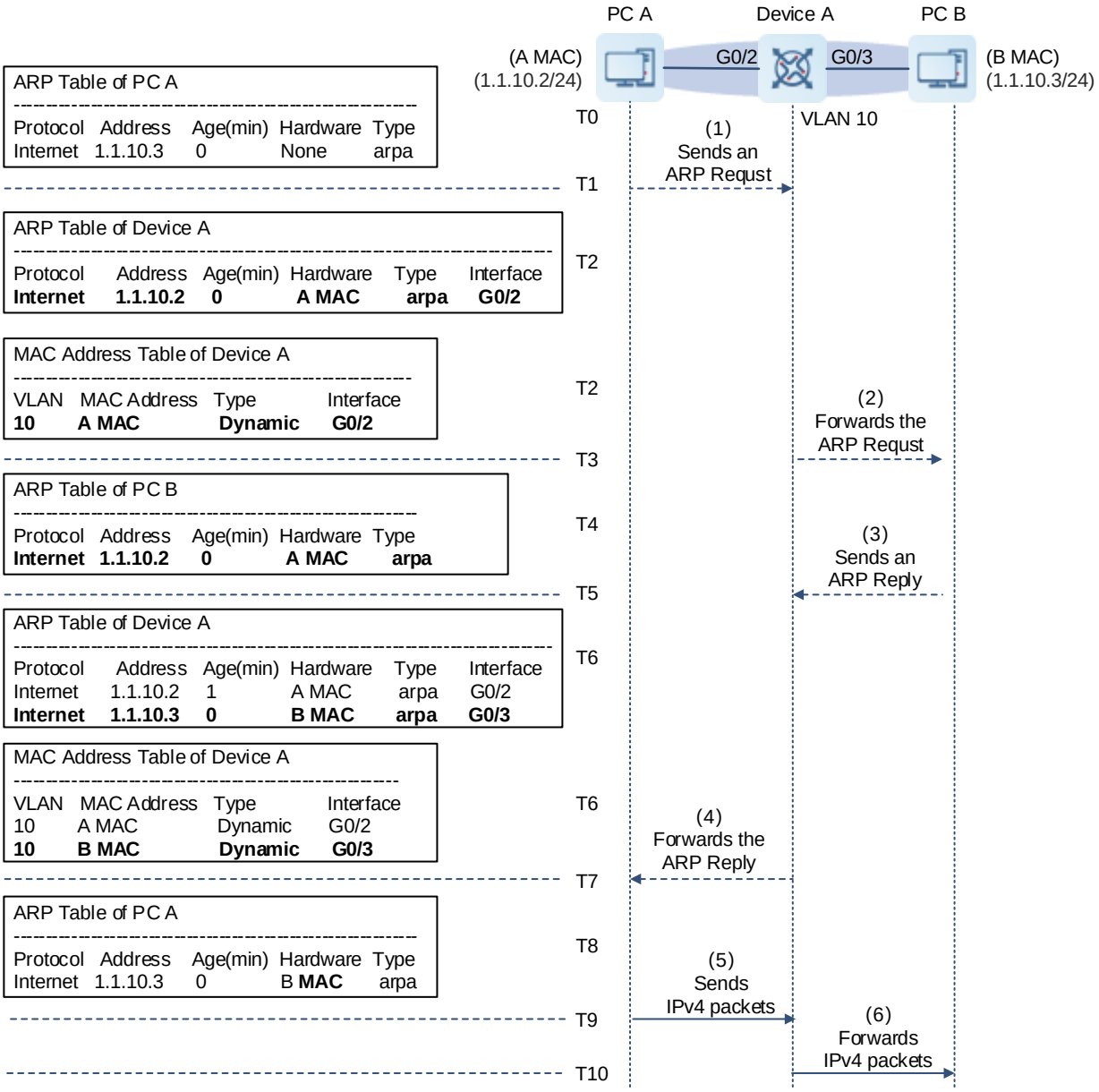
(4) Device A finds the MAC address of PC A in the ARP table and forwards the unicast ARP reply packet. PC A learns the MAC address of PC B through the ARP reply packet.

(5) PC A finds the MAC address of PC B (destination MAC address) in the ARP table and sends a unicast IPv4 packet to PC B.

(6) Device A finds the MAC address of PC B in the ARP table and forwards the unicast IPv4 packet.

Steps of intra-VLAN communication on one device are shown in Figure 1-3.

Figure 1-3 Steps of Intra-VLAN Communication

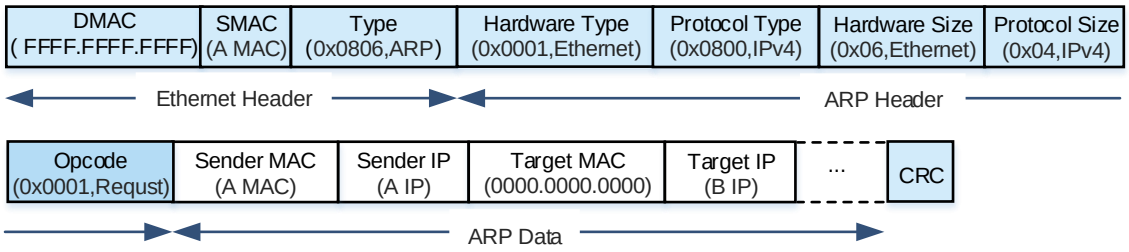


(1) PC A sends an ARP request packet to request the MAC address of PC B.

a. If PC A does not find the MAC address of PC B as shown in T0 in Figure 1-3, PC A sends an ARP request packet encapsulated in an Ethernet broadcast frame (as shown in Figure 1-4) to PC B as shown in T1 in Figure 1-3.

b. Upon receiving the ARP request packet of PC A from GigabitEthernet 0/2, if Device A finds that the packet is untagged, it determines that the packet comes from VLAN 10 and receives the packet. Device A learns the source IP address in the packet. Device A enters the mapping among the IP address, MAC address, port, and VLAN of PC A into the ARP table and enters the mapping among the VLAN, MAC address and port of PC A into the dynamic MAC address table, as shown in T2 in Figure 1-3.

Figure 1-4 The ARP Request Packet Sent by PC A for Requesting the MAC Address of PC B



(2) Device A forwards the ARP request packet.

a. Device A searches for the port that allows VLAN 10, and broadcasts the ARP request to GigabitEthernet 0/1 and GigabitEthernet 0/3 except the source port GigabitEthernet 0/2, as shown in T3 in Figure 1-3.

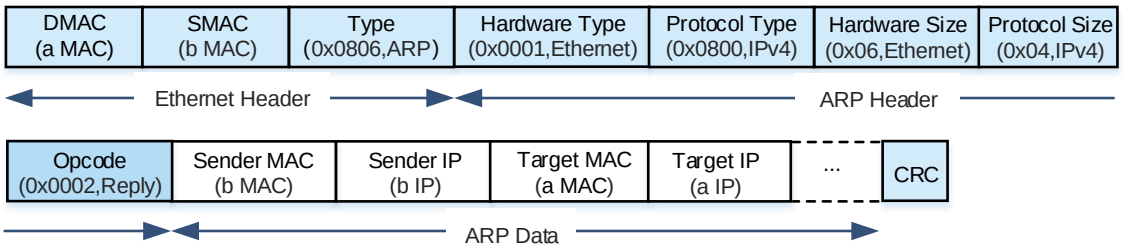
b. PC B receives the ARP request (as shown in Figure 1-4) of PC A from GigabitEthernet 0/3, learns the mapping between the IP address and MAC address of PC A, and enters the mapping into the ARP table.

(3) PC B sends a unicast ARP reply packet.

a. PC B finds that the destination IP address in the ARP request matches its own IP address, it sends a unicast ARP reply (as shown in Figure 1-5) to PC A, as shown in PC B in T5 in Figure 1-3.

b. Upon receiving the ARP reply of PC B from GigabitEthernet 0/3, Device A learns the source IP address. Device A enters the mapping among the IP address, MAC address, port, and VLAN of PC B into the ARP table and enters the mapping among the VLAN, MAC address and port of PC B into the dynamic MAC address table, as shown in T6 in Figure 1-3.

Figure 1-5 The ARP Reply Packet Sent by PC B to PC A



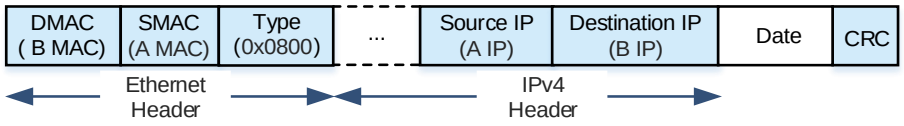
(4) Device A forwards the ARP reply packet.

a. Device A queries the MAC address table, matches MAC address of PC A, and forwards the ARP reply of PC B to GigabitEthernet 0/2, as shown in T7 in Figure 1-3.

b. PC A receives the ARP reply (as shown in Figure 1-5) of PC B from GigabitEthernet 0/2, learns the mapping between the IP address and MAC address of PC B, and enters the mapping into the ARP table, as shown in T8 in Figure 1-3.

PC A sends an IPv4 packet.PC A finds the MAC address of PC B in the ARP table, generates a packet with the MAC address of PC B as the destination MAC address and the MAC address of PC A as the source MAC address, and sends the packet (as shown in Figure 1-6) to PC B, as shown in T9 in Figure 1-3 .

Figure 1-6 The IPv4 Packet Sent from PC A to PC B



(5) Device A forwards the IPv4 packet.

Upon receiving the packet from PC A, Device A refreshes the MAC address table based on the source IP address of PC A, queries the MAC address table and matches the destination MAC address of PC B, and forwards the packet (as shown in Figure 1-6) to GigabitEthernet 0/3, as shown in T10 in Figure 1-3.

1.1.7 Intra-VLAN Communication Across Devices

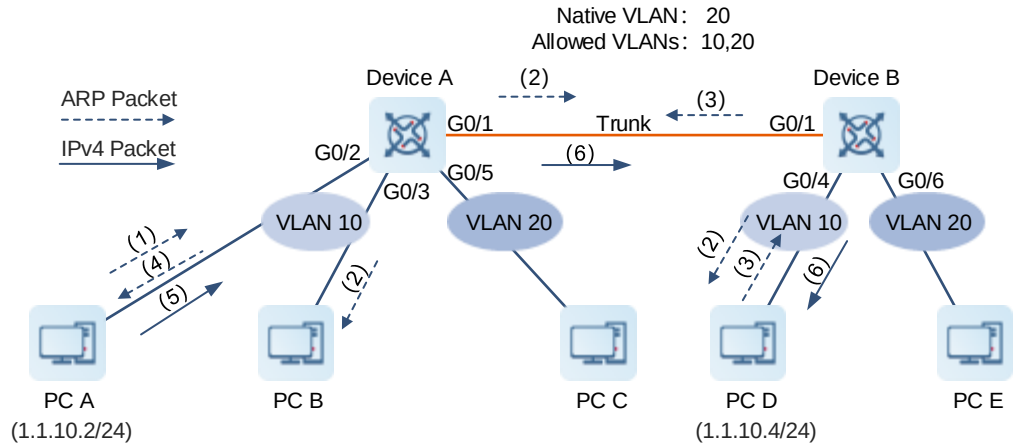
Ports on different devices can be assigned to the same VLAN. The procedure of intra-VLAN communication across devices is similar to the procedure of intra-VLAN communication on one device, except for the following:

- The trunk port sends a packet. If the packet comes from the native VLAN, the packet is untagged. Otherwise, the packet is tagged.
- When receiving a packet, the trunk port determines that the packet comes from the native VLAN if it is untagged. If the native VLAN is an allowed VLAN of the port, the packet is allowed to pass through the port. Otherwise, the packet is not allowed to pass through the port. If the packet is tagged, the port check whether the packet comes from an allowed VLAN. If yes, the port receives the packet and forwards it. If not, the port discards the packet.

As shown in Figure 1-9, Device A and Device B is connected through the trunk port GigabitEthernet 0/1. PC A is connected to VLAN 10 through Device A, and PC d is connected to VLAN 10 through Device B. The communication between PC A and PC d is intra-VLAN communication across devices.

To explain the process of adding and removing a tag, assume that VLAN 20 is the native VLAN of GigabitEthernet 0/1.

Figure 1-7 An Example of Intra-VLAN Communication Across Devices



PC A determines that PC d is on the same network segment as the device. PC A needs to use the MAC address of PC d as the destination MAC address for communication. The communication steps are as follows:

(1) PC A does not find the MAC address of PC d (the destination MAC address) in the ARP table. PC A sends a broadcast ARP request packet to request the MAC address of PC d. Device A learns the MAC address of PC A through the ARP request packet.

(2) Device A forwards the broadcast ARP request packet. As VLAN 10 is not the native VLAN of the trunk port GigabitEthernet 0/1, add a VLAN 10 tag to the packet and send the packet. When the GigabitEthernet 0/1 port of Device B receives the packet, the port finds that the packet carries a VLAN 10 tag and VLAN 10 is an allowed VLAN. The port allows the packet to pass through. The port removes the VLAN 10 tag from the packet and forwards the ARP request packet to the GigabitEthernet 0/4 port. Device B and PC d learn the MAC address of PC A through the ARP request packet.

(3) PC d sends a unicast ARP reply packet. The GigabitEthernet 0/1 port of Device B adds a VLAN tag to the ARP reply packet and forwards the packet. Device B and Device A learn the MAC address of PC d through the ARP reply packet.

(4) When the GigabitEthernet 0/1 port of Device A receives the packet, the port finds that the packet carries a VLAN 10 tag and VLAN 10 is an allowed VLAN. The port allows the packet to pass through. Device A queries the ARP table and matches the MAC address of PC A. The port removes the VLAN 10 tag from the packet and forwards the unicast ARP request packet to the GigabitEthernet 0/2 port.

(5) PC A finds the MAC address of PC d (the destination MAC address) in the ARP table and sends a unicast IPv4 packet to PC d.

(6) Device A queries the ARP table, matches the MAC address of PC d, adds a VLAN 10 tag to the IPv4 packet and forwards the packet. When receiving the packet, .Device B queries the ARP table, matches the MAC address of PC d, removes the VLAN 10 tag from the IPv4 packet and forwards the packet.

1.1.8 Inter-VLAN Communication

In practice, devices from different VLANs may need to communicate with each other. Although VLAN broadcast domains are isolated in Layer 2, the devices from different VLANs can communicate with each other by using a Layer 3 device. But this method requires routed ports. To address this problem, a routing module can be added to a Layer 2 switch to form a Layer 3 switch.

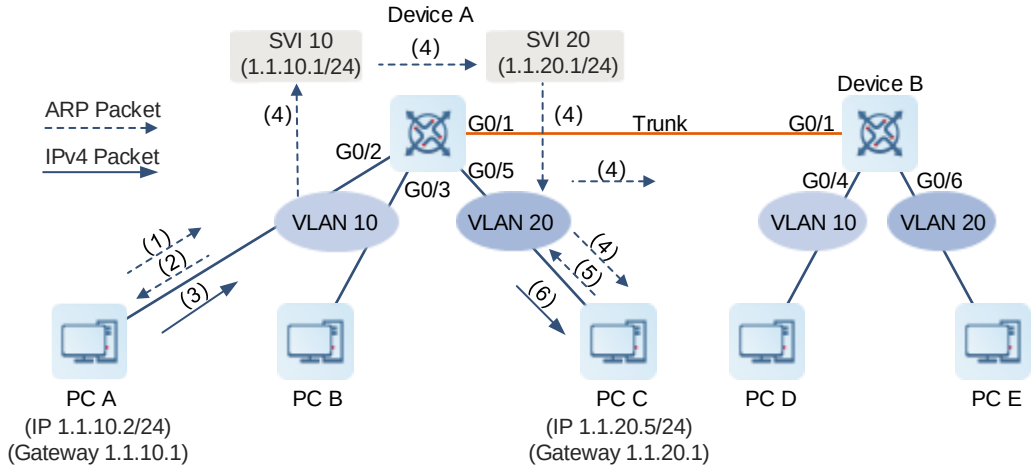
On a Layer 3 switch, one VLAN corresponds to one SVI. The Layer 3 switch takes the SVI as an independent interface, creates an ARP entry for the interface, and records the mapping between the IP address and MAC address of the SVI.

An STA in each VLAN uses the IP address of an SVI corresponding to this VLAN as the default gateway. When the STA in the VLAN tries to communicate with an STA in another VLAN, the STA queries the MAC address of the gateway to be used as the destination MAC address, and then forwards a packet to the gateway.

When the Layer 3 switch receives the packet whose MAC address matches the SVI, the switch enables Layer 3 forwarding. If the Layer 3 switch does not know the MAC address of the destination host, the Layer 3 switch queries the routing table in the software, determines the egress VLAN of the destination host, and sends an ARP request in the egress VLAN. Upon obtaining an ARP reply from the destination host, the Layer 3 switch learns the IP address, MAC address, port, and VLAN of the destination host, updates the Layer 3 routing table, replaces the source MAC address in the packet with the MAC address of the gateway and the destination MAC address with the MAC address of the destination host, and forwards the packet to the destination host. Later, when the Layer 3 switch receives packets with the same destination MAC address, the Layer 3 switch directly queries the Layer 3 routing table, replaces the source and destination MAC addresses, and forwards the packets.

As shown in Figure 1-8, PC A belongs to VLAN 10 and PC C belongs to VLAN 20. They communicate with each other across the VLANs.

Figure 1-8 An Example of Inter-VLAN Communication



After an SVI is created, Device A generates the mappings of IP addresses to MAC addresses and a directly connected route to the Layer 3 SVI. As shown in T0 in Figure 1-9, the gateway MAC address is the MAC address of Device A.

The IP address of PC A is 1.1.10.2/24 and its corresponding gateway IP address is 1.1.10.1/24. The IP address of PC C is 1.1.20.5/24 and its corresponding gateway IP address is 1.1.20.1/24. To communicate with PC C, PC A first determines that they are on different network segments. PC A takes the gateway MAC address as the destination address for communication. The communication steps are as follows:

PC A queries the ARP table for the gateway MAC address corresponding to the gateway IP address (1.1.10.1/24). If PC A does not find the MAC address of the gateway, goes from step (1). If PC A finds the MAC address of the gateway, goes from step (3).

(1) PC A does not find the MAC address of the gateway, it sends a broadcast ARP request packet to request the MAC address of the gateway. Device A learns the MAC address of PC A through the ARP request packet.

(2) Device A sends a unicast ARP reply packet. PC A learns the MAC address of the gateway through the ARP reply packet.

(3) PC A finds the MAC address of the gateway of in the ARP table, uses the gateway MAC address as the destination MAC address, and the IP address of PC C as the destination IP address, and sends a unicast IPv4 packet to PC C. When receiving the packet, if Device A does not find the MAC address of PC C, it performs step (4). If Device A finds the MAC address of PC C, it performs step (6) directly.

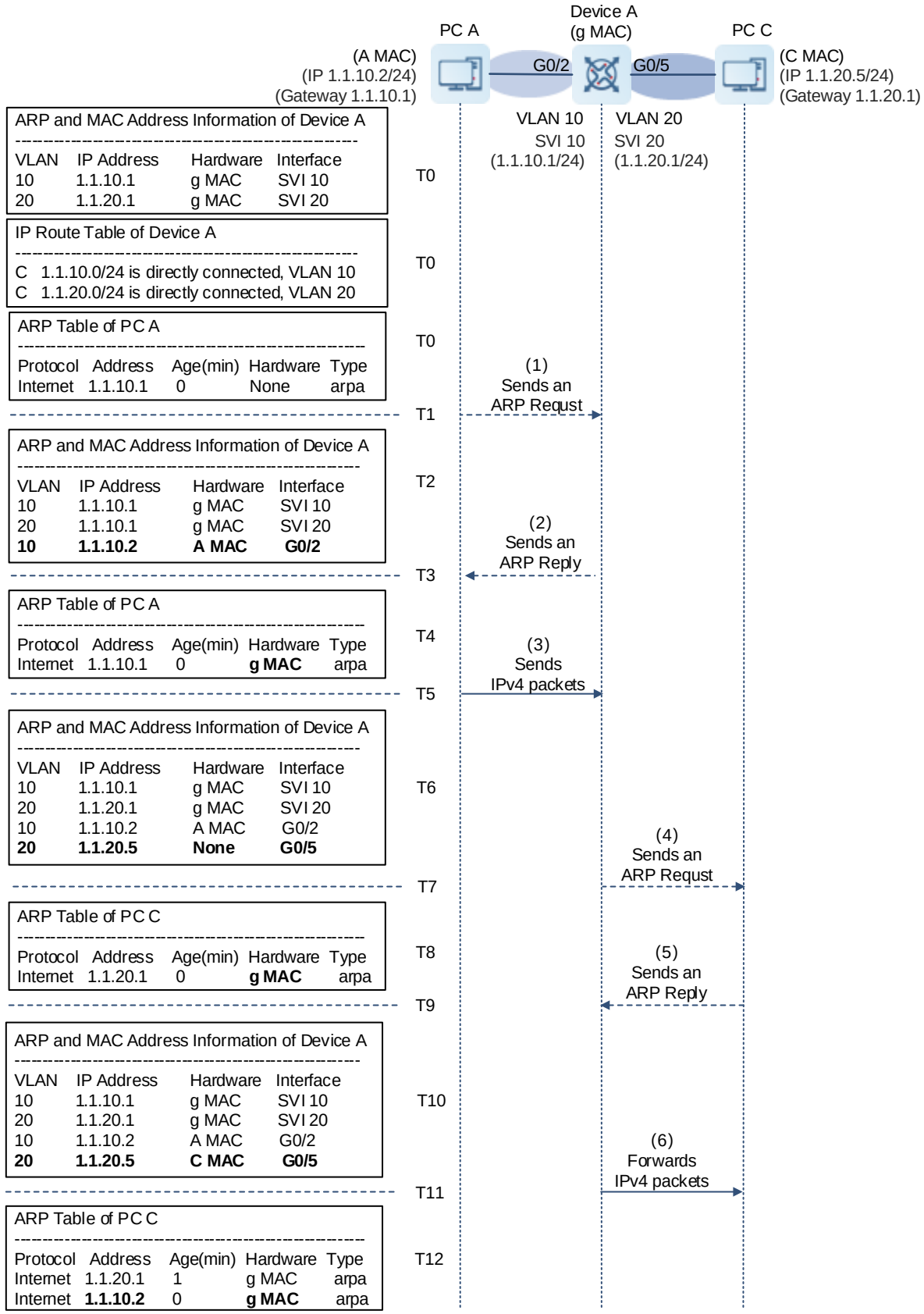
(4) Device A identifies that the destination MAC address is the MAC address of the gateway, and sends the packet to SVI 10 for forwarding. The CPU queries the software routing table, and finds that the egress VLAN corresponding to the destination IP address of PC C is VLAN 20 through the directly connected route of Layer 3 SVIs. SVI 20 in VLAN 20 sends a broadcast ARP request to request the MAC address corresponding to the IP address of PC C. PC C learns the mapping of the MAC address of the gateway and the IP address of SVI 20 through the packet.

(5) PC C sends a unicast ARP reply. Device A learns the MAC address of PC C.

(6) For the packet sent from PC A to PC C, Device A changes its source MAC address from a MAC to g MAC, and changes its destination MAC address from g MAC to c MAC, and then forwards it. PC C learns the mapping of the MAC address of the gateway and the IP address of PC A through the packet.

Steps of inter-VLAN communication are shown in Figure 1-9.

Figure 1-9 Steps of Inter-VLAN Communication

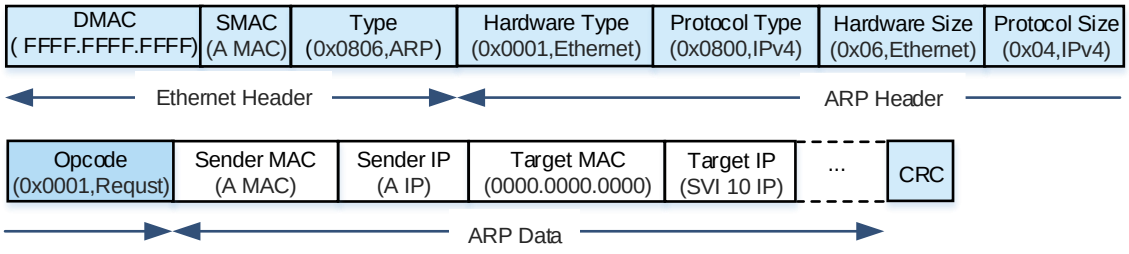


(7) PC A sends an ARP request packet to request the MAC address of the gateway.

a. If PC A does not find the MAC address of the gateway IP address (1.1.10.1) as shown in T0 in Figure 1-9, PC A sends an ARP request packet encapsulated in an Ethernet broadcast frame (as shown in Figure 1-10) to the gateway as shown in T1 in Figure 1-9. The gateway of PC A is SVI 10 whose IP address is 1.1.10.1/24. Therefore, the gateway uses the IP address of SVI 10 as a target IP address.

b. When receiving the ARP request packet of PC A from GigabitEthernet 0/2, if Device A finds that the packet is untagged, it determines that the packet comes from VLAN 10 and receives the packet. Device A learns the source IP address in the packet, and enters the mapping among the IP address, MAC address, port, and VLAN of PC A into the ARP table. Device A enters the mapping among the VLAN, MAC address, and port of PC A into the dynamic MAC address table. The ARP table and the MAC address table contain information about Layer 3 forwarding and such information is called as a Layer 3 routing table, as shown in T2 in Figure 1-9. The routing table cannot be queried but its content is reflected in the ARP table and MAC address table.

Figure 1-10 The ARP Request Packet Sent from PC A for the MAC Address of the Gateway

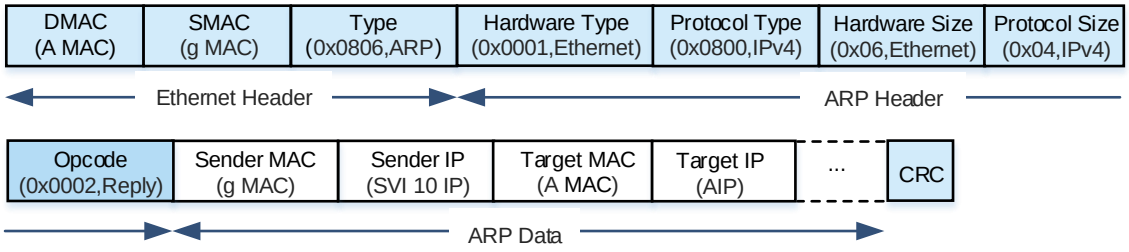


(2) Device A sends an ARP reply packet.

a. If Device A finds that the packet is an ARP packet and the destination IP address in the packet is the IP address of the SVI 10 of this device, it returns a unicast ARP reply and forwards the reply packet through GigabitEthernet 0/2, as shown in T3 in Figure 1-11Figure 1-9. The reply packet (as shown in Figure 1-11) includes the IP address and MAC address of SVI 10.

b. Upon receiving the ARP reply, PC A learns the MAC address corresponding to the gateway IP address of the SVI 10 and updates the ARP table of PC A, as shown in T4 in Figure 1-9.

Figure 1-11 ARP Reply Packet Sent by the Gateway to PC A

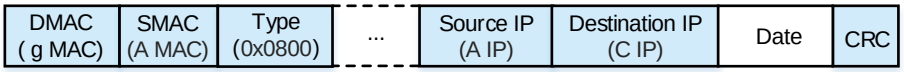


(3) PC A sends an IP packet to PC C.

a. If PC A finds the MAC address of the gateway in the ARP table, it generates a packet with the gateway MAC address as the destination MAC address and the MAC address of PC A as the source MAC address, encapsulates the packet in an Ethernet frame (as shown in Figure 1-12), and forwards the packet to the gateway, as shown in T5 in Figure 1-11Figure 1-9.

b. Upon receiving the packet of PC A from GigabitEthernet 0/2 (as shown in Figure 1-12), Device A finds that the packet is untagged. Device A determines that the packet comes from VLAN 10 and updates the MAC address entry of PC A. The destination MAC address (gateway MAC address) is the gateway MAC address of the local device. Therefore, the packet must be forwarded at Layer 3. If Device A finds the MAC address of PC C corresponding to the IP address of PC C, the device goes to step (6).

Figure 1-12 IPv4 Packet Sent by PC A to PC C



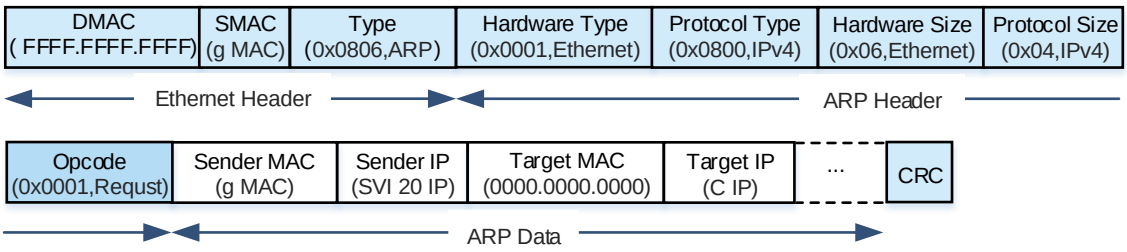
(4) The gateway sends an ARP request for the MAC address of PC C.

a. If Device A does not find the MAC address corresponding to the IP address of PC C (1.1.20.5), it sends the packet to SVI 10 to forward the packet to the CPU.

b. The CPU queries the software routing table, and finds that the IP address matches a direct route to 1.1.20.0/24 and the egress VLAN is VLAN 20. In the egress VLAN (VLAN 20), SVI 20 sends a broadcast ARP request packet for the MAC address of PC C, as shown in T7 in Figure 1-11Figure 1-9. The gateway of PC C is SVI 20. Therefore the gateway uses the IP address of SVI 20, 1.1.20.1/24, as its IP address.

c. The ARP request is sent from GigabitEthernet 0/1 and GigabitEthernet 0/5 that allow VLAN 20. PC C receives the ARP request (as shown in Figure 1-13) of Device A from GigabitEthernet 0/5, learns the mapping between the gateway IP address and gateway MAC address in VLAN 20 on Device A, and enters the mapping into the ARP table, as shown in T8 in Figure 1-11Figure 1-9.

Figure 1-13 ARP Request Packet Sent by the Gateway for Requesting MAC Address of PC C

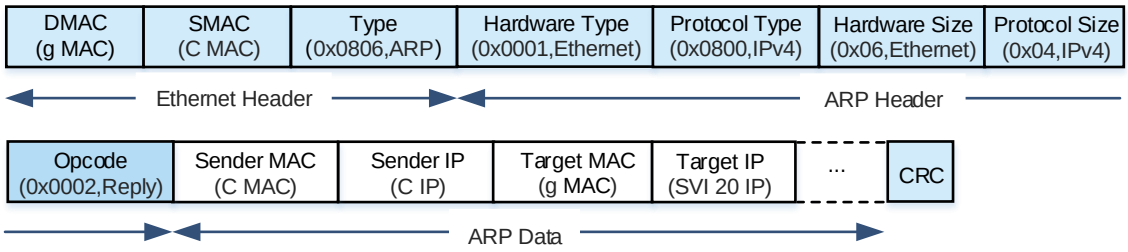


(5) PC C sends an ARP reply packet.

a. If PC C finds that the destination IP address in the ARP request matches its own IP address, it sends an ARP reply packet in unicast mode (as shown in Figure 1-14), as shown in T9 in Figure 1-11Figure 1-9.

b. Upon receiving the ARP reply (as shown in Figure 1-14) of PC C, Device A learns the mapping among the IP address, MAC address, port, and VLAN of PC C, and enters the mapping into the ARP table, as shown in T10 in Figure 1-11Figure 1-9.

Figure 1-14 ARP Reply Packet Sent by PC C to the Gateway

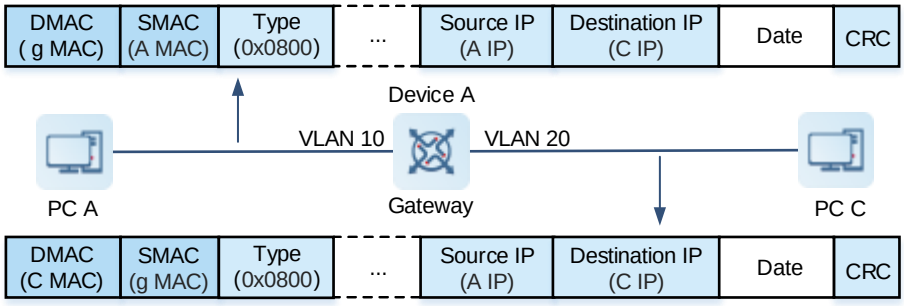


(6) The gateway forwards the IP packet.

a. If Device A finds the MAC address of PC C in the routing table, it replaces the MAC address of PC A with the gateway MAC address and the gateway MAC address with the MAC address of PC C, and forwards the packet of PC A to PC C (as shown in Figure 1-15), as shown in T11 in Figure 1-11Figure 1-9.

b. PC C receives this packet (as shown in Figure 1-15) from GigabitEthernet 0/5, learns the mapping between the IP address of PC A and the gateway MAC address, and enters the mapping into the ARP table, as shown in T12 in Figure 1-11Figure 1-9. Later, when PC C sends the packet to PC A, it can query the gateway MAC address corresponding to the IP address of PC A (1.1.10.2), and encapsulates the packet for PC A by using the gateway MAC address as the destination MAC address.

Figure 1-15 Packet Forwarded by the Gateway



1.1.9 VLAN Assignment Methods

1. Port-based Assignment

- Port-based VLAN

Port-based VLAN assignment is the most basic VLAN assignment method. After the administrator configures a port with a Port Default VLAN ID (PVID), the port is assigned to a VLAN. The VLAN ID range is from 1 to 4094. In case of no VLAN assignment, all ports are assigned to VLAN 1 by default.

- Super VLAN

Super VLAN is also referred to as VLAN aggregation, which aggregates multiple VLANs into one IP address segment. Only one IP address is assigned to a super VLAN that contains multiple sub-VLANs, and this saves address resources and facilitates network management. A VLAN accessible by an STA is subject only to an access port of the STA. That is, the STA is assigned to a VLAN based on the STA's access port. For more information, see *Super VLAN*.

- Private VLAN

Private VLAN is a technology that can divide an Layer 2 broadcast domain of a VLAN into multiple sub domains, and each sub domain is composed of one private VLAN pair: primary VLAN and secondary VLAN. This technology can increase the number of users that can be supported by an operator network and reduce waste of IP address resources. A VLAN accessible by an STA is subject only to an access port of the STA. That is, the STA is assigned to a VLAN based on the STA's access port. For more information, see *Private VLAN*.

2. Packet-based Assignment

- MAC-based VLAN

MAC-based VLAN is a VLAN assignment method based on MAC addresses. When an STA is physically relocated, it can re-access the network from another device and there is no need to reconfigure a VLAN for the access port. If different STAs access the network from the same port, they can be assigned to different VLANs based on their MAC addresses. For more information, see *MAC VLAN*.

Usually, MAC-based VLAN is used with the 802.1x VLAN delivery function to ensure secure and flexible access of 802.1x STAs.

- Voice VLAN

A voice VLAN is a VLAN dedicated for voice data streams of users. With this technology, data streams and voice streams are transmitted in the data VLAN and the voice VLAN, respectively, to prevent mutual interference between voice calls and service packets. This function identifies voice data and assigns the voice data to the voice VLAN. This method belongs to packet content-based VLAN assignment. For more information, see *Voice VLAN*.

- Protocol-based VLAN

Protocol-based VLAN is a VLAN assignment technology based on the protocol type or IP subnet in packets. The service types supported in a network are bound to VLANs to facilitate management and maintenance. This method belongs to packet-based VLAN assignment. For more information, see *Protocol-based VLAN*.

1.1.10 Protocols and Standards

IEEE 802.1Q: IEEE Standard for Local and Metropolitan Area Networks-Virtual Bridged Local Area Networks

1.2 Restrictions and Guidelines

- Before you configure a port type or add a port to a VLAN, ensure that the port is a Layer 2 interface.
- Run the **display interface switch-port** command to show the port type. If the **Switch-port** field corresponding to the port is set to **enabled**, the port is a Layer 2 interface. Otherwise, the port is a Layer 3 interface.
- Run the **switch-port** command to configure a port as a Layer 2 interface. Run the **unset switch-port mode** command to restore the default configuration of the Layer 2 interface mode. Then run the **unset switch-port** command to configure the Layer 2 interface as a Layer 3 interface. When you configure a port as a Layer 2 or Layer 3 interface, all attributes of the interface are restored to the default values in the corresponding interface mode.

1.3 Configuration Task Summary

VLAN configuration includes the following tasks:

(1) Configuring VLANs

(2) Creating and Configuring VLANs

- (Optional) Configuring VLAN Names
- (Optional) Configuring SVIs

(3) Configure a port type and add ports to a VLAN. Configure at least one of the following tasks.

- Configuring Access Ports
- Configuring Trunk Ports
- Configuring Uplink Ports
- Configuring Hybrid Ports

1.4 Configuring VLANs

1.4.1 Overview

Create a VLAN, and enter the VLAN configuration mode.

1.4.2 Configuration Tasks

VLAN configuration includes the following tasks:

(1) Creating and Configuring VLANs

(2) (Optional) Configuring VLAN Names

(3) (Optional) Configuring SVIs

1.4.3 Creating and Configuring VLANs

1. Overview

Create a VLAN, and enter the VLAN configuration mode.

A VLAN is identified based on a VLAN ID, and the VLAN ID range is from 1 to 4094. VLAN 1 is created automatically and cannot be deleted. VLANs 2 to 4094 can be created, deleted, and edited.

2. Restrictions and Guidelines

- When you create a VLAN, the system will prompt the VLAN creation failure if hardware resource is insufficient.
- Run the **unset vlan** *vlan-id* command to delete a common static VLAN. VLANs such as VLAN 1, VLANs with SVIs configured, and private VLANs cannot be deleted.
- To delete a VLAN with a created SVI, run the **unset interface vlan** *vlan-id* command to delete the SVI and then the **unset vlan** *vlan-id* command to delete the VLAN.
- Check created VLANs and its VLAN status. If you enter a new VLAN ID, a corresponding VLAN is created. If you enter an existing VLAN ID, the system enters the configuration mode of the corresponding VLAN, and you can modify the VLAN settings.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a VLAN and enter the VLAN configuration mode.

vlan *vlan-id*

Only VLAN 1 exists by default.

1.4.4 Configuring VLAN Names

1. Overview

Name a VLAN.

2. Restrictions and Guidelines

- You cannot rename a VLAN the same as the default name of another VLAN.
- You are advised to run the **display vlan** command to check all VLAN names in case of name duplication before you name a VLAN.
- A VLAN is named as VLAN*vlan-id*. For example, the name of VLAN 2 is VLAN0002.
- To restore the VLAN name to a default value, run the **unset name** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VLAN configuration mode.

vlan *vlan-id*

(4) Configure a VLAN name.

name *vlan-name*

The name of a VLAN is VLAN *vlan-id* by default.

1.4.5 Configuring SVIs

1. Overview

You can create an SVI of a VLAN and enter the SVI configuration mode. Configure an IP address for the SVI so that devices in the VLAN can communicate at Layer 3.

2. Restrictions and Guidelines

- You cannot create SVIs for sub-VLANs of a super VLAN or secondary VLANs of a private VLAN.
- In the **interface vlan** *interface-number* command, *interface-number* indicates the SVI ID. An SVI ID must follow the ID of the corresponding VLAN, which ranges from 1 to 4,094.
- To delete a VLAN with a created SVI, run the **unset interface vlan** *interface-number* command to delete the SVI and then run the **unset vlan** *vlan-id* command to delete the VLAN.
- To configure an IPv4 address, see *Configuring IPv4 Basics*.
- To configure an IPv6 address, see *Configuring IPv6 Basics*.

3. Prerequisites

Create a VLAN.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an SVI of a VLAN and enter the SVI configuration mode.

interface vlan *interface-number*

No SVI is configured for a VLAN by default.

(4) (Optional) Configure an IP address for an SVI.

ip address { *ipv4-address mask* | *ipv4-address/mask-length* }

No IPv4 address is configured for an SVI by default.

1.5 Configuring Access Ports

1.5.1 Overview

Configure a Layer 2 interface as an access port and add the interface to a VLAN.

1.5.2 Restrictions and Guidelines

- When you run the **switch-port access vlan** *vlan-id* command to add an access port to a VLAN, the command takes effect only on access ports. If you add the access port to an inexistent VLAN, the VLAN is automatically created. Run the **unset switch-port access vlan** command to restore the VLAN for the access port to VLAN 1.
- In the VLAN configuration mode, run the **add interface** { *interface-type interface-number* | **range** *interface-type interface-range* } command to add an access port to the current VLAN. The command takes effect only on access ports and has the same effect as that of the **switch-port access vlan** *vlan-id* command. Run the **unset add interface** { *interface-type interface-number* | **range** *interface-type interface-range* } command to restore the VLAN (other than VLAN 1) of a specified access port to VLAN 1.

1.5.3 Procedure (Interface Configuration Mode)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure the Layer 2 interface as an access port.

switch-port mode access

The Layer 2 interface mode is **access** by default.

The Layer 2 interface mode is **hybrid** by default.

(5) Add an access port to a specified VLAN. The VLAN can be automatically created.

switch-port access vlan *vlan-id*

An access port belongs to VLAN 1 by default.

An access port belongs to no VLAN by default.

1.5.4 Procedure (VLAN Configuration Mode)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure the Layer 2 interface as an access port.

switch-port mode access

(5) Return to the global configuration mode.

exit

(6) Create a VLAN and enter the VLAN configuration mode.

vlan *vlan-id*

Only VLAN 1 exists by default.

(7) Add an access port to a VLAN. The VLAN can be automatically created. Select a method to configure the interface or the group of interfaces.

- Add a Layer 2 Ethernet interface to a VLAN.
  - **add interface** { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

  An access port belongs to VLAN 1 by default.

- Add a Layer 2 link aggregation to a VLAN.

  add interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

1.6 Configuring Trunk Ports

1.6.1 Overview

A trunk link can transmit traffic of multiple VLANs. To enable a port to transmit traffic of multiple VLANs, you can configure the port as a trunk port.

You can configure a Layer 2 Ethernet interface or aggregate interface as a trunk port. For more information about aggregate interface, see *Configuring Aggregate Interface* .

A trunk port can belong to one native VLAN and several allowed VLANs.

- Native VLAN: When the trunk port receives an untagged packet or a priority packet, the trunk port considers that the packet comes from the native VLAN of this port. A priority packet is a tagged packet, but its VLAN ID is 0 and contains only priority information. Packets of the native VLAN sent by the trunk port are untagged. When you configure a trunk link, ensure that the trunk ports of the link are bound to the same native VLAN.
- Allowed VLAN: When the trunk port receives a tagged packet, it determines whether the packet comes from an allowed VLAN of this port. If yes, the trunk port receives the packet. Otherwise, the trunk port discards the packet. Packets of an allowed VLAN sent by the trunk port are tagged.

1.6.2 Restrictions and Guidelines

- Before you configure the native VLAN and allowed VLANs for the trunk port, create these VLANs first. *When you* set the native VLAN to an inexistent VLAN, this VLAN is not created automatically.
- The native VLAN of an interface can be excluded from the allowed VLAN list for this interface. In this case, traffic of the native VLAN cannot pass through this interface.
- Run the **unset switch-port mode** command to restore the Layer 2 interface mode to the default configuration.

  Run the **unset switch-port trunk pvid vlan** command to restore the native VLAN of the trunk port to a default value.

  Run the **unset switch-port trunk allow-pass vlan** command to restore the allowed VLANs of the trunk port to default values.

1.6.3 Procedure

(1) Create a VLAN.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create one VLAN or a group of VLANs as the native VLAN and allowed VLANs of the trunk port.

vlan { *vlan-id* | **range** *vlan-range* }

Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

(2) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(3) Configure the Layer 2 interface as a trunk port.

switch-port mode trunk

(4) (Optional) Configure a native VLAN for the trunk port.

switch-port trunk pvid vlan *vlan-id*

The native VLAN of a trunk port is VLAN 1 by default.

(5) (Optional) Configure an allowed VLAN list for the trunk port.

switch-port trunk allow-pass vlan { **all** | **add** *vlan-list* | **remove** *vlan-list* | **except** *vlan-list* | **only** *vlan-list* }

The allowed VLAN of a trunk port is VLAN 1 by default.

The allowed VLANs of a trunk port are VLANs 1 to 4094 by default.

1.7 Configuring Uplink Ports

1.7.1 Overview

You can configure an interface as an uplink port to transmit only tagged frames of multiple VLANs. An uplink port can belong to one native VLAN and several allowed VLANs.

The uplink port is generally used in QinQ scenarios as an interface for a PE to connect to a public network.

- Native VLAN: When the uplink port receives an untagged packet or a priority packet, the uplink port considers that the packet comes from the native VLAN of this port. Different from the trunk port, packets of the native VLAN sent by the uplink port are tagged.
- Allowed VLAN: When the uplink port receives a tagged packet, it determines whether the packet comes from the native VLAN of this port. If yes, the uplink port receives the packet. Otherwise, the uplink port discards the packet. Packets of the allowed VLANs sent by the uplink port are tagged.

1.7.2 Restrictions and Guidelines

- Before you configure the native VLAN and allowed VLANs for the uplink port, create these VLANs first.
- Run the **unset switch-port mode** command to restore the Layer 2 interface mode to the default configuration.

  Run the **unset switch-port trunk pvid vlan** command to restore the native VLAN of the uplink port to a default value.

  Run the **unset switch-port trunk allow-pass vlan** command to restore the allowed VLANs of the uplink port to default values.
- Run the **display vlan** [**id** *vlan-id*] and **display interface switch-port** commands to check whether the configurations take effect.

1.7.3 Procedure

(1) Create a VLAN.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create one VLAN or a group of VLANs as the native VLAN and allowed VLANs of the uplink port.

vlan { *vlan-id* | **range** *vlan-range* }

Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

(2) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(3) Configure the Layer 2 interface as an uplink port.

switch-port mode uplink

(4) (Optional) Configure a native VLAN for the uplink port.

switch-port trunk pvid vlan *vlan-id*

The native VLAN of an uplink port is VLAN 1 by default.

(5) (Optional) Configure an allowed VLAN list for the uplink port.

switch-port trunk allow-pass vlan { **all** | **add** *vlan-list* | **remove** *vlan-list* | **except** *vlan-list* | **only** *vlan-list* }

The allowed VLANs of the uplink port are VLANs 1 to 4,094 by default.

1.8 Configuring Hybrid Ports

1.8.1 Overview

You can configure an interface as a hybrid port to transmit traffic of multiple VLANs. A hybrid port can belong to one native VLAN and several allowed VLANs.

- Native VLAN: When the hybrid port receives an untagged packet or a priority packet, the hybrid port considers that the packet comes from the native VLAN of this port. Packets of the native VLAN sent by the hybrid port are untagged.
- Allowed VLAN: When the hybrid port receives a tagged packet, it determines whether the packet comes from an allowed VLAN of this port. If yes, the hybrid port receives the packet. Otherwise, the hybrid port discards the packet. Different from the trunk port, the hybrid port can send tagged or untagged packets of the allowed VLANs (other than the native VLAN) based on configurations.

1.8.2 Restrictions and Guidelines

- Before you configure the native VLAN and allowed VLANs for a hybrid port, create these VLANs first.
- Run the **unset switch-port mode** command to restore the Layer 2 interface mode to the default

configuration.

Run the **unset switch-port hybrid pvid vlan** command to restore the native VLAN of the hybrid port to a default value.

Run the **unset switch-port hybrid allow-pass vlan** command to restore the allowed VLANs of the hybrid port to default values.

- Run the **display vlan** [**id** *vlan-id*] and **display interface switch-port** commands to check whether the configurations take effect.

1.8.3 Procedure

(1) Create a VLAN.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create one VLAN or a group of VLANs as the native VLAN and allowed VLANs of the hybrid port.

vlan { *vlan-id* | **range** *vlan-range* }

Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

(2) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(3) Configure the Layer 2 interface as a hybrid port.

switch-port mode hybrid

(4) Configure a native VLAN for the hybrid port.

switch-port hybrid pvid vlan *vlan-id*

The native VLAN of the hybrid port is VLAN 1 by default.

(5) Configure an allowed VLAN list for the hybrid port.

switch-port hybrid allow-pass vlan { [**add**] **tagged** *vlan-list* | [**add**] **untagged** *vlan-list* | **only tagged** *vlan-list* | **remove** *vlan-list* }

By default, the allowed VLANs of a hybrid port are untagged VLAN 1 and tagged VLANs 2 to 4,094.

1.9 Monitoring

This section describes the display commands used for checking the running status of a configured function to verify the configuration effect.

This section also describes the debugging command used for outputting debugging information.

Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging function immediately after use.

You can run the cleanup commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-3 Monitoring

Command	Purpose
display vlan [**vlan-id** *vlan-id*]	Displays VLAN configurations.
display interface switch-port	Displays configurations of Layer 2 interfaces.
debugging bridge vlan	Enables the debugging function of a VLAN.

1.10 Configuration Examples

1.10.1 Configuring Layer 2 Isolation and Layer 3 Interconnection of VLANs

1. Requirements

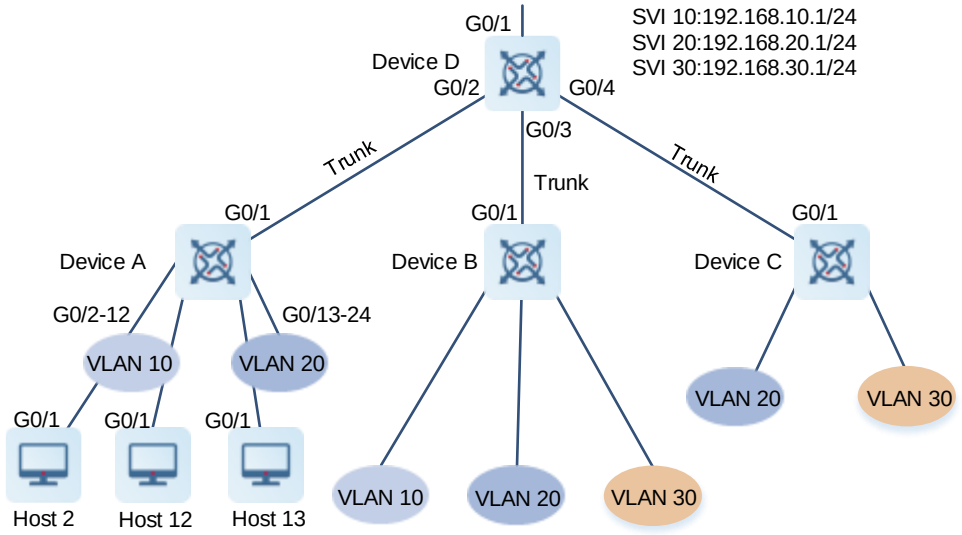
There are three departments in a user's intranet. The networks of the three departments are required to be isolated in Layer 2 and interconnected in Layer 3.

Therefore, create VLANs on the three access devices, Devices A, B, and C. The user intranet is divided into VLAN 10, VLAN 20, and VLAN 30, and they are isolated in Layer 2. Add a port to a VLAN, and all clients connected to this port belong to the VLAN. To change the VLAN of a client, modify the VLAN configuration of the port instead of moving the client.

Configure three VLANs on the core device, Device D. Configure the ports connected to Devices A, B, and C as trunk ports, and specify the allowed VLAN lists. Configure three SVIs as the gateway interfaces of the subnets of the three VLANs. The three subnets corresponding to the VLANs are 192.168.10.0/24, 192.168.20.0/24, and 192.168.30.0/24 respectively. The subnets of the three VLANs are interconnected through IP forwarding of Layer 3 core devices.

2. Topology

Figure 1-16 Configuring VLANs



3. Notes

- On the Layer 3 core device Device D, configure the three VLANs, configure the ports for connecting Devices A, B, and C as trunk ports, and specify an allowed VLAN list to realize Layer 2 isolation.
- Configure three SVIs on Device D, which are used as the gateway interfaces of the IP subnets corresponding to the three VLANs. Configure the IP addresses for these SVIs. Configure a default gateway for hosts based on the VLANs of the hosts.
- Create the VLANs respectively on the three access devices, Devices A, B, and C, assign an access port for each VLAN, and configure the port for connecting Device D as a trunk port.

4. Procedure

(1) Create and configure VLANs.

Create VLANs on Device D and rename VLAN 10.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# vlan 10
DeviceD(cfg-vlan)# name office
DeviceD(cfg-vlan)# vlan range 20,30
DeviceD(cfg-vlan-range)# exit
```

Create a VLAN on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan range 10,20
DeviceA(cfg-vlan-range)# exit
```

Create a VLAN on Devices B.

```
DeviceB> privilege
```

```
DeviceB# config-system terminal
DeviceB(cfg)# vlan range 10,20,30
DeviceB(cfg-vlan-range)# exit
```

Create a VLAN on Devices C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# vlan range 20,30
DeviceC(cfg-vlan-range)# exit
```

(2) Configure IP addresses for the Layer 3 SVIs of the VLANs on Device D.

```
DeviceD(cfg)# interface vlan 10
DeviceD(cfg-if-VLAN 10)# ip address 192.168.10.1 255.255.255.0
DeviceD(cfg-if-VLAN 10)# exit
DeviceD(cfg)# interface vlan 20
DeviceD(cfg-if-VLAN 20)# ip address 192.168.20.1 255.255.255.0
DeviceD(cfg-if-VLAN 20)# exit
DeviceD(cfg)# interface vlan 30
DeviceD(cfg-if-VLAN 30)# ip address 192.168.30.1 255.255.255.0
DeviceD(cfg-if-VLAN 30)# exit
```

(3) Configure the downlink port of Device D as a trunk port and configure an allowed VLAN list for the port.

```
DeviceD(cfg)# interface range gigabitethernet 0/2-4
DeviceD(cfg-if-list)# switch-port
DeviceD(cfg-if-list)# switch-port mode trunk
DeviceD(cfg-if-list)# exit
DeviceD(cfg)# interface gigabitethernet 0/2
DeviceD(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
DeviceD(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan only 10,20
DeviceD(cfg-if-GigabitEthernet 0/2)# exit
DeviceD(cfg)# interface gigabitethernet 0/3
DeviceD(cfg-if-GigabitEthernet 0/3)# switch-port trunk pvid vlan 1
DeviceD(cfg-if-GigabitEthernet 0/3)# switch-port trunk allow-pass vlan only 10,20,30
DeviceD(cfg-if-GigabitEthernet 0/3)# exit
DeviceD(cfg-if)# interface gigabitethernet 0/4
DeviceD(cfg-if-GigabitEthernet 0/4)# switch-port trunk pvid vlan 1
DeviceD(cfg-if-GigabitEthernet 0/4)# switch-port trunk allow-pass vlan only 20,30
DeviceD(cfg-if-GigabitEthernet 0/4)# end
DeviceD# write
```

(4) Configure the uplink port of the access device as a trunk port configure an allowed VLAN list for the port. The following example configures the uplink port on Device A. The configuration on Devices B and C is similar to that on Device A.

Configure the uplink port of Device A as a trunk port and allow the traffic of all VLANs to pass through the port.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
```

```
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

(5) Configure the downlink port of the access device as an access port and add the port to a specified VLAN. The following example configures the downlink port on Device A. The configuration on Devices B and C is similar to that on Device A.

Configure the downlink port of Device A as an access port and add the port to different VLANs.

```
DeviceA(cfg)# interface list gigabitethernet 0/2-12
DeviceA(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode access
DeviceA(cfg-if-list)# switch-port access vlan 10
DeviceA(cfg-if-list)# interface list gigabitethernet 0/13-24
DeviceA(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode access
DeviceA(cfg-if-list)# switch-port access vlan 20
DeviceA(cfg-if-list)# end
DeviceA# write
```

5. Verification

(1) Run the **display vlan** [**vlan-id** *vlan-id*] command on Device D to display VLAN information, including the VLAN ID, name, status and ports.

```
DeviceD# display vlan
VLAN  Name          Status                      Ports
---- --------  --------  --------------------------------
   1 VLAN0001   STATIC      Gi0/1, Gi0/5, Gi0/6, Gi0/7
                            Gi0/8, Gi0/9, Gi0/10, Gi0/11
                            Gi0/12, Gi0/13, Gi0/14, Gi0/15
                            Gi0/16, Gi0/17, Gi0/18, Gi0/19
                            Gi0/20, Gi0/21, Gi0/22, Gi0/23
                            Gi0/24
  10 office     STATIC      Gi0/2, Gi0/3
  20 VLAN0020  STATIC      Gi0/2, Gi0/3, Gi0/4
  30 VLAN0030  STATIC      Gi0/3, Gi0/4
```

(2) Check the port configuration and port status.

Run the **display interface switch-port** command on Device D to display the VLAN status of ports.

```
DeviceD# display interface switch-port
Interface                          Switch-port Mode      Access Native Protected VLAN lists
---------------------------------- ---------- --------- ------ ------ --------- --
GigabitEthernet 0/2                enabled    TRUNK     1      Disabled  10,20
GigabitEthernet 0/3                enabled    TRUNK     1      Disabled  10,20,30
GigabitEthernet 0/4                enabled    TRUNK     1      Disabled  20,30
```

Run the **display interface description** command on Device D to check whether the port status is up.

```
DeviceD# display interface description
Interface                                Status     Administrative Description
```

```
---------------------------------------- -------- -------------- -----------
GigabitEthernet 0/2                      up       up
GigabitEthernet 0/3                      up       up
GigabitEthernet 0/4                      up       up

VLAN 10                                    up       up
VLAN 20                                    up       up
VLAN 30                                    up       up
```

(3) Run the **display ip route-config** command on Device D to display the direct route information of SVIs.

Packets whose destination IP address matches 192.168.10.0/24 are forwarded to VLAN 10. Packets whose destination IP address matches 192.168.20.0/24 are forwarded to VLAN 20. Packets whose destination IP address matches 192.168.30.0/24 are forwarded to VLAN 30.

```
DeviceD# display ip route-config
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPF, B - BGP, I - IS-IS, V - Overflow route
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area, EV - BGP EVPN, A - Arp to host
         LA - Local aggregate route
         * - candidate default

Gateway of last resort is no set
…
C       192.168.10.0/24 is directly connected, VLAN 10
C       192.168.10.1/32 is local host.
C       192.168.20.0/24 is directly connected, VLAN 20
C       192.168.20.1/32 is local host.
C       192.168.30.0/24 is directly connected, VLAN 30
C       192.168.30.1/32 is local host.
```

Ping the SVI address of a VLAN (for example, VLAN 10) on Device D.

```
DeviceD# ping 192.168.10.1
Sending 5, 100-byte ICMP Echoes to 192.168.10.1, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

(4) The following example describes how to ping Hosts 2, 12, and 13.

Configure the default gateway of Host 2 in VLAN 10 as 192.168.10.1, the IP address of the interface as 192.168.10.2, and the mask as 255.255.255.0.

```
Host2> privilege
Host2# config-system terminal
Host2(cfg)# ip route-config 0.0.0.0 0.0.0.0 192.168.10.1
Host2(cfg)# interface gigabitethernet 0/1
Host2(cfg-if-GigabitEthernet 0/1)# unset switch-port
```

```
Host2(cfg-if-GigabitEthernet 0/1)# ip address 192.168.10.2 255.255.255.0
```

Configure the default gateway of Host 12 in VLAN 10 as 192.168.10.1, the IP address of the interface as 192.168.10.12, and the mask as 255.255.255.0.

```
Host2> privilege
Host2# config-system terminal
Host2(cfg)# ip route-config 0.0.0.0 0.0.0.0 192.168.10.1
Host2(cfg)# interface gigabitethernet 0/1
Host2(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host2(cfg-if-GigabitEthernet 0/1)# ip address 192.168.10.12 255.255.255.0
```

Configure the default gateway of Host 13 in VLAN 20 as 192.168.20.1, the IP address of the interface as 192.168.20.13, and the mask as 255.255.255.0.

```
Host13> privilege
Host13# config-system terminal
Host13(cfg)# ip route-config 0.0.0.0 0.0.0.0 192.168.20.1
Host13(cfg)# interface gigabitethernet 0/1
Host13(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host13(cfg-if-GigabitEthernet 0/1)# ip address 192.168.20.13 255.255.255.0
```

Run the **display ip route-config** command on the hosts to display the default gateway. The default gateway must be configured on the transmit and receive ends. Inter-VLAN communication fails to be implemented if either end is not configured with the default gateway. Take Host 2 as an example. Due to the default route S*0.0.0.0/0 [1/0] via 192.168.10.1, a host sends a packet to the gateway 192.168.10.1 when the host tries to communicate across VLANs.

```
Host2# display ip route-config
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPF, B - BGP, I - IS-IS, V - Overflow route
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area, EV - BGP EVPN, A - Arp to host
         LA - Local aggregate route
         * - candidate default

Gateway of last resort is 192.168.10.1 to network 0.0.0.0
S*      0.0.0.0/0 [1/0] via 192.168.10.1
C       192.168.10.0/24 is directly connected, GigabitEthernet 0/1
C       192.168.10.2/32 is local host.
```

Ping the IP address of a host (for example, Host 2) on Device D.

```
DeviceD# ping 192.168.10.2
Sending 5, 100-byte ICMP Echoes to 192.168.10.2, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/3/4 ms.
```

Hosts in the same VLAN (for example, VLAN 10) can ping each other.

```
Host2# ping 192.168.10.12
Sending 5, 100-byte ICMP Echoes to 192.168.10.12, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms.
```

Hosts in different VLANs can also ping each other. If you delete the IP address of an SVI on Device D or delete the default gateway on a host, hosts in different VLANs cannot ping each other.

```
Host2# ping 192.168.20.13
Sending 5, 100-byte ICMP Echoes to 192.168.10.3, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms.
```

6. Configuration Files

- Device D configuration file

```
vlan 10
name office
vlan range 1,20,30
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 10,20
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 10,20,30
!
interface GigabitEthernet 0/4
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 20,30
!
interface VLAN 10
 ip address 192.168.10.1 255.255.255.0
interface VLAN 20
 ip address 192.168.20.1 255.255.255.0
interface VLAN 30
 ip address 192.168.30.1 255.255.255.0
```

- Device A configuration file. Similar configurations of interfaces are represented by ellipsis (...). For the configuration of Devices B and C, see the Device A configuration procedure.

```
interface GigabitEthernet 0/1
 switch-port mode trunk
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode access
 switch-port access vlan 10
…
interface GigabitEthernet 0/12
 switch-port
 switch-port mode access
 switch-port access vlan 10
!
interface GigabitEthernet 0/13
 switch-port
 switch-port mode access
 switch-port access vlan 20
…
interface GigabitEthernet 0/24
 switch-port
 switch-port mode access
 switch-port access vlan 20
```

7. Common Errors

- If no IP address is configured for a gateway (SVI), or no default gateway is configured for an STA, or either the transmit or receive end is not configured with the default gateway, STAs in different VLANs cannot ping each other.
- If connected trunk ports have different native VLANs, data communication may fail.
- If interfaces connected to STAs are not configured as access ports, data communication may fail.
- If only one VLAN needs to be configured for a trunk port but the default allowed VLAN of the trunk port is not deleted after the VLAN creation command is run, the configuration result is inconsistent with the configuration plan.
- The trunk port of the local device sends untagged packets of the native VLAN. If the trunk port of a device produced by another manufacture only receive tagged packets, the untagged packets sent by the local device may be discarded on the receiving end. In this case, on the local device, configure the port connected to the device of another manufacture as an uplink port so that packets sent by this port are tagged.

Contents

1 Configuring MAC VLAN 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Background and Function 1
1.1.3 Principles 1
1.1.4 MAC VLAN Rule 2
1.1.5 Protocols and Standards 3
1.2 Configuring MAC VLAN 3
1.2.1 Overview 3
1.2.2 Restrictions and Guidelines 3
1.2.3 Prerequisites 4
1.2.4 Procedure 4
1.3 Monitoring 5
1.4 Configuration Examples 6
1.4.1 Configuring MAC VLAN 6

1 Configuring MAC VLAN

Note

MAC VLAN is supported by only the S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, or S6250-48XS8CQ.

1.1 Introduction

1.1.1 Overview

The media access control (MAC) virtual local area network (VLAN) is a VLAN assignment technology based on MAC addresses. When the physical location of an STA changes and causes a change in the access port of the STA, MAC VLAN can assign the STA to a specified VLAN based on the MAC address of the STA, freeing you to reconfigure a VLAN for the STA's current access port.

1.1.2 Background and Function

1. Background

As mobile office is popularized, an STA connects to a network through different ports. It may connect to a network through port 1 at a time and then connect to the network through port 2 at another time. If VLANs are assigned based on ports and port 1 and port 2 belong to different VLANs, when the STA connects to the network in the second time through port 2, it is assigned to a different VLAN. In this case, the STA cannot use the resources in the VLAN to which port 1 belongs. If port 1 and port 2 belong to the same VLAN, when a second STA is connected through port 2, security issues may arise. To enable hosts of different VLANs to connect to a network through the same port, the MAC VLAN function is developed.

2. Function

The MAC VLAN function is to assign STAs to VLANs based on MAC addresses of the STAs. MAC VLAN is a VLAN assignment method based on users' MAC addresses.

This function is generally used in combination with the VLAN delivery function of IEEE 802.1X. After you configure or deliver MAC VLAN entries for a Layer 2 switch, VLANs can be assigned based on MAC addresses of STAs, thereby ensuring secure and flexible access of IEEE 802.1X-compliant STAs.

1.1.3 Principles

The MAC VLAN function must be configured on hybrid ports, and a MAC VLAN rule takes effect on all hybrid ports, on which the MAC VLAN function is enabled.

- The MAC VLAN function is available to untagged packets only. When a hybrid port receives an untagged packet, the port tries to match the source MAC address in the packet with the specified MAC address in the MAC VLAN rule.
  a. A MAC VLAN rule has a higher priority than the voice VLAN, subnet-based VLAN, protocol-based VLAN, and port VLAN rules. If the MAC VLAN function is enabled, when receiving an untagged packet,

the port matches the source MAC address in the packet with the specified MAC address in the MAC VLAN rule to check whether the packet complies with the MAC VLAN rule.

- If they match, the port changes the VLAN in the packet to the specified VLAN in the MAC VLAN rule. Then, the subsequent processing of the packet is based on this specified VLAN. For example, after an untagged packet matches a MAC VLAN rule, the VLAN learned based on the MAC address is the one redirected according to the MAC VLAN rule.

b. If they do not match, the port will match the packet against VLAN assignment rule based on packet content. The rule involves the voice VLAN, subnet-based VLAN and protocol-based VLAN. If the untagged packet matches both the MAC VLAN rule and packet content-based VLAN assignment rules, the packet is redirected to the VLAN specified in the MAC VLAN rule.

- If the port fails to match the packet with any packet content-based VLAN assignment rule, the port, according to the assignment rule of the port VLAN (common VLAN), considers that the packet belongs to the native VLAN of this port, receives the packet, and marks the packet with the native VLAN tag.

- The MAC VLAN function is unavailable to tagged packets.
- When a port receives a tagged packet, the port follows the frame receiving rules for common VLANs, and receives the packet if the port allows the packet to pass. Otherwise, the port discards the packet.
- The MAC VLAN function is unavailable to priority packets, either. In a priority packet, the VLAN ID in the VLAN tag is 0, indicating that the packet does not carry VLAN information and carries only a Class of Service (CoS) priority.

1.1.4 MAC VLAN Rule

- Configure a static MAC VLAN rule by running commands on the command line interface (CLI).

  A MAC VLAN rule reflects the mapping between a MAC address and a VLAN.

  You can configure a mapping between a MAC address and a VLAN on the CLI on a local device. Run the **display mac-vlan** command to display a MAC VLAN rule. The rule marked with "S" (static) indicates a static MAC VLAN rule.

- Configure a dynamic MAC VLAN rule by using an authentication server.

  The VLAN delivery function of IEEE 802.1X can automatically configure and deliver a mapping between a MAC address and a VLAN. Run the **display mac-vlan** command. The rule marked with "D" (dynamic) indicates a dynamic MAC VLAN rule.

  After an STA passes IEEE 802.1X authentication, the device creates a dynamic mapping between the STA's MAC address and a VLAN based on the information provided by the authentication server. After the STA goes offline, the device automatically deletes the mapping immediately.

  To use this method, configure mappings between MAC addresses and VLANs. For more information about the VLAN delivery function of IEEE 802.1X, see *Configuring IEEE 802.1X*.

- You can configure MAC VLAN rules on the local device and authentication server. If the MAC VLAN rules are consistent, they take effect concurrently. If they are inconsistent, the rule configured earlier takes effect and the rule configured later can be displayed but does not take effect. To facilitate management, you are advised to adopt one configuration method.

Caution

- If the fail VLAN function is configured in the IEEE 802.1X function, when a user fails to pass the authentication, the user packet is assigned to the fail VLAN, for example, VLAN 100. However, the static MAC VLAN rule (VLAN 200) has a higher priority. Therefore, the user packet is redirected to VLAN 200. As a result, the user cannot make communication in the fail VLAN 100. To use the fail VLAN function of the IEEE 802.1X function, do not configure a static MAC VLAN rule on the device.

1.1.5 Protocols and Standards

IEEE 802.1Q: IEEE Standard for Local and Metropolitan Area Networks-Virtual Bridged Local Area Networks

1.2 Configuring MAC VLAN

1.2.1 Overview

You can configure a static MAC VLAN entry, create a mapping between a MAC address and a VLAN, and then configure a MAC VLAN on a port so that the MAC VLAN entry takes effect on the port.

1.2.2 Restrictions and Guidelines

- The MAC VLAN function takes effect on a hybrid port only. Only a Layer 2 interface can be configured as a hybrid port.
  - You can run the **display interface switch-port** command to display the interface type. If the **Switch-port** field corresponding to the interface is **enabled**, this interface is a Layer 2 interface. If the **Switch-port** field corresponding to the interface is **disabled**, this interface is a Layer 3 interface.
  - You can run the **switch-port** command to configure the interface as a Layer 2 interface. Run the **unset switch-port mode** command to restore the Layer 2 interface mode to the default mode and run the **unset switch-port** command to configure a Layer 2 interface as a Layer 3 interface. When you configure an interface as a Layer 2 or Layer 3 interface, all attributes of the interface are restored to the default values in this interface mode.
- Run the **mac-vlan** command to configure static MAC VLAN entries, run the **unset** *mac*-**vlan** to delete static MAC VLAN entries.
- **mac-***address mac-address* indicates a MAC address. The MAC address specified in a MAC VLAN entry must be a unicast address.
- **mask** *mac-mask* indicates the mask of a MAC address. When a port with the MAC VLAN function enabled receives a packet and the packet matches a MAC VLAN entry whose mask is all F and a MAC VLAN entry whose mask is not all F, the port forwards the packet based on the latter one (indicating a larger range of MAC addresses).
  - Run the **unset mac-vlan mac-address** *mac-address* [**mask** *mac-mask*] command to delete static MAC VLAN entries of a specified MAC address.
- **vlan** *vlan-id* indicates the VLAN, to which a packet is forwarded if the MAC address in the packet matches the MAC VLAN rule. Before you configure a static MAC VLAN entry or a dynamic MAC VLAN entry is generated, make sure that the specified VLAN already exists. The specified VLAN in a MAC VLAN entry cannot be a super VLAN, a remote VLAN, or a primary VLAN. However, it can be a sub VLAN or a secondary VLAN. If an untagged packet matches both a MAC VLAN rule and a voice VLAN rule, the packet

is redirected to the VLAN in the MAC VLAN rule. Run the **unset mac-vlan vlan** *vlan-id* command to delete the static MAC VLAN entries of a specified VLAN.

- **priority** *priority-value* indicates the 802.1p priority. The default value is **0**. When a packet matches a MAC VLAN rule, the receiving port of the packet changes the CoS priority of the packet to the value of *priority-value*. If the packet priority needs to be modified, ensure that the Quality of Service (QoS) module trusts the packet priority modified by an interface. Otherwise, the modification is invalid. If an untagged packet matches both a MAC VLAN rule and a voice VLAN rule, the packet priority is modified according to the voice VLAN rule.

  The packet trust mode of the QoS module is disabled by default. This causes the QoS module to change the priority of all packets to **0**, and overwrites the packet priority modified by the MAC VLAN function. Run the **mls qos trust** { **cos** | **ip-precedence** | **dscp** | **exp** } command in interface configuration mode, and select the **cos** parameter to enable the QoS module to trust the CoS modification, that is, the packet priority modified by the MAC VLAN function.

- Run the **unset mac-vlan all** command to delete all MAC VLAN entries.
- Run the **mac-vlan enable** command to enable the MAC VLAN function on ports so that the function can take effect. Otherwise, the MAC VLAN function does not take effect even if there are global MAC VLAN entries.
- You must complete interface configuration before you enable the MAC VLAN function on a port. Otherwise, the allowed VLAN list of the port cannot be modified.

1.2.3 Prerequisites

1.2.4 Procedure

(1) Create a MAC VLAN.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create a MAC VLAN.

vlan { *vlan-id* | **range** *vlan-range* }

By default, no VLAN other than VLAN 1 exists.

d Exit the VLAN configuration mode.

exit

(2) Configure a static MAC VLAN rule in global configuration mode, and specify a mapping between a MAC address and a VLAN in the rule.

mac-vlan mac-address *mac-address* [**mask** *mac-mask*] **vlan** *vlan-id* [**priority** *priority-value*]

No static MAC VLAN rule exists by default.

a Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.
  - **interface** { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.
  - **interface** { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

b Configure the Layer 2 interface as a hybrid port.

switch-port mode hybrid

The default Layer 2 interface mode is the access mode.

c Add the MAC VLAN to the allowed untagged VLAN list of the hybrid port.

switch-port hybrid allow-pass vlan untagged *vlan_list*

The allowed untagged VLAN of the hybrid port is VLAN 1 by default.

(3) Configure the hybrid port to trust the CoS priority of packets.

mls qos trust cos

The default trust mode of a port is untrusted.

(4) Enable the MAC VLAN function on the port.

mac-vlan enable

The port-based MAC VLAN function is disabled by default.

1.3 Monitoring

This section describes the **display** commands used for checking the running status of a configured function to verify the configuration effect.

This section also describes the **debugging** command used for outputting debugging information.

⚠ Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging function immediately after use.

You can run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
display mac-vlan all	Displays all MAC VLAN entries, including static and dynamic MAC VLAN entries.
display mac-vlan dynamic	Displays dynamic MAC VLAN entries.
display mac-vlan static	Displays static MAC VLAN entries.
display mac-vlan vlan vlan-id	Displays MAC VLAN entries of a specified VLAN.

Command	Purpose
display mac-vlan interface	Displays ports on which the MAC VLAN function is enabled.
display mac-vlan mac-address *mac-address* [**mask** *mac-mask*]	Displays MAC VLAN entries of a specified MAC address.
debugging bridge mvlan	Debugs the MAC VLAN function.

1.4 Configuration Examples

1.4.1 Configuring MAC VLAN

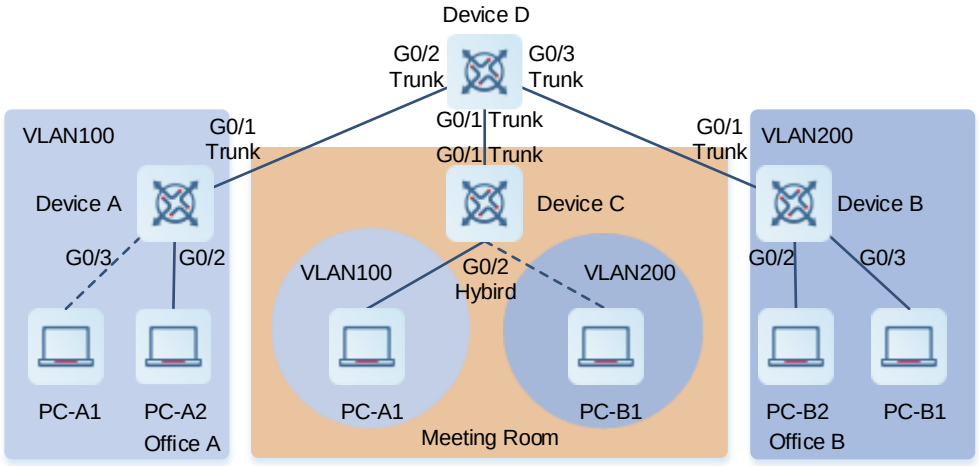
1. Requirements

A company provide a temporary office in a meeting room where public access ports are assigned to VLANs based on the MAC addresses of employees' portable PCs. Employee PCs connected to any port are automatically assigned to their respective department VLANs.

2. Topology

PC-A1 and PC-A2 belong to department A and they are assigned to VLAN 100. PC-B1 and PC-B2 belong to department B and they are assigned to VLAN 200. The meeting room uses public ports. PC-A1 and PC-A2 can be assigned only to VLAN 100 after access and PC-B1 and PC-B2 can be assigned only to VLAN 200.

Figure 1-1 Configuring MAC VLAN



3. Notes

- On Device D, configure the ports connecting to access devices as trunk ports.
- Create MAC VLANs on Device C and configure static MAC VLAN rules in global configuration mode.
  - Department A uses a PC with the MAC address of 54bf.6450.dd10 for access in the meeting room, and the PC is assigned to VLAN 100 with a priority of 1.
  - Department B uses multiple PCs for access in the meeting room. Their MAC addresses are in the range from 54bf.6450.dd20 to 54bf.6450.dd2F (represented with 54bf.6440.dd20 plus the mask of ffff.ffff.fff0).

The PCs are assigned to VLAN 200 with a priority of 2.

- On Device C, configure the ports connecting to PCs as hybrid ports, and add the MAC VLANs to the allowed untagged VLAN list of the ports. Configure the ports to trust the CoS priority of packets. Enable the MAC VLAN function on the ports.
- Configure the uplink ports on Device A and Device B as trunk ports, and the ports connecting to PCs as access ports. Add the access ports on Device A to VLAN 100, and add the access ports on Device B to VLAN 200.

4. Procedure

(1) Create VLANs.

Configure VLAN 100 and VLAN 200 on Device D.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# vlan range 100,200
DeviceD(cfg-vlan-range)# exit
```

Configure VLAN 100 and VLAN 200 on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# vlan rang 100,200
DeviceC(cfg-vlan-range)# exit
```

Configure VLAN 100 on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan 100
DeviceA(cfg-vlan-range)# exit
```

Configure VLAN 200 on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan 200
DeviceB(cfg-vlan-range)# exit
```

(2) Configure trunk ports.

Configure the downlink ports GigabitEthernet 0/1–3 of Device D as trunk ports.

```
DeviceD(cfg)# interface list gigabitethernet 0/1-3
DeviceD(cfg-if-list)# switch-port
DeviceD(cfg-if-list)# switch-port mode trunk
DeviceD(cfg-if-list)# switch-port trunk pvid vlan 1
DeviceD(cfg-if-list)# switch-port trunk allow-pass vlan all
DeviceD(cfg-if-list)# end
DeviceD# write
```

Configure the uplink ports GigabitEthernet 0/1 of Devices A/B/C as trunk ports. The configurations on Devices A, B, and C are similar. The following examples takes Device A as an example.

```
DeviceA(cfg)# interface gigabitethernet 0/1
```

```
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

(3) Add user ports to static VLANs on Devices A and B.

Add the user ports GigabitEthernet 0/2–3 on Device A to VLAN 100.

```
DeviceA(cfg)# interface list gigabitethernet 0/2-3
DeviceA(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode access
DeviceA(cfg-if-list)# switch-port access vlan 100
DeviceA(cfg-if-list)# end
DeviceA# write
```

Add the user ports GigabitEthernet 0/2–3 on Device B to VLAN 200.

```
DeviceB(cfg)# interface list gigabitethernet 0/2-3
DeviceB(cfg-if-list)# switch-port
DeviceB(cfg-if-list)# switch-port mode access
DeviceB(cfg-if-list)# switch-port access vlan 200
DeviceB(cfg-if-list)# end
DeviceB# write
```

(4) Add the user port GigabitEthernet 0/2 on Device C to VLANs based on the MAC addresses of PCs.

Create a static MAC VLAN rule in global configuration mode.

```
DeviceC(cfg)# mac-vlan mac-address 54bf.6450.dd10 vlan 100 priority 1
DeviceC(cfg)# mac-vlan mac-address 54bf.6450.dd20 mask ffff.ffff.fff0 vlan 200 priority 2
```

Configure the downlink port GigabitEthernet 0/2 as a hybrid port, and add the MAC VLANs to the allowed untagged VLAN list of the port.

```
DeviceC(cfg)# interface gigabitethernet 0/2
DeviceC(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/2)# switch-port mode hybrid
DeviceC(cfg-if-GigabitEthernet 0/2)# switch-port hybrid pvid vlan 1
DeviceC(cfg-if-GigabitEthernet 0/2)# switch-port hybrid allow-pass vlan add untagged 1,100,200
```

Configure the port to trust the CoS priority of packets.

```
DeviceC(cfg-if-GigabitEthernet 0/2)# mls qos trust cos
```

Enable the MAC VLAN function on the port.

```
DeviceC(cfg-if-GigabitEthernet 0/2)# mac-vlan enable
DeviceC(cfg-if-GigabitEthernet 0/2)# end
DeviceC# write
```

5. Verification

(1) Run the **display mac-vlan interface** command to display ports on which the MAC VLAN function is enabled.

```
DeviceC# display mac-vlan interface
```

```
MAC VLAN is enabled on following interface:
---------------------------------------
GigabitEthernet 0/2
```

(2) Run the **display interface** command to verify that packets sent by port GigabitEthernet 0/2 to the MAC VLAN do not carry tags.

```
DeviceC(cfg-if-GigabitEthernet 0/2)# display interface gigabitethernet 0/2
  Bridge attributes:
    Port-type: hybrid
    Tagged vlan id: 2-99,101-199,201-4094
Untagged vlan id: 1,100,200
```

(3) Check static MAC VLAN entries on Device C.

```
DeviceC# display mac-vlan static
The following MAC VLAN address exist:
S: Static     D: Dynamic
MAC ADDR                MASK                VLAN ID     PRIO     STATE
-------------------------------------------------------
54bf.6450.dd10          ffff.ffff.ffff      100         1        S
54bf.6450.dd20          ffff.ffff.fff0      200         2        S
Total MAC VLAN address count: 2
```

(4) To verify that the MAC VLAN function assigns STAs to different VLANs to implement Layer 2 isolation, assign PC-A1, PC-A2, and PC-B2 to the same network segment and ping the STAs. Only STAs in the same VLAN can be pinged successfully.

Connect PC-A2 to the network through office A, assign PC-A2 to VLAN 100, and set its IP address and mask to 192.168.5.12/24.

```
PC-A2> privilege
PC-A2# config-system terminal
PC-A2(cfg)# interface gigabitethernet 0/1
PC-A2(cfg-if-GigabitEthernet 0/1)# unset switch-port
PC-A2(cfg-if-GigabitEthernet 0/1)# ip address 192.168.5.12/24
```

Connect PC-B2 to the network through office B, assign PC-B2 to VLAN 200, and set its IP address and mask to 192.168.5.22/24.

```
PC-B2> privilege
PC-B2# config-system terminal
PC-B2(cfg)# interface gigabitethernet 0/1
PC-B2(cfg-if-GigabitEthernet 0/1)# unset switch-port
PC-B2(cfg-if-GigabitEthernet 0/1)# ip address 192.168.5.22/24
```

Connect PC-A1 to the network through the meeting room and set its IP address and mask to 192.168.5.11/24. Verify that PC-A1 can ping PC-A2 in VLAN 100 but cannot ping PC-B2 in VLAN 200. This indicates that PC-A1 is assigned to VLAN 100 by the MAC VLAN function.

```
PC-A1> privilege
PC-A1# config-system terminal
PC-A1(cfg)# interface gigabitethernet 0/1
PC-A1(cfg-if-GigabitEthernet 0/1)# unset switch-port
```

```
PC-A1(cfg-if-GigabitEthernet 0/1)# ip address 192.168.5.11/24
PC-A1(cfg-if-GigabitEthernet 0/1)# end
PC-A1# ping 192.168.5.12
Sending 5, 100-byte ICMP Echoes to 192.168.5.12, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/22/103 ms.
PC-A1# ping 192.168.5.22
Sending 5, 100-byte ICMP Echoes to 192.168.5.22, timeout is 2 seconds:
   < press Ctrl+C to break >
....
Success rate is 0 percent (0/5).
```

6. Configuration Files

- Device C configuration file

```
vlan range 1,100,200
!
mac-vlan mac-address 54bf.645c.dd10 vlan 100 priority 1
mac-vlan mac-address 54bf.6450.dd20 mask ffff.ffff.fff0 vlan 200 priority 2
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode hybrid
 switch-port hybrid pvid vlan 1
 switch-port hybrid allow-pass vlan only tagged 2-99,101-199,201-4094
 switch-port hybrid allow-pass vlan add untagged 1,100,200
 mls qos trust cos
 mac-vlan enable
```

- Device A configuration file

```
vlan range 1,100
!
interface GigabitEthernet 0/1
   switch-port
switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
```

```
 switch-port mode access
 switch-port access vlan 100
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode access
 switch-port access vlan 100
```

- Device B configuration file

```
vlan range 1,200
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
```

- Device D configuration file

```
vlan range 1,100,200
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/3
 switch-port
switch-port mode trunk
```

```
switch-port trunk pvid vlan 1
switch-port trunk allow-passvlan only 1-4094
```

7. Common Errors

- An Ethernet interface or aggregated port is not configured as a Layer 2 interface before the MAC VLAN function is configured on the interface.
- The QoS module does not trust the packet priority modified by an interface. As a result, the MAC VLAN priority configuration does not take effect.

Contents

1 Configuring Protocol-based VLAN 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Subnet-based VLAN 1
1.1.3 Protocol-based VLAN 1
1.1.4 Principles 1
1.1.5 Protocols and Standards 2
1.2 Restrictions and Guidelines 2
1.3 Configuration Task Summary 2
1.4 Configuring the Subnet-based VLAN Function 3
1.4.1 Overview 3
1.4.2 Restrictions and Guidelines 3
1.4.3 Prerequisites 3
1.4.4 Procedure 3
1.5 Configuring the Protocol-based VLAN Function 4
1.5.1 Overview 4
1.5.2 Restrictions and Guidelines 5
1.5.3 Prerequisites 6
1.5.4 Procedure 6
1.6 Monitoring 7
1.7 Configuration Examples 7
1.7.1 Configuring Subnet-based VLANs 7

1.7.2 Configuring Protocol-based VLANs 12

1 Configuring Protocol-based VLAN

Note

Protocol-based VLAN is supported by only the S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ.

1.1 Introduction

1.1.1 Overview

Protocol-based VLAN is a VLAN technology based on protocol types. When a device receives a packet without a VLAN ID from a port, the device automatically forwards the packet to a VLAN specified in a user-defined rule if the packet matches the user-defined rule.

The protocol VLAN is divided into subnet-based VLAN and protocol-based VLAN.

1.1.2 Subnet-based VLAN

VLAN assignment based on source IP address and subnet mask in a packet is referred to as subnet-based VLAN.

The subnet-based VLAN function assigns users to different VLANs based on their IP addresses. For example, in a shared office area, IP addresses of users are fixed, but the access ports are dynamic. To control user access to network resources, you can assign users to different VLANs based on their IP addresses.

Caution

Subnet-based VLANs take effect on trunk and hybrid ports, available to packets without VLAN ID only.

1.1.3 Protocol-based VLAN

VLAN assignment based on packet type and Ethernet type is referred to as protocol-based VLAN.

The protocol-based VLAN function is used to assign packets of different service types to different VLANs to facilitate management and maintenance.

Caution

Protocol-based VLANs take effect on trunk and hybrid ports. This function is available to packets without VLAN ID only.

1.1.4 Principles

When a port receives a packet, the port assigns the packet to a VLAN so that the packet belongs to this VLAN only. The procedure of determining a VLAN for a packet is as follows:

(1) If the packet carries a tag and the tag contains a VLAN ID, the packet belongs to the VLAN specified by the VLAN ID. The port receives the packet if the port allows the packet to pass. Otherwise, the port discards the packet.

(2) If the packet does not carry a VLAN ID (untagged packet or a priority packet with the VLAN ID of 0), the port tries to match the packet with VLAN assignment policies based on priority in descending order. If the packet matches a rule, the packet is assigned to the VLAN specified in the rule. If the packet matches multiple rules at the same time, the packet is forwarded based on the rule with the highest priority.

a A MAC VLAN policy has the highest priority. A packet with the MAC address matching a MAC VLAN rule is assigned to the MAC VLAN specified in the rule. The MAC VLAN policy is unavailable to priority packets.

b If a voice VLAN policy exists, a device identifies a voice packet based on the MAC address or voice organizationally unique identification (OUI) and then forwards the packet to the specified voice VLAN.

c If the IP address in a packet matches an IP address in a rule of a subnet-based VLAN policy, the packet is assigned to the subnet-based VLAN specified in the rule.

d If the packet type and Ethernet type of a packet match those configured in a rule of a protocol-based VLAN policy, the packet is assigned to the protocol-based VLAN specified in the rule.

(3) A packet is assigned to the port VLAN (access port) or native VLAN (trunk/hybrid/uplink port) in any one of the following cases: The policy-based VLAN assignment function is disabled, no policy exists, or the packet does not match any policy.

1.1.5 Protocols and Standards

- IEEE 802.1Q: IEEE Standard for Local and Metropolitan Area Networks-Virtual Bridged Local Area Networks

1.2 Restrictions and Guidelines

- The subnet-based VLAN and protocol-based VLAN function only take effect on a Layer 2 interface and the Layer 2 interface must be a trunk or hybrid port.
- You can run the **display interface switch-port** command to display the interface type. If the **Switch-port** field corresponding to the interface is enabled, this interface is a Layer 2 interface. If the **Switch-port** field corresponding to the interface is disabled, this interface is a Layer 3 interface.
- You can run the **switch-port** command to configure the interface as a Layer 2 interface. Run the **unset switch-port mode** command to restore the Layer 2 interface mode to the default mode and run the **unset switch-port** command to configure a Layer 2 interface as a Layer 3 interface. When you configure an interface as a Layer 2 or Layer 3 interface, all attributes of the interface are restored to the default values in this interface mode.

1.3 Configuration Task Summary

Protocol-based VLAN configuration include the following optional tasks. Select the configuration tasks as required.

- Configuring Subnet-based VLAN
- Configuring Protocol-based VLAN

1.4 Configuring Subnet-based VLAN

1.4.1 Overview

You can configure VLAN assignment based on IP addresses so that packets from a specified network segment or a specified IP address are forwarded to the specified VLAN.

1.4.2 Restrictions and Guidelines

- The subnet-based VLAN function takes effect on trunk or hybrid ports only. You must configure a port as a trunk or hybrid port and then run the **protocol-vlan ipv4** command to configure the subnet-based VLAN function.
- To enable the subnet-based VLAN function on a trunk or hybrid port, add all subnet-based VLANs to the allowed VLAN list of the trunk or hybrid port.
- You can enable the subnet-based VLAN function in interface configuration mode without global subnet-based VLAN configuration. However, you must enable the subnet-based VLAN function in both the global and interface configuration mode so that the subnet-based VLAN function takes effect.

1.4.3 Prerequisites

Configure a Layer 2 interface by running the **switch-port** command.

1.4.4 Procedure

(1) Configure subnet-based VLAN rules in global configuration mode.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create a VLAN by configuring a policy.

vlan { *vlan-id* | **range** *vlan-range* }

Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

e Configure subnet-based VLAN rules one by one in global configuration mode, and assign a specified subnet to the specified VLAN.

protocol-vlan ipv4 *ipv4-address* **mask** *mask* **vlan** *vlan-id*

No global subnet-based VLAN rule is configured by default.

(2) Enter the interface configuration mode and configure subnet-based VLAN.

a Enter the Layer 2 Ethernet interface configuration mode.

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

o Enter the Layer 2 aggregate interface configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(3) Configure the port as a trunk or hybrid port.

switch-port mode { **trunk | hybrid** }

The default Layer 2 interface mode is the access mode.

(4) Add subnet-based VLANs to the allowed VLAN list of an interface. Configure one of the following tasks.

- Configure an allowed VLAN list for the trunk port.

  switch-port trunk allow-pass vlan { **all** | **add** *vlan-list* | **remove** *vlan-list* | **except** *vlan-list* | **only** *vlan-list* }

  The default allowed VLANs of the trunk port are VLANs 1 to 4094.

- Configure an allowed VLAN list for the hybrid port.

  switch-port hybrid allow-pass vlan { [**add**] **tagged** *vlan-list* | [**add**] **untagged** *vlan-list* | **only tagged** *vlan-list* | **remove** *vlan-list* }

  The default allowed VLANs of the hybrid port are untagged VLAN 1 and tagged VLANs 2 to 4094.

(5) Enable the subnet-based VLAN function on the interface.

protocol-vlan ipv4

The subnet-based VLAN function is disabled by default.

1.5 Configuring Protocol-based VLAN

1.5.1 Overview

Protocol-based VLAN assigns packets of different service types to different VLANs to facilitate management and maintenance. Figure 1-1 shows the frame format.

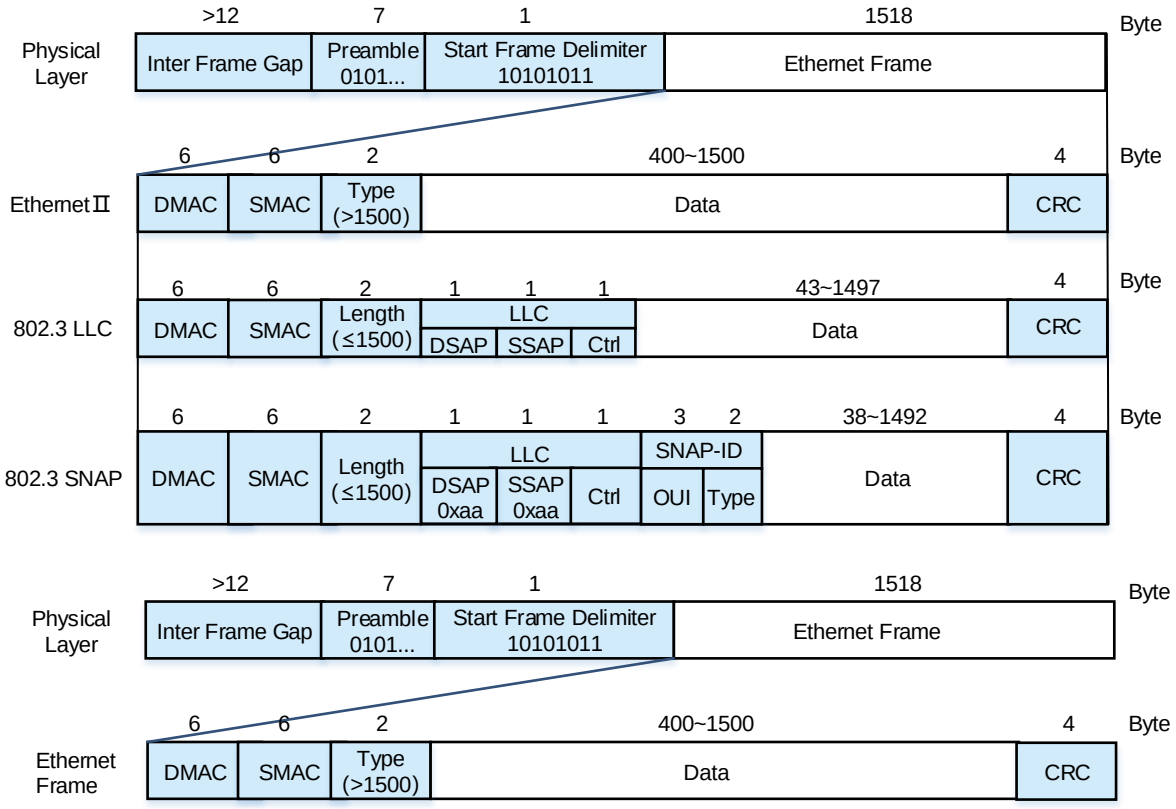
This function uses **frame-type** to indicate a frame type, and supports the **EtherII**, **SNAP** and **LLC** frame types.

If the frame type is **EtherII** or **SNAP**, the **ether-type** *type* parameter is used to configure the Ethernet packet type.

If the frame type is **LLC**, the **dsap** *value* **ssap** *value* parameter is used to configure the type of a service access point (SAP).

This function uses the **ether-type** *type* parameter to configure the Ethernet packet type.

Figure 1-1 Ethernet Frame Format



1.5.2 Restrictions and Guidelines

- The protocol-based VLAN function takes effect on trunk or hybrid ports only. You must configure an interface as a trunk or hybrid port and then run the **protocol-vlan profile vlan** command to configure the protocol-based VLAN function.
- To configure and enable the protocol-based VLAN function on a trunk or hybrid port, add all protocol-based VLANs to the allowed VLANs of the trunk or hybrid port.
- Configure the protocol-based VLAN packet type in global configuration mode, configure specified VLANs in the interface configuration mode, and enable the protocol-based VLAN function. Then the protocol-based VLAN function takes effect. When you delete the global configuration, the protocol-based VLAN configuration of all interfaces related to profile number is automatically deleted.
- **profile** *profile-number* specifies the number of a protocol-based VLAN rule (profile). The value range is from 1 to 16.
- **frame-type**: Specifies a frame type. The values include **EtherII**, **SNAP**, and **LLC**. If the frame type is **EtherII** or **SNAP**, configure the **ether-type** *type* parameter to specify the corresponding Ethernet type. When the frame type is **LLC**, configure the **dsap** *value* **ssap** *value* parameter to specify the corresponding SAP type.
- **ether-type** *type* specifies an Ethernet type of frames. The value range is from 1501 to 65535 (in decimal notation of the protocol ID). The value can also be a protocol name or the protocol ID in hexadecimal notation. The supported protocols include AARP (0x80f3), AMBER (0x6008), APPLETALK (0x809b), ARP (0x0806), CFM (0x8902), DECNET-IV (0x6003), DIAGNOSTIC (0x6005), EAPOL (0x888e), FC_ETH, HEX6000 (0x6000), HEX8042 (0x8042), IPv4 (0x0800), IPv6 (0x86DD), IPX (0x8137), LAT (0x6004),

LAVC-SCA (0x6007), MAC_IN_MAC (0x88E7), MOP-CONSOLE (0x6002), MOP-DUMP (0x6001), MUMPS (0x6009), NETBIOS (0x8040), SNX-IDP (0x0600), VINES-ECHO (0x0baf), and X-25 (0x0805).

- In the **dsap** *value* **ssap** *value* parameter, **dsap** specifies a destination service access point (DSAP), and **ssap** specifies a source service access point (SSAP). The range of *value* is from 0 to 255, corresponding to 0x00 to 0xFF of one byte in hexadecimal notation. For example, if an IP packet is encapsulated into a Logical Link Control (LLC) frame, the DSAP value and SSAP value are both 0x06.
- *vlan-id* specifies a VLAN ID. The value range is from 1 to 4094.

1.5.3 Prerequisites

Configure an interface as a Layer 2 interface by running the **switch-port** command.

1.5.4 Procedure

(1) Create a VLAN assigned by the protocol-based VLAN policy.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create a VLAN by configuring a policy.

vlan { *vlan-id* | **range** *vlan-range* }

Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

(2) Configure protocol VLAN rules one by one in global configuration mode.

protocol-vlan profile *profile-number* **frame-type** { **EtherII ether-type** *type* | **SNAP ether-type** *type* | **LLC dsap** *value* **ssap** *value* }

No global protocol-based VLAN rule is configured by default.

(3) Enter the interface configuration mode and configure protocol-based VLAN.

a Enter the Layer 2 Ethernet interface configuration mode.

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure the port as a trunk or hybrid port.

switch-port mode { **trunk | hybrid** }

The default Layer 2 interface mode is the access mode.

(5) Add protocol-based VLANs to the allowed VLAN list of an interface. Configure one of the following tasks.

- Configure an allowed VLAN list for the trunk port.

switch-port trunk allow-pass vlan { **all** | **add** *vlan-list* | **remove** *vlan-list* | **except** *vlan-list* | **only** *vlan-list* }

The default allowed VLANs of the trunk port are VLANs 1 to 4094.

- Configure an allowed VLAN list for the hybrid port.

  switch-port hybrid allow-pass vlan { [**add**] **tagged** *vlan-list* | [**add**] **untagged** *vlan-list* | **only tagged** *vlan-list* | **remove** *vlan-list* }

  The default allowed VLANs of the hybrid port are untagged VLAN 1 and tagged VLANs 2 to 4094.

(6) Configure protocol-based VLAN rules one by one on the interface so that packets that match the *num* rule are assigned to the VLAN specified by VLAN *vlan-id*.

protocol-vlan profile *profile-number* **vlan** *vlan-id*

The protocol-based VLAN function is disabled by default.

1.6 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

⚠ Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging function immediately after use.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** command may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
display protocol-vlan	Displays configurations of all protocol-based VLANs.
display protocol-vlan profile	Displays the protocol-based VLAN configuration.
display protocol-vlan ipv4	Displays the subnet-based VLAN configuration.
debugging bridge protvlan	Debugs the protocol-based VLAN function.

1.7 Configuration Examples

1.7.1 Configuring Subnet-based VLAN

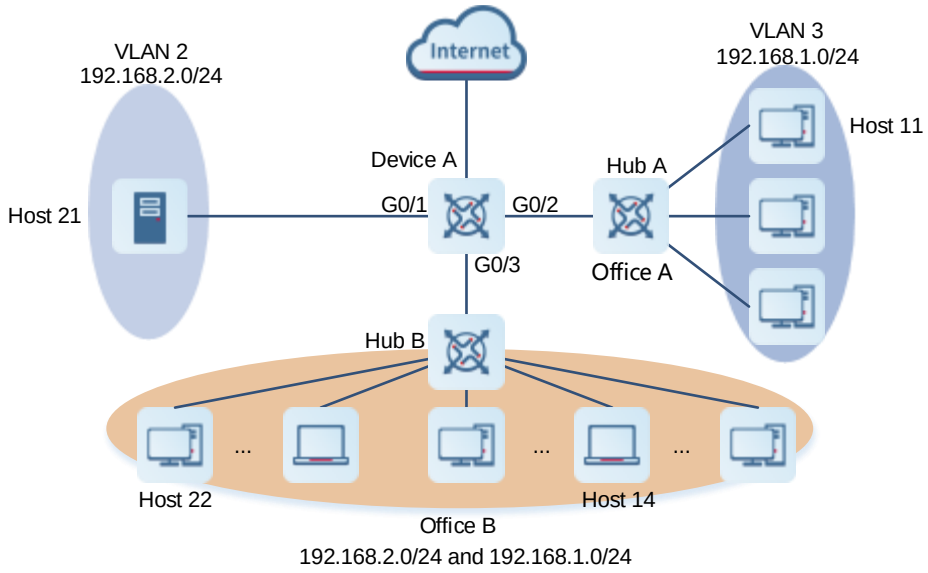
1. Requirements

Office area A and office area B are connected to Layer 3 device A through hubs. In office area A, hosts belong to a fixed network segment and are assigned to VLANs based on ports. In shared office area B, there are two network segments of hosts. Such hosts are not connected through fixed access ports and therefore they cannot

be assigned to VLANs based on ports. Device A must assign hosts in office area B to VLANs based on the IP address segments, to which user packets belong.

2. Topology

Figure 1-2 Configuring Subnet-based VLAN



3. Notes

- Configure VLAN 2 and VLAN 3 on device A.
- Configure GigabitEthernet 0/1 as an access port and assign the port to VLAN 2. Configure GigabitEthernet 0/2 as an access port and assign the port to VLAN 3.
- Configure a subnet-based VLAN rule in global configuration mode on device A and assign the IP address segment 192.168.2.0/24 to VLAN 2 and the IP address segment 192.168.1.0/24 to VLAN 3.
- Configure GigabitEthernet 0/3 as a hybrid port, add untagged VLAN 2 and VLAN 3 to the allowed VLAN list of the port, and enable the subnet-based VLAN function on GigabitEthernet 0/3.

4. Procedure

(1) Create VLANs to be assigned based on subnets. Configure a global subnet-based VLAN rule, in which packets with IP addresses in the network segment of 192.168.1.0/24 are assigned to VLAN 3 and packets with IP addresses in the network segment of 192.168.2.0/24 are assigned to VLAN 2.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan range 2-3
DeviceA(cfg-vlan)# exit
DeviceA(cfg)# protocol-vlan ipv4 192.168.1.0 mask 255.255.255.0 vlan 3
DeviceA(cfg)# protocol-vlan ipv4 192.168.2.0 mask 255.255.255.0 vlan 2
```

(2) Configure GigabitEthernet 0/1 and GigabitEthernet 0/2 as access ports and add them to different VLANs.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port
```

```
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port access vlan 2
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 3
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

(3) Configure GigabitEthernet 0/3 that will run the subnet-based VLAN function as a hybrid port, and allow packets from the subnet-based VLANs to pass.

```
DeviceA(cfg)# interface gigabitethernet 0/3
DeviceA(cfg-if-GigabitEthernet 0/3)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/3)# switch-port mode hybrid
DeviceA(cfg-if-GigabitEthernet 0/3)# switch-port hybrid allow-pass vlan untagged 2-3
```

(4) Enable the subnet-based VLAN function on GigabitEthernet 0/3 that needs to run the subnet-based VLAN function.

```
DeviceA(cfg-if-GigabitEthernet 0/3)# protocol-vlan ipv4
```

5. Verification

(1) Run the **display protocol-vlan ipv4** command to check whether the protocol-based VLAN configuration on the device is correct.

```
DeviceA# display protocol-vlan ipv4
ip                  mask                vlan
--------------- --------------- ----
192.168.1.0       255.255.255.0     3
192.168.2.0       255.255.255.0     2
interface               ipv4 status
------------------- -----------
Gi0/3                     enable
```

(2) Take hosts 21, 22, 11, and 14 as examples to verify the subnet-based VLAN assignment effect.

Set the default gateway address of host 21 in VLAN 2 to 192.168.2.1, and the IP address to 192.168.2.21/24.

```
Host21> privilege
Host21# config-system terminal
Host21(cfg)# ip route-config 0.0.0.0 0.0.0.0 192.168.2.1
Host21(cfg)# interface gigabitethernet 0/1
Host21(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host21(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.21/24
Host21(cfg-if-GigabitEthernet 0/1)# end
```

Set the default gateway address of host 22 in VLAN 2 to 192.168.2.1, and the IP address to 192.168.2.22/24.

```
Host22> privilege
Host22# config-system terminal
Host22(cfg)# ip route-config 0.0.0.0 0.0.0.0 192.168.2.1
Host22(cfg)# interface gigabitethernet 0/1
```

```
Host22(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host22(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.22/24
Host22(cfg-if-GigabitEthernet 0/1)# end
```

Set the default gateway address of host 11 in VLAN 3 to 192.168.1.1, and the IP address to 192.168.1.11/24.

```
Host11> privilege
Host11# config-system terminal
Host11(cfg)# ip route-config 0.0.0.0 0.0.0.0 192.168.1.1
Host11(cfg)# interface gigabitethernet 0/1
Host11(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host11(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.11/24
Host11(cfg-if-GigabitEthernet 0/1)# end
```

Set the default gateway address of host 14 in VLAN 3 to 192.168.1.1, and the IP address to 192.168.1.14/24.

```
Host14> privilege
Host14# config-system terminal
Host14(cfg)# ip route-config 0.0.0.0 0.0.0.0 192.168.1.1
Host14(cfg)# interface gigabitethernet 0/1
Host14(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host14(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.14/24
Host14(cfg-if-GigabitEthernet 0/1)# end
```

Verify that host 22 is assigned to VLAN 2 by the subnet-based VLAN function, and host 22 can ping host 21 in VLAN 2 but cannot ping host 11 in VLAN 3.

```
Host22# ping 192.168.2.21
Sending 5, 100-byte ICMP Echoes to 192.168.2.21, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/4 ms.
Host22# ping 192.168.1.11
Sending 5, 100-byte ICMP Echoes to 192.168.1.11, timeout is 2 seconds:
  < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5).
```

Verify that host 14 is assigned to VLAN 3 by the subnet-based VLAN function, and host 14 can ping host 11 in VLAN 3 but cannot ping host 21 in VLAN 2.

```
Host14# ping 192.168.1.11
Sending 5, 100-byte ICMP Echoes to 192.168.1.11, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/4 ms.
Host14# ping 192.168.2.21
Sending 5, 100-byte ICMP Echoes to 192.168.2.21, timeout is 2 seconds:
  < press Ctrl+C to break >
.....
```

```
Success rate is 0 percent (0/5).
```

Configure the IP addresses of the SVI gateways for VLAN 2 and VLAN 3 on device A to implement Layer 3 connection between the VLANs.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface vlan 2
DeviceA(cfg-if-VLAN 2)# ip address 192.168.2.1/24
DeviceA(cfg-if-VLAN 2)# exit
DeviceA(cfg)# interface vlan 3
DeviceA(cfg-if-VLAN 3)# ip address 192.168.1.1/24
DeviceA(cfg-if-VLAN 2)# end
```

Verify that host 14 in VLAN 3 can ping host 21 in VLAN 2. Likewise, verify that host 22 in VLAN 3 can ping host 21 in VLAN 2.

```
Host14# ping 192.168.2.21
Sending 5, 100-byte ICMP Echoes to 192.168.2.21, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 3/19/79 ms.
```

6. Configuration Files

Device A configuration file

```
protocol-vlan ipv4 192.168.1.0 mask 255.255.255.0 vlan 3
protocol-vlan ipv4 192.168.2.0 mask 255.255.255.0 vlan 2
!
vlan range 2-3
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode access
 switch-port access vlan 2
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode access
 switch-port access vlan 3
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode hybrid
 switch-port hybrid allow-pass vlan only tagged 1,4-4094
 switch-port hybrid allow-pass vlan add untagged 2-3
 protocol-vlan ipv4
```

7. Common Errors

- A port connected to the device is not a trunk or hybrid port.

- The allowed VLAN list of the port connected to the device does not contain the user communication VLANs.
- The subnet-based VLAN function is disabled on a port.
- Packets match a VLAN assignment policy with a higher priority (for example, MAC VLAN or voice VLAN) on the port, causing the failure of the subnet-based VLAN policy to take effect.

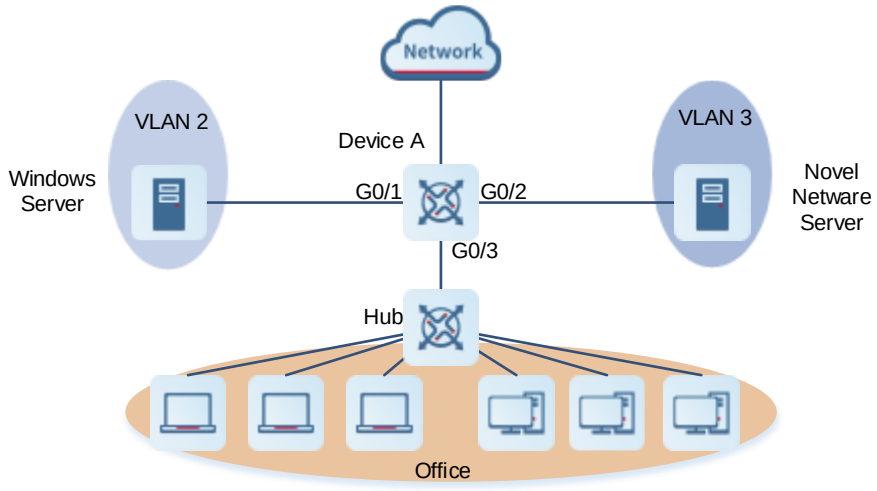
1.7.2 Configuring Protocol-based VLAN

1. Requirements

In an office area, some PCs run the Windows NT operating system (OS) and support the IP protocol, and some run the Novell Netware OS and support the Internetwork Packet Exchange (IPX) protocol. Two servers in the office network are used to isolate hosts running different OSs at Layer 2 so that the hosts running Windows NT can access only the Windows NT server and the hosts running Novell Netware can access only the Novell Netware server. The purpose is to reduce network traffic.

2. Topology

Figure 1-3 Configuring Protocol-based VLAN



3. Notes

- Configure VLAN 2 and VLAN 3 on device A.
- Configure GigabitEthernet 0/1 as an access port and assign the port to VLAN 2. Configure GigabitEthernet 0/2 as an access port and assign the port to VLAN 3.
- Configure protocol-based VLAN rules in global configuration mode on device A so that IP packets match rule 1 and IPX packets match rule 2.
- Configure GigabitEthernet 0/3 as a hybrid port and add VLAN 2 and VLAN 3 to the allowed untagged VLAN list of the port. Enable the protocol-based VLAN function on GigabitEthernet 0/3, and associate rule 1 with VLAN 2 and rule 2 with VLAN 3.

4. Procedure

(1) Create VLAN 2 and VLAN 3 to be assigned according to protocol rules.

```
DeviceA> privilege
```

```
DeviceA# config-system terminal
DeviceA(cfg)# vlan range 2-3
DeviceA(cfg-vlan)# exit
```

(2) Configure protocol-based VLAN rules in global configuration mode.

Apply rule 1 to IP packets whose frame type is EthernetII and Ethernet type is 0x0800.

```
DeviceA(cfg)# protocol-vlan profile 1 frame-type ETHERII ether-type 0x0800
```

Apply rule 2 to IPX packets whose frame type is EthernetII and Ethernet type is 0x8137.

```
DeviceA(cfg)# protocol-vlan profile 2 frame-type ETHERII ether-type 0x8137
```

Apply rule 1 to IP packets whose Ethernet type is 0x0800.

```
DeviceA(cfg)# protocol-vlan profile 1 ether-type 0x0800
```

(3) Configure ports that connect to servers as access ports and add the ports to different VLANs.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port access vlan 2
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 3
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

(4) Configure the port that needs to run the protocol-based VLAN function as a hybrid port, and add VLAN 2 and VLAN 3 to the allowed VLAN list of the port.

```
DeviceA(cfg)# interface gigabitethernet 0/3
DeviceA(cfg-if-GigabitEthernet 0/3)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/3)# switch-port mode hybrid
DeviceA(cfg-if-GigabitEthernet 0/3)# switch-port hybrid allow-pass vlan untagged 2-3
```

(5) Enable the protocol-based VLAN function on GigabitEthernet 0/3, and assign packets that match rule 1 to VLAN 2 and packets that match rule 2 to VLAN 3.

```
DeviceA(cfg-if-GigabitEthernet 0/3)# protocol-vlan profile 1 vlan 2
DeviceA(cfg-if-GigabitEthernet 0/3)# protocol-vlan profile 2 vlan 3
```

5. Verification

Run the **display protocol-vlan profile** [*profile-number*] command to check whether the protocol-based VLAN configuration on the device is correct.

```
DeviceA(cfg)# display protocol-vlan profile
profile frame-type        ether-type/DSAP+SSAP    interface        vlan
------- ---------------- ---------------------- --------------- ----
1       ETHERII           0x0800                  Gi0/3            2
2       ETHERII           0x8137                  Gi0/3            3
```

6. Configuration Files

Device A configuration file

```
protocol-vlan profile 1 frame-type ETHERII ether-type IPV4
```

```
protocol-vlan profile 2 frame-type ETHERII ether-type IPX
!
vlan range 2-3
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode access
 switch-port access vlan 2
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode access
 switch-port access vlan 3
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode hybrid
 switch-port hybrid allow-pass vlan only tagged 1,4-4094
 switch-port hybrid allow-pass vlan add untagged 2-3
 protocol-vlan profile 1 vlan 2
 protocol-vlan profile 2 vlan 3
```

7. Common Errors

- A port connected to the device is not a trunk or hybrid port.
- The allowed VLAN list of the port connected to the device does not contain the user communication VLANs.
- The protocol-based VLAN function is disabled on a port.
- Packets match a VLAN assignment policy with a higher priority (for example, MAC VLAN, voice VLAN, or subnet-based VLAN) on the port, causing the failure of the protocol-based VLAN policy to take effect.

Contents

1 Configuring Private VLAN 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Concepts 1
1.1.3 Packet Forwarding Rules 3
1.2 Restrictions and Guidelines 4
1.3 Configuration Task Summary 4
1.4 Configuring PVLAN and Layer 2 Association 5
1.4.1 Overview 5
1.4.2 Restrictions and Guidelines 5
1.4.3 Procedure 6
1.5 Configuring a Layer 3 Association for a PVLAN 6
1.5.1 Overview 6
1.5.2 Restrictions and Guidelines 7
1.5.3 Prerequisites 7
1.5.4 Procedure 7
1.6 Configuring an Isolated Port or a Community Port 7
1.6.1 Overview 7
1.6.2 Restrictions and Guidelines 7
1.6.3 Prerequisites 8
1.6.4 Procedure 8
1.7 Configuring a Promiscuous Port 8

1.7.1 Overview 8

1.7.2 Restrictions and Guidelines 8

1.7.3 Prerequisites 9

1.7.4 Procedure 9

1.8 Monitoring 9

1.9 Configuration Examples 10

1.9.1 Configuring PVLAN Across Layer 2 Devices 10

1.9.2 Configuring PVLAN on a Single Layer 3 Device 18

1 Configuring Private VLAN

Note

Private VLAN is not supported by the S6980-64QC, S6920-4C, or S6580-24DC8QC.

1.1 Introduction

1.1.1 Overview

Carriers hope to isolate users to authenticate and charge each user, avoid virus attacks and broadcast storms, and strengthen user security. VLANs are isolated at Layer 2. If each user is assigned with a VLAN, the authentication, billing, and security requirements can be met. However, a device supports 4094 VLANs at most. Assigning a VLAN to each user limits the number of users supported by a device. What's more, it is infeasible for a Layer 3 device to assign one IP address to each VLAN in consideration of scarce IP addresses. To solve the problem, the private VLAN (PVLAN) technology is developed.

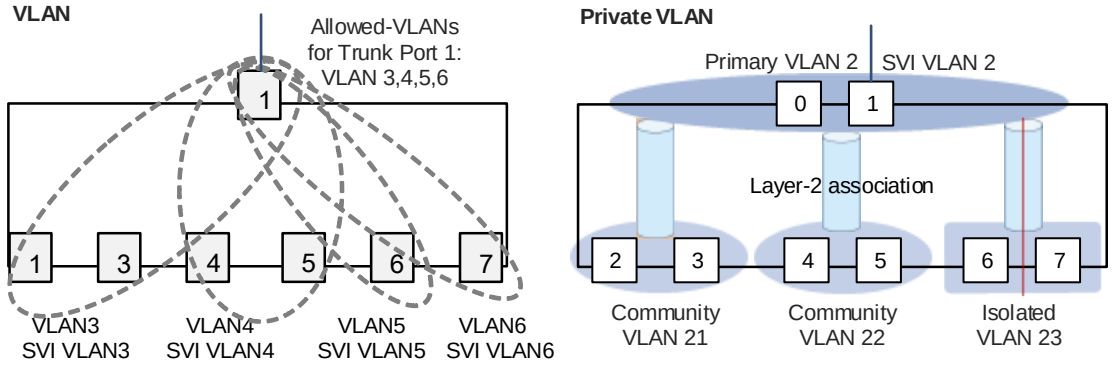
1.1.2 Basic Concepts

The PVLAN technology divides a Layer 2 broadcast domain of a VLAN into multiple sub-domains to form a dual-layer VLAN structure. The outer layer is referred to as primary VLAN, and the inner layer is referred to as secondary VLAN. Secondary VLANs can be further divided to community VLANs and isolated VLANs.

1. Features

- You can configure multiple PVLANs on a device. A PVLAN is a PVLAN domain.
- Each PVLAN can contain only one primary VLAN and multiple secondary VLANs. The primary VLAN corresponds to multiple secondary VLANs. The primary VLAN and one secondary VLAN form a PVLAN pair, which represents a PVLAN sub-domain. For example, VLAN 2/21 represents a PVLAN sub-domain. A PVLAN domain can have multiple PVLAN sub-domains. The secondary VLANs of different sub-domains are different, but they share the same primary VLAN.
- Both the primary VLAN and secondary VLANs can contain ports. As shown in Figure 1-1, ports 0 and 1 within the primary VLAN belong to primary VLAN 2; and ports 2 and 3 belong to community VLAN 21. After a Layer 2 association is configured between primary VLAN 2 and community VLAN 21, ports 0, 1, 2, and 3 belong to the same PVLAN sub-domain VLAN 2/21. They can make Layer 2 communication with each other.
- In a PVLAN, Layer 3 switch virtual interfaces (SVIs) can be created only in the primary VLAN, and Layer 3 SVIs cannot be created in secondary VLANs. You can make Layer 3 communication with the port in a primary VLAN through a port in the secondary VLAN only after a Layer 3 association is configured between the secondary VLAN and the primary VLAN. Otherwise, the hosts can make only Layer 2 communication.

Figure 1-1 Logic Diagram of PVLAN



2. Primary VLAN and Promiscuous Port

Ports belonging to the primary VLAN are referred to as promiscuous ports. A PVLAN can have one or more promiscuous ports.

- In a PVLAN domain, a promiscuous port can communicate with any port, including another promiscuous port, a community VLAN port, and an isolated VLAN port. A promiscuous port is used as an uplink port for connecting to an external network or a port for connecting to a shared server so that hosts in all secondary VLANs can access the network and shared server.
- Outside a PVLAN domain, a promiscuous port belongs to the primary VLAN, allows common packets with the primary VLAN ID to pass through, and discards the other common VLAN packets.

3. Community VLAN and Community Port

A PVLAN domain can have multiple community VLANs. The ports in a community VLAN are referred to as community ports.

- Community ports in a community VLAN can communicate with each other but community ports of different community VLANs cannot communicate with each other.
- Community ports can communicate with promiscuous ports.
- Community ports cannot communicate with isolated ports.

4. Isolated VLAN and Isolated Port

A PVLAN domain can have only one isolated VLAN. Ports in an isolated VLAN are referred to as isolated ports.

- Isolated ports cannot communicate with each other. You can assign required hosts to an isolated VLAN for isolation.
- Isolated ports can communicate with only promiscuous ports. Packets can be forwarded from isolated ports to the promiscuous ports and vice versa.
- Packets can be forwarded from an isolated port to a trunk port. When receiving a packet (packet from the promiscuous port) with the VLAN ID of the primary VLAN ID, the trunk port can forward the packet to the isolated port. This enables the isolated port to communicate with an external network through a promiscuous port. Hosts in an isolated VLAN cannot communicate with each other. Therefore, when receiving a packet with the VLAN ID of the isolated VLAN, the trunk port cannot forward the packet to the isolated port.

1.1.3 Packet Forwarding Rules

Table 1-1 lists packet forwarding rules between ports in a PVLAN and VLAN tag changes.

Table 1-1 Packet Forwarding Rules Between Different Types of Ports and VLAN Tag Changes

Input Port	Output Port	Packet Forwarding Rules and VLAN Tag Changes
Promiscuous Port	Promiscuous Port	Reachable, VLAN tag unchanged
	Community Port	Reachable, VLAN tag unchanged
	Isolated Port	Reachable, VLAN tag unchanged
	Trunk Port	Reachable, a primary VLAN tag added
Community Port	Promiscuous Port	Reachable, VLAN tag unchanged
	Community Port	● Reachable in the same community VLAN, VLAN tag unchanged ● Unreachable in different community VLANs
	Isolated Port	Unreachable
	Trunk Port	Reachable, a community VLAN tag added
	Promiscuous Port	Reachable, VLAN tag unchanged
	Community Port	Unreachable
	Isolated Port	Unreachable
	Trunk Port	Reachable. Packets are from isolated VLANs, and isolated-VLAN tags are added to the packets and then forwarded through trunk ports.
Isolated Port	Promiscuous Port	Reachable, VLAN tag unchanged
	Community Port	Unreachable
	Isolated Port	Unreachable
	Trunk Port	Reachable. An isolated VLAN tag is added to an isolated VLAN packet and the packet is forwarded through isolated ports.
	Promiscuous Trunk Port	● Promiscuous trunk port VLAN tag is added to a PVLAN packet. ● The port follows common VLAN packet forwarding rules to forward a public VLAN packet and the VLAN tag of the public VLAN packet is unchanged.
	Isolated Trunk Port	Unreachable
Trunk Port	Promiscuous Port	Reachable, VLAN tag removed

Input Port	Output Port	Packet Forwarding Rules and VLAN Tag Changes
	Community Port	● Reachable in the primary VLAN, VLAN tag removed ● Reachable in the same community VLAN, VLAN tag removed ● Unreachable in different community VLANs ● Unreachable in an isolated VLAN
	Isolated Port	● Reachable in the primary VLAN, VLAN tag removed ● Unreachable in a community VLAN ● Unreachable in an isolated VLAN
	Trunk Port	Reachable, VLAN tag unchanged
Switch CPU	Promiscuous Port	Reachable, no VLAN tag
	Community Port	Reachable, no VLAN tag
	Isolated Port	Reachable, no VLAN tag
	Trunk Port	Reachable, a primary VLAN tag added

1.2 Restrictions and Guidelines

- You can only configure a Layer 2 Ethernet interface or a Layer 2 aggregate interface as an isolated, community or promiscuous port within a PLVAN and add the port to the PVLAN.
- You can only configure a Layer 2 Ethernet interface or a Layer 2 aggregate interface as a trunk port. You can only configure the trunk port as an isolated or promiscuous trunk port within a PLVAN and add the port to the PVLAN.
- You can run the **display interface switch-port** command to display the interface type. If the **Switch-port** field corresponding to the interface is **enabled**, this interface is a Layer 2 interface. If the **Switch-port** field corresponding to the interface is **disabled**, this interface is a Layer 3 interface.
- You can run the **switch-port** command to configure the interface as a Layer 2 interface. Run the **unset switch-port mode** command to restore the Layer 2 interface mode to the default mode and run the **unset switch-port** command to configure a Layer 2 interface as a Layer 3 interface. When you configure an interface as a Layer 2 or Layer 3 interface, all attributes of the interface are restored to the default values in this interface mode.

1.3 Configuration Task Summary

The configuration includes the following tasks:

(1) Configure PVLAN.

a Configuring PVLAN and Layer 2 Association

b Configuring a Layer 3 Association for a PVLAN

(2) Configure port types and add the ports to PVLAN pairs.

a Configuring an Isolated Port or a Community Port

b Configuring a Promiscuous Port

1.4 Configuring PVLAN and Layer 2 Association

1.4.1 Overview

The primary VLAN and secondary VLANs of a PVLAN cannot exist separately. You must add the secondary VLANs to the primary VLAN to form a Layer 2 association and multiple PVLAN pairs, which is also called PVLAN sub-domains.

You can add isolated ports and community ports to the primary VLAN and secondary VLANs only after a Layer 2 association is configured between the secondary VLANs and the primary VLAN. If you cancel a Layer 2 association, the PVLAN pairs, to which the isolated ports and promiscuous ports are added, are canceled as well.

After a Layer 2 association is configured between the secondary VLANs and the primary VLAN, you can configure a Layer 3 association so that the multiple secondary VLANs use the same IP gateway to save IP address resources.

Use the **isolate-user-vlan l2-mapping** command to configure a Layer 2 association between the primary VLAN and the secondary VLANs.

1.4.2 Restrictions and Guidelines

- The PVLAN supports only host and promiscuous ports. Before you configure a common VLAN as a PVLAN, ensure that the ports in the PVLAN are valid. Otherwise, the configuration may fail. If a trunk, hybrid or uplink port exists in a common VLAN, the common VLAN can be configured as a PVLAN, but you are advised to remove the unused port from the PVLAN. If an access or dot1q-tunnel port exists in a common VLAN, the common VLAN cannot be configured as a PVLAN.
- **primary**: Configure a common VLAN as the primary VLAN of a PVLAN. A primary VLAN specifies a PVLAN. The device support creating multiple PVLANs and configuring multiple primary VLANs.
- **community**: Configure a common VLAN as a community VLAN. A Private VLAN has one or more community VLANs. Therefore, each primary VLAN can be associated with one or more community VLANs.
- **isolated**: Configure a common VLAN as an isolated VLAN. A Private VLAN has only one isolated VLAN. Therefore, each primary VLAN can be associated with only one isolated VLAN.
- *secondary-vlan-list* indicates a list of secondary VLANs. **add** *secondary-vlan-list* is used to associate with a secondary VLAN. **remove** *secondary-vlan-list* is used to remove the association specified by *secondary-vlan-list*.
- You cannot delete a PVLAN by using the **unset vlan** command. You must run the **unset isolate-user-vlan** { **community** | **isolated** | **primary** } command to restore a PVLAN to a common static VLAN, and be sure to keep the original PVLAN attributes unchanged.
- Before you modify PVLAN attributes of a port, you must restore the PVLAN to a common static VLAN and then configure new PVLAN attributes for the port. When you restore a PVLAN to a common static VLAN, all the associations in the PVLAN are canceled. You need to reconfigure associations for the PVLAN and add ports to the PVLAN.

1.4.3 Procedure

(1) Configure a secondary VLAN for a PVLAN.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create a secondary VLAN and enter the VLAN configuration mode.

vlan { *vlan-id* | **range** *vlan-range* }

Only VLAN 1 exists by default.

d Configure attributes of the secondary VLAN.

isolate-user-vlan { **community** | **isolated** }

A VLAN is a common VLAN and has no PVLAN attribute by default.

e Return to the global configuration mode.

exit

(2) Configure a primary VLAN for the PVLAN.

a Create a primary VLAN and enter the VLAN configuration mode.

vlan *vlan-id*

Only VLAN 1 exists by default.

b Configure attributes of the primary VLAN.

isolate-user-vlan primary

A VLAN is a common VLAN and has no PVLAN attribute by default.

(3) Add the secondary VLAN to the primary VLAN to form a Layer 2 association.

isolate-user-vlan l2-mapping { *secondary-vlan-list* | **add** *secondary-vlan-list* | **remove** *secondary-vlan-list* }

No Layer 2 association is configured between the primary VLAN and the secondary VLANs by default.

1.5 Configuring a Layer 3 Association for a PVLAN

1.5.1 Overview

To enable hosts in a secondary VLAN domain to access an external network, configure a Layer 3 SVI for the primary VLAN, configure an IP address for the SVI, and then configure a Layer 3 association between the primary VLAN and the secondary VLAN on the SVI of the primary VLAN. Then, hosts in the secondary VLAN can make Layer 3 communication through the SVI of the primary VLAN. If no Layer 3 association is configured between the primary VLAN and the secondary VLAN, the hosts in the secondary VLAN cannot make Layer 3 communication.

You can use the **isolate-user-vlan l3-mapping** command to configure/remove a Layer 3 association between the primary VLAN and the secondary VLAN.

1.5.2 Restrictions and Guidelines

- You can create an SVI only for a primary VLAN. Run the **interface vlan** *interface-number* command to create an SVI. The value of *interface-number* must be the VID of a primary VLAN.
- To configure an IPv4 address, see *Configuring IPv4 Basics*.
- To configure an IPv6 address, see *Configuring IPv6 Basics*.

1.5.3 Prerequisites

Create a PVLAN and a Layer 2 association between a secondary VLAN and the primary VLAN.

1.5.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an SVI for the primary VLAN and enter the SVI configuration mode.

interface vlan *interface-number*

No SVI exists in a VLAN by default.

(4) Configure an IP address for the SVI.

ip address { *ipv4-address mask* | *ipv4-address/mask-length* }

An SVI has no IP address by default.

(5) Configure a Layer 3 association for the PVLAN.

isolate-user-vlan l3-mapping { *secondary-vlan-list* | **add** *secondary-vlan-list* | **remove** *secondary-vlan-list* }

No Layer 3 association is configured between the primary VLAN and the secondary VLANs by default.

1.6 Configuring an Isolated Port or a Community Port

1.6.1 Overview

After the primary VLAN and secondary VLAN as well as a Layer 2 association are configured, you need to configure ports connected to hosts. User ports are configured as host ports. If an enterprise has only one host, you can configure the user port as an isolated port and add the port to an isolated VLAN. If an enterprise has multiple hosts, you can configure the user ports as community ports and add the ports to a community VLAN. The host ports added to an isolated VLAN are isolated ports, and the host ports added to community VLANs are community ports.

1.6.2 Restrictions and Guidelines

- The Layer 2 interfaces connected to hosts must be configured as host ports before they are associated with PVLAN pairs to add the host ports to a secondary VLAN of a PVLAN.
- *primary-vlan-id* specifies the primary VLAN ID in a PVLAN pair. *secondary-vlan-id* specifies a secondary VLAN ID in a PVLAN pair. The VLANs specified by *primary-vlan-id* and *secondary-vlan-id* must be a PVLAN

pair that has a Layer 2 association.

- One host port can be associated with only one PVLAN pair. To modify the association of a host port with a PVLAN pair, use the **unset switch-port isolate-user-vlan l2-mapping** command to cancel the association between the host port and the PVLAN pair.

1.6.3 Prerequisites

- Configure an interface as a Layer 2 interface.

Create a PVLAN and a Layer 2 association between a secondary VLAN and the primary VLAN.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 link aggregation configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure a Layer 2 interface as a PVLAN host port.

switch-port mode isolate-user-vlan host

A Layer 2 interface is an access port by default.

(5) Associate the host port with a PVLAN pair that has a Layer 2 association and add the host port to a secondary VLAN of the PLVAN.

switch-port isolate-user-vlan l2-mapping *primary-vlan-id secondary-vlan-id*

A port belongs to VLAN 1 and is not associated with a Layer 2 PVLAN pair by default.

1.7 Configuring a Promiscuous Port

1.7.1 Overview

To enable a port connected to an external network or a server to be accessible to community ports and isolated ports, configure the port as a promiscuous port so that uplink and downlink packets can be normally forwarded.

1.7.2 Restrictions and Guidelines

- You must configure a Layer 2 interface as a promiscuous port and then associate the promiscuous port with a PVLAN pair to add the port to the primary VLAN of the PLVAN.
- A promiscuous port can be associated with multiple PVLAN pairs that share the same primary VLAN. However, a host port can be associated with one PVLAN pair only.

1.7.3 Prerequisites

- Configure an interface as a Layer 2 interface.

Create a PVLAN and a Layer 2 association between a secondary VLAN and the primary VLAN.

1.7.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 link aggregation configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure a Layer 2 interface as a PVLAN promiscuous port.

switch-port mode isolate-user-vlan promiscuous

A Layer 2 interface is an access port by default.

(5) Associate the promiscuous port with one or more PVLAN pairs that have Layer 2 associations to add the port to the primary VLAN of the PLVAN.

switch-port isolate-user-vlan l3-mapping *primary-vlan-id* { *secondary-vlan-list* | **add** *secondary-vlan-list* | **remove** *secondary-vlan-list* }

A port belongs to VLAN 1 and is not associated with a Layer 2 PVLAN pair by default.

1.8 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

⚠ Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging function immediately after use.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-2 Monitoring

Command	Purpose
display vlan isolate-user-vlan [**community** | **primary** | **isolated**]	Displays PVLAN configuration.
debugging bridge pvlan	Debugs the PVLAN function.

1.9 Configuration Examples

1.9.1 Configuring PVLAN Across Layer 2 Devices

1. Requirements

As shown in Figure 1-2, in the hosting service operation network, enterprise hosts are connected to the network through device A or device B. Main requirements are as follows:

- Hosts of an enterprise can communicate with each other but the host communication between enterprises is isolated. All enterprise hosts share the same gateway address and can communicate with the external network through this gateway address.

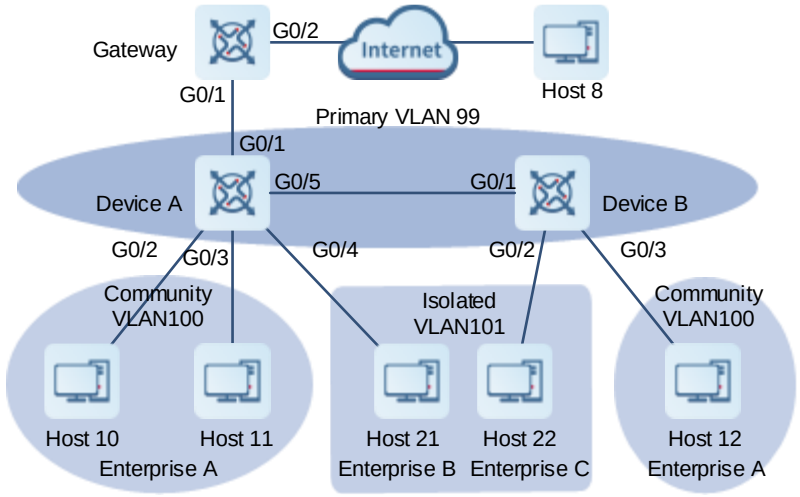
The deployment mode is as follows:

- Configure the PVLAN function on device A and device B and the interconnection ports are all trunk ports. Add all enterprise hosts to the same PVLAN (for example, primary VLAN 99).
- All the enterprise hosts communicate with an external network through a shared SVI of the primary VLAN 99. The external network perceives the primary VLAN 99 only and does not perceive secondary VLANs. Configure GigabitEthernet 0/1 of device A as a promiscuous port and connect this interface to the gateway. The promiscuous port only allows primary VLAN packets from the PVLAN to pass through and outputs untagged packets.
- GigabitEthernet 0/1 of the gateway can function as a trunk or hybrid port. Because packets forwarded by a promiscuous port are untagged, set the native VLAN of GigabitEthernet 0/1 of the gateway to the primary VLAN of the PVLAN.
- If an enterprise has multiple hosts, allocate the hosts to different community VLANs and configure the ports connected to the hosts as host ports of the community VLANs so that the hosts in the enterprise can communicate with each other but host communication between enterprises is isolated.

  If an enterprise has only one host, allocate the host to an isolated VLAN and configure the port connected to the host as the host port of the isolated VLAN, to implement isolation of communication between hosts of different enterprises.

2. Topology

Figure 1-2 Configuring PVLAN Applied Across Layer 2 Devices



3. Notes

- Add all enterprise hosts to a PVLAN (primary VLAN 99). All enterprise hosts share the same SVI through this PVLAN to communicate with the external network. Enterprise A has multiple hosts. Assign the hosts to community VLAN 100 so that hosts in the enterprise can communicate with each other and the host communication between enterprises is isolated. Enterprise B and enterprise C has only one host each. Assign the hosts to isolated VLAN 101. The hosts in enterprise B and enterprise C cannot communicate with each other. Configure a PVLAN on Device A and Device B. The PVLAN contains one primary VLAN 99, one community VLAN 100, and one isolated VLAN 101.
- Configure a Layer 3SVI for the primary VLAN on Device A, and configure a Layer 3 association between the primary VLAN and the secondary VLANs.
- Configure GigabitEthernet 0/2-0/4 of Device A as host ports. Configure GigabitEthernet 0/2 and 0/3 of Device B as host ports.
- Configure GigabitEthernet 0/1 of Device A connected to a gateway as a promiscuous port. Only packets from the primary VLAN can pass through this port. No tag is carried when the promiscuous port outputs packets from the primary VLAN.
- Because the PVLAN runs across devices, the ports connecting different devices must be configured as trunk ports. Configure GigabitEthernet 0/5 of Device A and GigabitEthernet 0/1 of Device B as trunk ports. GigabitEthernet 0/1 of the gateway can be configured as a trunk port or hybrid port. Because packets forwarded by a promiscuous port do not carry tags, the native VLAN of GigabitEthernet 0/1 must be set to the primary VLAN of the PVLAN.

4. Procedure

(1) Configure a PVLAN and Layer 2 associations.

Configure a Layer 2 association between the secondary VLANs (VLAN 100 and VLAN 101) and the primary VLAN 99 on Device A.

```
DeviceA> privilege
```

```
DeviceA# config-system terminal
DeviceA(cfg)# vlan 100
DeviceA(cfg-vlan)# isolate-user-vlan community
DeviceA(cfg-vlan)# exit
DeviceA(cfg)# vlan 101
DeviceA(cfg-vlan)# isolate-user-vlan isolated
DeviceA(cfg-vlan)# exit
DeviceA(cfg)# vlan 99
DeviceA(cfg-vlan)# isolate-user-vlan primary
DeviceA(cfg-vlan)# isolate-user-vlan l2-mapping 100-101
DeviceA(cfg-vlan)# exit
```

Configure a Layer 2 association between the secondary VLANs (VLAN 100 and VLAN 101) and the primary VLAN 99 on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan 100
DeviceB(cfg-vlan)# isolate-user-vlan community
DeviceB(cfg-vlan)# exit
DeviceB(cfg)# vlan 101
DeviceB(cfg-vlan)# isolate-user-vlan isolated
DeviceB(cfg-vlan)# exit
DeviceB(cfg)# vlan 99
DeviceB(cfg-vlan)# isolate-user-vlan primary
DeviceB(cfg-vlan)# isolate-user-vlan l2-mapping 100-101
DeviceB(cfg-vlan)# exit
```

(2) Configure a Layer 3 association for the PVLAN.

Configure an IP address and a mask for the Layer 3 SVI of the primary VLAN on Device A, and configure a Layer 3 association between the primary VLAN and the secondary VLANs.

```
DeviceA(cfg)# interface vlan 99
DeviceA(cfg-vlan)# ip address 192.168.99.1 255.255.255.0
DeviceA(cfg-vlan)# isolate-user-vlan mapping 100,101
DeviceA(cfg-vlan)# exit
```

(3) Add the host ports to the secondary VLANs.

Configure the ports of Device A connected to hosts as host ports and associate the ports with PVLAN pairs to add the ports to the secondary VLANs.

```
DeviceA(cfg)# interface list gigabitethernet 0/2-3
DeviceA(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode isolate-user-vlan host
DeviceA(cfg-if-list)# switch-port isolate-user-vlan l2-mapping 99 100
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# interface gigabitethernet 0/4
DeviceA(cfg-if-GigabitEthernet 0/4)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/4)# switch-port mode isolate-user-vlan host
DeviceA(cfg-if-GigabitEthernet 0/4)# switch-port isolate-user-vlan l2-mapping 99 101
```

```
DeviceA(cfg-if-GigabitEthernet 0/4)# exit
```

Configure the ports of Device B connected to hosts as host ports and associate the ports with PVLAN pairs to add the ports to the secondary VLANs.

```
DeviceB(cfg)# interface gigabitethernet 0/2
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port mode isolate-user-vlan host
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port isolate-user-vlan l2-mapping 99 101
DeviceB(cfg-if-GigabitEthernet 0/2)# exit
DeviceB(cfg)# interface gigabitethernet 0/3
DeviceB(cfg-if-GigabitEthernet 0/3)# switch-port
DeviceB(cfg-if-GigabitEthernet 0/3)# switch-port mode isolate-user-vlan host
DeviceB(cfg-if-GigabitEthernet 0/3)# switch-port isolate-user-vlan l2-mapping 99 100
DeviceB(cfg-if-GigabitEthernet 0/3)# exit
```

(4) Configure GigabitEthernet 0/1 of Device A as a promiscuous port and associate the port with the primary VLAN 99 and the secondary VLANs 100 and 101 to add the port to the primary VLAN.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port mode isolate-user-vlan promiscuous
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port isolate-user-vlan l3-mapping 99 add 100-101
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

(5) Configure the ports connecting devices in the PVLAN as common trunk ports.

Configure GigabitEthernet 0/5 of Device A as a trunk port.

```
DeviceA(cfg)# interface gigabitethernet 0/5
DeviceA(cfg-if-GigabitEthernet 0/5)# switch-port
 DeviceA(cfg-if-GigabitEthernet 0/5)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet 0/5)# switch-port trunk pvidvlan 1
DeviceA(cfg-if-GigabitEthernet 0/5)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-GigabitEthernet 0/5)# end
DeviceA# write
```

Configure GigabitEthernet 0/1 of Device B as a trunk port.

```
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid  vlan 1
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-GigabitEthernet 0/1)# end
DeviceB# write
```

(6) Configure the downlink port GigabitEthernet 0/1 of the gateway as a trunk port and set the native VLAN of this port to the primary VLAN.

```
GW> privilege
GW# config-system terminal
GW(cfg)# interface gigabitethernet 0/1
GW(cfg-if-GigabitEthernet 0/1)# switch-port
```

```
GW(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
GW(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid    vlan 99
GW(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
GW(cfg-if-GigabitEthernet 0/1)# end
GW# write
```

5. Verification

(1) Check whether the VLAN and port configurations are correct.

Check whether the VLAN and port configurations on Device A are correct. GigabitEthernet 0/1 is a promiscuous port and belongs to primary VLAN 99. GigabitEthernet 0/2 and GigabitEthernet 0/3 are community ports and belong to community VLAN 100. GigabitEthernet 0/4 is an isolated port and belongs to isolated VLAN 101. GigabitEthernet 0/5 is a trunk port and belongs to all VLANs. A Layer 3 SVI address is configured on Device A and the routing function is enabled.

```
DeviceA# display vlan isolate-user-vlan
VLAN    Type            Status      Routed        Ports                          Associated VLANs
----    ----          -------     ------        ------------------          ------------------
99       primary       active      Enabled       Gi0/1, Gi0/5                  100-101
100      community     active      Enabled       Gi0/2, Gi0/3, Gi0/5           99
101      isolated      active      Enabled       Gi0/4, Gi0/5                  99
```

Check whether the VLAN and port configurations on Device B are correct.

```
DeviceB# display vlan isolate-user-vlan
VLAN    Type            Status      Routed        Ports                          Associated VLANs
----    ----          -------     ------        ------------------          ------------------
99       primary       active      Disabled      Gi0/1                          100-101
100      community     active      Disabled      Gi0/1, Gi0/3                   99
101      isolated      active      Disabled      Gi0/1, Gi0/2                   99
```

Check whether the gateway configuration is correct.

```
GW# display interface gigabitethernet 0/1 switch-port
Interface                                 Switch-port Mode          Access Native Protected VLAN lists
--------------------------------- ---------- --------- ------ ------ --------- ----------
GigabitEthernet 0/1                       enabled       TRUNK        1          99         Disabled    ALL
```

(2) Configure IP addresses and masks for hosts. To facilitate verification, add the hosts to the same network segment.

Configure an IP address and a mask for host 10 in the community VLAN.

```
Host10> privilege
Host10# config-system terminal
Host10(cfg)# interface gigabitethernet 0/1
Host10(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host10(cfg-if-GigabitEthernet 0/1)# ip address 192.168.99.10 255.255.255.0
```

Configure an IP address and a mask for host 11 in the community VLAN.

```
Host11> privilege
Host11# config-system terminal
Host11(cfg)# interface gigabitethernet 0/1
```

```
Host11(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host11(cfg-if-GigabitEthernet 0/1)# ip address 192.168.99.11 255.255.255.0
```

Configure an IP address and a mask for host 12 in the community VLAN.

```
Host12> privilege
Host12# config-system terminal
Host12(cfg)# interface gigabitethernet 0/1
Host12(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host12(cfg-if-GigabitEthernet 0/1)# ip address 192.168.99.12 255.255.255.0
```

Configure an IP address and a mask for host 21 in the isolated VLAN.

```
Host21> privilege
Host21# config-system terminal
Host21(cfg)# interface gigabitethernet 0/1
Host21(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host21(cfg-if-GigabitEthernet 0/1)# ip address 192.168.99.21 255.255.255.0
```

Configure an IP address and a mask for host 22 in the isolated VLAN.

```
Host22> privilege
Host22# config-system terminal
Host22(cfg)# interface gigabitethernet 0/1
Host22(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host22(cfg-if-GigabitEthernet 0/1)# ip address 192.168.99.22 255.255.255.0
```

To simulate external network data, connect host 8 to the gateway interface GigabitEthernet 0/2 and add GigabitEthernet 0/2 to VLAN 99.

```
GW> privilege
GW# config-system terminal
GW(cfg)# interface gigabitethernet 0/2
GW(cfg-if-GigabitEthernet 0/2)# switch-port
GW(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 99
```

Configure an address and a mask for host 8.

```
Host8> privilege
Host8# config-system terminal
Host8(cfg)# interface gigabitethernet 0/1
Host8(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host8(cfg-if-GigabitEthernet 0/1)# ip address 192.168.99.8 255.255.255.0
```

(3) Check whether packets from the hosts in the PVLAN are forwarded in accordance with the forwarding rules of the PVLAN ports.

Hosts in the same community VLAN can communicate with each other even if the PVLAN is configured across devices.

```
Host10# ping 192.168.99.11
Sending 5, 100-byte ICMP Echoes to 192.168.99.11, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms.
Host10# ping 192.168.99.12
```

```
Sending 5, 100-byte ICMP Echoes to 192.168.99.12, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms.
```

Hosts in an isolated VLAN cannot communicate with each other.

```
Host21# ping 192.168.99.22
Sending 5, 100-byte ICMP Echoes to 192.168.99.22, timeout is 2 seconds:
   < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5).
```

Hosts in a community VLAN cannot communicate with hosts in an isolated VLAN.

```
Host10# ping 192.168.99.22
Sending 5, 100-byte ICMP Echoes to 192.168.99.22, timeout is 2 seconds:
   < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5).
```

Hosts in a community VLAN and hosts in an isolated VLAN can communicate with an external network through the promiscuous port and gateway.

```
Host10# ping 192.168.99.8
Sending 5, 100-byte ICMP Echoes to 192.168.99.8, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/4 ms.
Host22# ping 192.168.99.8
Sending 5, 100-byte ICMP Echoes to 192.168.99.8, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/4 ms.
```

6. Configuration Files

- Device A configuration file

```
vlan 99
 isolate-user-vlan primary
 isolate-user-vlan l2-mappingadd 100-101
!
vlan 100
 isolate-user-vlan community
!
vlan 101
 isolate-user-vlan isolated
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode isolate-user-vlan promiscuous
```

```
 switch-port isolate-user-vlan l3-mapping 99 add 100-101
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping    99 100
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping 99 100
!
interface GigabitEthernet 0/4
 switch-port
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping    99 101
!
interface GigabitEthernet 0/5
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface VLAN 99
 isolate-user-vlan mapping add 100,101
 ip address 192.168.99.1 255.255.255.0
```

- Device B configuration file

```
vlan 99
 isolate-user-vlan primary
 isolate-user-vlan l2-mapping add 100-101
!
vlan 100
 isolate-user-vlan community
!
vlan 101
 isolate-user-vlan isolated
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode isolate-user-vlan host
```

```
 switch-port isolate-user-vlan l2-mapping    99 101
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping 99 100
```

- Gateway configuration file

```
vlan range 1,99
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 99
 switch-port trunk allow-pass vlan only 1-4094
```

7. Common Errors

- No Layer 2 association is configured between the primary VLAN and the secondary VLANs of a PVLAN. As a result, a VLAN list fails to be added to an isolated port, a promiscuous port, or a community port when you configure the port.
- A host has been associated with a PVLAN pair when it is associated with another PVLAN pair, which will cause a configuration error.
- No Layer 3 association is configured for a PVLAN.

1.9.2 Configuring PVLAN on a Single Layer 3 Device

1. Requirements

As shown in Figure 1-3, in the hosting service operation network, enterprise hosts are connected to the network through the Layer 3 Device A. Main requirements are as follows:

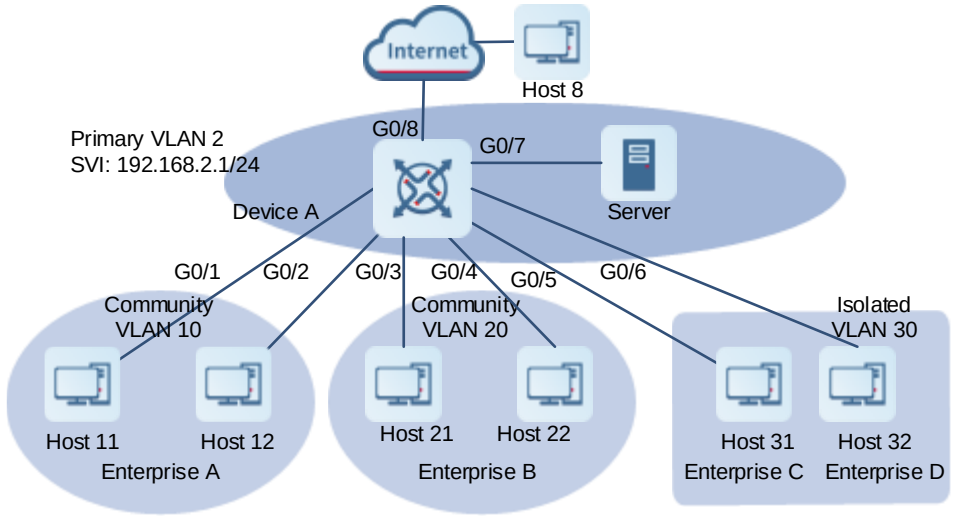
- Hosts of an enterprise can communicate with each other but the host communication between enterprises is isolated.
- All enterprise hosts share the same gateway address and can communicate with the external network through this gateway address.
- All enterprise hosts can access the server.

The deployment mode is as follows:

- Add all enterprise hosts to the same PVLAN (for example, primary VLAN 2). Assign hosts of enterprise A to community VLAN 10, hosts of enterprise B to community VLAN 20, and hosts of enterprises C and D to isolated VLAN 30.
- Configure the SVI gateway (for example, 192.168.2.1/24) of the PVLAN on the Layer 3 device A and configure the mappings between the primary VLAN and the secondary VLANs on the Layer 3. Then, all enterprise hosts can communicate with the external network through this gateway address.
- Configure the port directly connected to the server as a promiscuous port. All enterprise hosts can communicate with the server through the promiscuous port.

2. Topology

Figure 1-3 Configuring PVLAN Applied on a Single Layer 3 Device



3. Notes

- Configure the PVLAN function on Device A. VLAN 2 is the primary VLAN. VLAN 10, VLAN 20, and VLAN 30 are secondary VLANs.
- Configure GigabitEthernet 0/7 that is directly connected to the server as a promiscuous port. Then, all enterprise hosts can communicate with the server through the promiscuous port.
- Configure GigabitEthernet 0/8 of Device A that is connected to an external network as a promiscuous port. Set the IP address of the gateway SVI to 192.168.2.1/24 and configure a Layer 3 associations between the primary VLAN 2 and the secondary VLANs (VLAN 10, VLAN 20, and VLAN 30). All enterprise hosts can communicate with an external network based on this gateway IP address.

4. Procedure

(1) Configure a Layer 2 association between the primary VLAN and the secondary VLANs.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan 10
DeviceA(cfg-vlan)# isolate-user-vlan community
DeviceA(cfg-vlan)# exit
DeviceA(cfg)# vlan 20
DeviceA(cfg-vlan)# isolate-user-vlan community
DeviceA(cfg-vlan)# exit
DeviceA(cfg)# vlan 30
DeviceA(cfg-vlan)# isolate-user-vlan isolated
DeviceA(cfg-vlan)# exit
DeviceA(cfg)# vlan 2
DeviceA(cfg-vlan)# isolate-user-vlan primary
DeviceA(cfg-vlan)# isolate-user-vlan l2-mapping10,20,30
```

```
DeviceA(cfg-vlan)# exit
```

(2) Configure the ports connected to hosts as host ports and associate the ports with the primary VLAN and secondary VLANs to add the ports to the secondary VLANs.

```
DeviceA(cfg)# interface list gigabitethernet 0/1-2
DeviceA(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode isolate-user-vlan host
DeviceA(cfg-if-list)# switch-port isolate-user-vlan l2-mapping 2 10
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# interface list gigabitethernet 0/3-4
DeviceA(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode isolate-user-vlan host
DeviceA(cfg-if-list)# switch-port isolate-user-vlan l2-mapping 2 20
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# interface list gigabitethernet 0/5-6
DeviceA(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode isolate-user-vlan host
DeviceA(cfg-if-list)# switch-port isolate-user-vlan l2-mapping 2 30
DeviceA(cfg-if-list)# exit
```

(3) Configure GigabitEthernet 0/7 connected to the server as a promiscuous port and associate the port with the primary VLAN and the secondary VLANs to add the port to the primary VLAN.

```
DeviceA(cfg)# interface gigabitethernet 0/7
DeviceA(cfg-if-GigabitEthernet 0/7)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/7)# switch-port mode isolate-user-vlan promiscuous
DeviceA(cfg-if-GigabitEthernet 0/7)# switch-port isolate-user-vlan l3-maping 2 10,20,30
DeviceA(cfg-if-GigabitEthernet 0/7)# exit
```

(4) Configure GigabitEthernet 0/8 connected to an external network as a promiscuous port and associate the port with the primary VLAN and the secondary VLANs to add the port to the primary VLAN.

```
DeviceA(cfg)# interface gigabitethernet 0/8
DeviceA(cfg-if-GigabitEthernet 0/8)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/8)# switch-port mode isolate-user-vlan promiscuous
DeviceA(cfg-if-GigabitEthernet 0/8)# switch-port isolate-user-vlan l3-mapping 2 10,20,30
DeviceA(cfg-if-GigabitEthernet 0/8)# exit
```

(5) Create an SVI for the primary VLAN, configure an SVI address, and configure a Layer 3 association for the PVLAN.

```
DeviceA(cfg)# interface vlan 2
DeviceA(cfg-if-VLAN 2)# ip address 192.168.2.1 255.255.255.0
DeviceA(cfg-if-VLAN 2)# isolate-user-vlan l3-mapping 10,20,30
DeviceA(cfg-if-VLAN 2)# end
DeviceA# write
```

5. Verification

(1) Check whether the PVLAN and port configurations are correct.

```
DeviceA# display vlan isolate-user-vlan
VLAN    Type        Status    Routed    Ports              Associated VLANs
```

```
----- ---------  -------  -------  ------------  --------------------
2     primary     active   Enabled  Gi0/7, Gi0/8   10,20,30
10    community   active   Enabled  Gi0/1, Gi0/2   2
20    community   active   Enabled  Gi0/3, Gi0/4   2
30    isolated    active   Enabled  Gi0/5, Gi0/6   2
```

(2) Configure IP addresses and masks for hosts.

Configure an IP address and a mask for host 11.

```
Host11> privilege
Host11# config-system terminal
Host11(cfg)# interface gigabitethernet 0/1
Host11(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host11(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.11 255.255.255.0
```

Configure an IP address and a mask for host 12.

```
Host12> privilege
Host12# config-system terminal
Host12(cfg)# interface gigabitethernet 0/1
Host12(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host12(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.12 255.255.255.0
```

Configure an IP address and a mask for host 21.

```
Host21> privilege
Host21# config-system terminal
Host21(cfg)# interface gigabitethernet 0/1
Host21(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host21(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.21 255.255.255.0
```

(3) Check whether packets from the hosts in the PVLAN are forwarded in accordance with the forwarding rules of the PVLAN ports.

Hosts in the same community VLAN can communicate with each other. Hosts in different community VLANs cannot communicate with each other.

```
Host11# ping 192.168.2.12
Sending 5, 100-byte ICMP Echoes to 192.168.2.12, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms.
Host11# ping 192.168.2.21
Sending 5, 100-byte ICMP Echoes to 192.168.2.21, timeout is 2 seconds:
  < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5).
```

Hosts in an isolated VLAN cannot communicate with each other (the specific process is omitted).

Hosts in a community VLAN cannot communicate with hosts in an isolated VLAN (the specific process is omitted).

Hosts in a community VLAN and hosts in an isolated VLAN can communicate with an external network through the promiscuous port (the specific process is omitted).

Promiscuous ports can communicate with each other (the specific process is omitted).

6. Configuration Files

Device A configuration file

```
vlan 2
 isolate-user-vlan primary
 isolate-user-vlan l2-mapping add 10,20,30
!
vlan 10
 isolate-user-vlan community
!
vlan 20
 isolate-user-vlan community
!
vlan 30
 isolate-user-vlan isolated
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping 2 10
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping 2 10
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping 2 20
!
interface GigabitEthernet 0/4
 switch-port
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping 2 20
!
interface GigabitEthernet 0/5
 switch-port
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping 2 30
!
interface GigabitEthernet 0/6
 switch-port
```

```
 switch-port mode isolate-user-vlan host
 switch-port isolate-user-vlan l2-mapping 2 30
!
interface GigabitEthernet 0/7
 switch-port
 switch-port mode isolate-user-vlan promiscuous
 switch-port isolate-user-vlan l3-mapping 2 add 10,20,30
!
interface GigabitEthernet 0/8
 switch-port
 switch-port mode isolate-user-vlan promiscuous
 switch-port isolate-user-vlan l3-mapping 2 add 10,20,30
!
interface VLAN 2
isolate-user-vlan l3-mapping add 10,20,30
ip address 192.168.2.1 255.255.255.0
```

7. Common Errors

- No Layer 2 association is configured between the primary VLAN and the secondary VLANs of a PVLAN. As a result, configuring a Layer 3 association fails.
- No Layer 3 association is configured. As a result, hosts cannot communicate with the external network.
- The ports connecting to the server and the external network are not configured as promiscuous ports. Therefore, uplink and downlink configurations are asymmetric, causing a packet forwarding failure.

Contents

1 Configuring Super VLAN....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Background and Function....1
1.1.3 Principles....3
1.2 Configuration Task Summary....5
1.3 Configuring Layer 2 Association for a Super VLAN....6
1.3.1 Overview....6
1.3.2 Restrictions and Guidelines....6
1.3.3 Procedure....6
1.4 Configuring Layer 3 Communication for a Super VLAN....7
1.4.1 Overview....7
1.4.2 Restrictions and Guidelines....7
1.4.3 Procedure....8
1.5 Configuring an IP Address Range for a Sub VLAN....8
1.5.1 Overview....8
1.5.2 Restrictions and Guidelines....8
1.5.3 Procedure....9
1.6 Configuring a Broadcast VLAN....9
1.6.1 Overview....9
1.6.2 Restrictions and Guidelines....9
1.6.3 Procedure....9

1.7 Monitoring 9

1.8 Configuration Examples..... 10

1.8.1 Configuring Super VLAN 10

1 Configuring Super VLAN

Note

Super VLAN function is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.1 Introduction

1.1.1 Overview

Super virtual local area network (VLAN), also referred to as VLAN aggregation, is a VLAN assignment method.

A common VLAN corresponds to a subnet and is used to complete Layer 3 communication with another VLAN based on a gateway IP address.

The super VLAN technology allows you to configure common VLANs as sub VLANs, aggregate them into one super VLAN, assign only one IP address to the super VLAN as the gateway IP address to implement inter-VLAN Layer 3 communication. This technology saves IP address resources and facilitates network management.

1.1.2 Background and Function

1. Background

A common VLAN corresponds to a logical subnet and is isolated from other VLANs at Layer 2. Each subnet must be assigned with a gateway IP address to implement inter-VLAN Layer 3 communication. In addition to the gateway IP address, each subnet also requires a subnet ID and a subnet-directed broadcast IP address as dedicated addresses. These three addresses cannot be assigned to hosts.

Suppose that a local area network (LAN) covers three departments and therefore requires three VLANs. The VLAN IDs and hosts of the three departments are as follows: VLAN 10 (31 hosts), VLAN 20 (44 hosts), and VLAN 30 (128 hosts). There are 203 hosts in total. The network segment 192.168.1.0/24 has 256 IP addresses, which are assigned to the three departments to meet network requirements. A device performs a logical "AND" operation on IP addresses and subnet masks in binary format to identify the subnet, to which the hosts belong. The VLAN assignment result is as follows:

- VLAN 10 requires 31 host addresses and 3 dedicated addresses, totaling 34 addresses. VLAN 10 needs to be assigned with 64 IP addresses (2^6) with the subnet mask of /26.

  IP address range: 192.168.1.0 to 192.168.1.63 = 11000000.10101000.00000001.(00000000 to 00111111)

  Subnet mask (/26): 255.255.255.192 = 11111111.11111111.11111111.11000000

  Network ID obtained via logical "AND" operation in binary format: 192.168.1.0 = 11000000.10101000.00000001.00000000

  Gateway IP address: 192.168.1.1 = 11000000.10101000.00000001.00000001

  Subnet-directed broadcast IP address: 192.168.1.63 = 11000000.10101000.00000001.00111111

  Host IP address: 192.168.1.2 to 192.168.1.62 (31 hosts must be assigned with 61 addresses, and therefore 30 addresses are wasted)

- VLAN 20 requires 44 host addresses and 3 dedicated addresses, totaling 47 addresses. VLAN 20 needs to be assigned with 64 addresses (2^6) with the subnet mask of /26.

  IP address range: 192.168.1.64 to 192.168.1.127 = 11000000.10101000.00000001.(01000000 to 01111111)

  Subnet mask (/26): 255.255.255.192 = 11111111.11111111.11111111.11000000

  Network ID obtained via logical "AND" operation in binary format: 192.168.1.64 = 11000000.10101000.00000001.01000000

  Gateway IP address: 192.168.1.65 = 11000000.10101000.00000001.01000001

  Subnet-directed broadcast IP address: 192.168.1.127 = 11000000.10101000.00000001.01111111

  Host IP address: 192.168.1.66 to 192.168.1.126 (44 hosts must be assigned with 61 addresses, and therefore 17 addresses are wasted)

- Only 125 addresses are available to VLAN 30 that includes 128 hosts.

  IP address range: 192.168.1.128 to 192.168.1.255 = 11000000.10101000.00000001.(10000000 to 11111111)

  Subnet mask (/25): 255.255.255.128 = 11111111.11111111.11111111.10000000

  Network ID obtained via logical "AND" operation in binary format: 192.168.1.128 = 11000000.10101000.00000001.10000000

  Gateway IP address: 192.168.1.129 = 11000000.10101000.00000001.10000001

  Subnet-directed broadcast IP address: 192.168.1.255 = 11000000.10101000.00000001.11111111

  Host IP address: 192.168.1.130 to 192.168.1.254 (128 hosts are assigned with only 125 addresses)

It can be inferred from the assignment method that if each VLAN is assigned with a subnet, 9 dedicated addresses are required and 47 (30 + 17) host addresses are wasted. Though the network segment 192.168.1.0/24 contains 256 addresses, they cannot meet the IP address requirements of the 203 hosts.

This assignment method also lacks flexibility. If you add hosts to VLAN 20, exceeding the limit 64, you have to add the new hosts to another VLAN and assign IP addresses of another network segment to the new hosts unless you change the original IP address assignment policy. As a result, hosts from the same department do not belong to the same network segment.

To avoid the IP address waste and address assignment management problems, the super VLAN technology is developed.

2. Function

The super VLAN technology is used to aggregate multiple sub VLANs into one super VLAN.

Layer 2 isolation is implemented by sub VLANs, and Layer 3 interconnection is implemented by super VLANs.

A super VLAN corresponds to a logical subnet, and all sub VLANs in the super VLAN belong to this subnet and share the same network ID, gateway IP address, and subnet-directed broadcast IP address. This saves IP address resources and improves network management flexibility.

3. Features

- Super VLAN

  A super VLAN corresponds to a network segment. Switch virtual interfaces (SVIs) can be created only in super VLANs, but no physical port can be added to the super VLANs. Super VLAN manages inter-VLAN

communication of sub VLANs by using the SVIs. A super VLAN cannot be considered as an IEEE 802.1Q VLAN. A super VLAN has only Layer 3 attributes of a common VLAN.

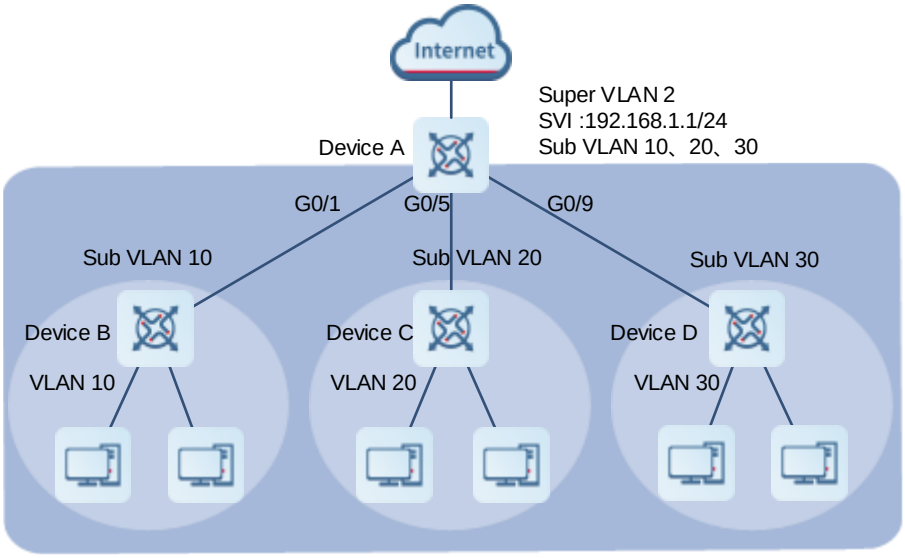
- Sub VLAN

  Each sub VLAN is an independent broadcast domain, and is isolated from other sub VLANs at Layer 2. A sub VLAN belongs to only one super VLAN. No SVI can be created in a sub VLAN, which can contain only physical ports. A sub VLAN has only Layer 2 attributes of a common VLAN. To implement Layer 3 communication across broadcast domains, sub VLANs must use the SVIs of the super VLAN.

4. Application

After the super VLAN technology is applied to the case described in Background, the network segment 192.168.1.0/24 can meet the requirement for IP addresses. As shown in Figure 1-1, on Device A, VLAN 2 is configured as a super VLAN, SVI address 192.168.1.1/24 as the gateway address, and VLAN 10, VLAN 20 and VLAN 30 as sub VLANs. Hosts are isolated at Layer 2 through sub VLANs and make Layer 3 communication with hosts in other network segments based on the common gateway IP address. The network segment accommodates IP addresses ranging from 192.168.1.2 to 192.168.1.254. A total of 253 IP addresses can be assigned to the 203 hosts in the sub VLANs, and there are some idle addresses left. Devices B, C, and D are used to expand the access ports and configured with VLAN 10, VLAN 20, and VLAN 30, respectively.

Figure 1-1 Application of Super VLAN



1.1.3 Principles

1. Communication in a Sub VLAN

A common VLAN corresponds to a subnet, and hosts in the subnet belong to the same broadcast domain. When a host (source host) in a common VLAN sends a packet, the host determines whether the destination host is in the same subnet as the source host. If yes, the host broadcasts an Address Resolution Protocol (ARP) request to request and obtain the MAC address of the destination host, and then completes Layer 2 communication with the destination host. If not, the host broadcasts an ARP request to request and obtain the MAC address of the gateway, and then forwards the packet to the gateway for Layer 3 communication.

In a super VLAN, when a host (source host) in a sub VLAN sends a packet, the host determines whether the destination host is in the same subnet as the source host and whether they belong to the same sub VLAN. If yes, the communication procedure is similar to the Layer 2 communication procedure of a common VLAN. The communication steps are as follows:

(1) The source host determines that the destination host is in the same subnet as the source host but the destination MAC address is unknown. The source host broadcasts an ARP request to request the destination MAC address.

(2) A switch learns the MAC address and forwards the ARP broadcast request to all ports except the source port in the sub VLAN.

(3) Upon receiving the ARP request, the destination host in the sub VLAN sends a unicast ARP reply to the source host.

(4) The switch learns the MAC address and forwards the ARP reply packet.

(5) Upon receiving the ARP reply packet, the source host obtains the destination MAC address and then can realize Layer 2 communication with the destination host.

2. Communication Between Different Sub VLANs in a Super VLAN

Though the sub VLANs in a super VLAN belong to the same subnet, they are isolated at Layer 2 and belong to different broadcast domains. When a host (source host) in a sub VLAN tries to communicate with a host (destination host) in another sub VLAN of the super VLAN, the communication initially follows a Layer 2 forwarding procedure because the source host does not know it is isolated from the destination host. In this case, communication between different sub VLANs relies on proxy ARP of a switch. The actual communication procedure is completed at Layer 3. The communication steps are as follows.

(1) The source host determines that the destination host is in the same subnet as the source host but the destination MAC address is unknown. The source host broadcasts an ARP request to request the destination MAC address.

(2) A switch learns the source MAC address and forwards the ARP request to all ports except the source port in the sub VLAN.

(3) Hosts in other sub VLANs cannot receive this ARP request. Therefore, no ARP reply is sent to the switch. After a specified period of time, the switch enables proxy ARP, replaces the source MAC address in the request packet with the gateway MAC address of the super VLAN, and broadcasts the ARP request in all sub VLANs of the super VLAN to request the destination MAC address.

Note

- The IP addresses in sub VLANs may be discontinuous and dynamic due to the address assignment policy of a Dynamic Host Configuration Protocol (DHCP) server or capacity expansion. This makes the super VLAN unable to identify the sub VLAN, to which the destination IP address belongs. Therefore, the broadcast packet is sent to all sub VLANs in the super VLAN by default.
- You can configure an IP address range for each sub VLAN so that the broadcast packet can be sent to a specific sub VLAN based on the IP address. In this case, the DHCP function must be disabled, because the IP addresses assigned by this function may fall out of the IP address range of a specific sub VLAN. The configured static IP addresses must be within the IP address range of a specific sub VLAN. Otherwise, communication between sub VLANs cannot be realized.
- You can configure a sub VLAN as a broadcast VLAN so that broadcast packets sent by the super VLAN are destined to this sub VLAN. However, this configuration may cause the failure of the destination host to receive the ARP broadcast request, resulting in a communication exception.

(4) Upon receiving the ARP request, the destination host in another sub VLAN learns the mapping between the gateway MAC address and the source IP address and sends a unicast ARP reply to the gateway.

(5) The switch learns the following mapping of the destination host: "IP address – MAC address – sub VLAN – port". The switch replaces the destination MAC address with the gateway MAC address and sends the unicast ARP reply to the source host.

(6) Upon receiving the ARP reply, the source host regards the gateway MAC address as the destination MAC address, learns the mapping between the gateway MAC address and the destination IP address, and completes Layer 3 communication through the proxy gateway.

The process in which the gateway replaces the source MAC address with the gateway MAC address to send an ARP request and the process in which the gateway replaces the destination MAC address with the gateway MAC address to send an ARP reply are referred to as proxy ARP. Communication across broadcast domains needs to rely on proxy ARP.

3. Communication Between Super VLANs

Different super VLANs correspond to different subnets and the destination host and source host in different super VLANs belong to different broadcast domains. The source host can determine that the destination host is in a different broadcast domain based on the IP address network segment. The communication process is similar to the communication between different network segments of common VLANs. In this communication procedure, the gateway also performs proxy ARP. However, the gateway is an SVI of the super VLAN rather than the SVI of a common VLAN.

4. Inter-Device Layer 2 Communication

Super VLAN works at Layer 3. The SVI of a super VLAN is used as a gateway only in Layer 3 communication across broadcast domains (across network segments or in the same network segment). The VLAN ID of a super VLAN is not used in Layer 2 communication.

No physical port can be added to a super VLAN. On a device where a super VLAN is configured, packets with containing VLAN ID of the super VLAN cannot pass through any port of the device.

As shown in Figure 1-1, VLAN 2 is configured as a super VLAN on Device A. Assume that VLAN 2 is configured as a common VLAN on Device B. When Device A receives a packet from VLAN 2 on Device B, Device A discards this packet.

The communication between sub VLAN 10 on Device A and VLAN 10 on Device B is inter-device Layer 2 communication. Their connected interfaces must allow packets of sub VLAN 10 to pass so that the packets can be forwarded by the interface. If packets are tagged, the VLAN ID in the tags is the VLAN ID of sub VLAN 10.

1.2 Configuration Task Summary

The configuration includes the following tasks:

(1) Configure a super VLAN and sub VLANs.

a Configuring Layer 2 Association for a Super VLAN

b Configuring Layer 3 Communication for a Super VLAN

c (Optional) Configuring an IP Address Range for a Sub VLAN

d (Optional) Configuring a Broadcast VLAN

(2) Configure ports and add the ports to the sub VLANs. For more information, see *Configuring VLAN*.

1.3 Configuring Layer 2 Association for a Super VLAN

1.3.1 Overview

You can configure a common VLAN as a super VLAN, configure other common VLANs as sub VLANs, and add the sub VLANs to the super VLAN. A sub VLAN is a broadcast domain, hosts in a sub VLAN can make Layer 2 communication with each other but are Layer 2 isolated from hosts outside the sub VLAN.

1.3.2 Restrictions and Guidelines

- Super VLAN
  - VLAN 1 cannot be configured as a super VLAN.
  - No physical port can be added to a super VLAN. After a common VLAN is configured as a super VLAN, the ports that have been added to the VLAN exit the VLAN and are added to VLAN 1.
  - A super VLAN does not belong to any physical port. Therefore, the device configured with the super VLAN cannot process packets that contain the super VLAN tag.
  - A super VLAN cannot be configured as a sub VLAN of another super VLAN. A sub VLAN of a super VLAN cannot be configured as a super VLAN.
  - You can use the **unset supervlan** command to restore a super VLAN and all its sub VLANs to common VLANs. You must restore a super VLAN to a common VLAN before deleting this super VLAN by running the **unset vlan** command.
- Sub VLAN
  - You must configure sub VLANs for a super VLAN so that the super VLAN is valid.
  - VLAN 1 cannot be configured as a sub VLAN.
  - You can configure multiple super VLANs on a device. A sub VLAN can be added to one super VLAN only and the sub VLANs in different super VLANs must be unique.
  - Avoid configuring too many sub VLANs. A broadcast packet of a super VLAN is sent to all its sub VLANs by default. The configuration of too many sub VLANs will degrade the performance.
  - Physical ports can be added to a sub VLAN. A common VLAN configured with a Layer 3 SVI cannot be configured as a sub VLAN.
  - You must restore a sub VLAN to a common VLAN by running the **unset subvlan** command before deleting this sub VLAN by running the **unset vlan** command. In super VLAN configuration mode, you can use the **unset subvlan** [*vlan-id-list*] command to restore a specific sub VLAN in the super VLAN to a common VLAN.

1.3.3 Procedure

(1) Create a common VLAN to be configured as a sub VLAN.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create a common VLAN to be configured as a sub VLAN.

vlan { *vlan-id* | **range** *vlan-range* }

Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

(2) Create a common VLAN to be configured as a super VLAN and enter the VLAN configuration mode.

vlan *vlan-id*

Only VLAN 1 exists by default.

(3) Configure a common VLAN as a super VLAN.

supervlan

No super VLAN is configured by default.

(4) Add the common VLAN to the super VLAN to make the common VLAN a sub VLAN of the super VLAN.

subvlan *vlan-id-list*

No common VLAN is added to a super VLAN as a sub VLAN by default.

1.4 Configuring Layer 3 Communication for a Super VLAN

1.4.1 Overview

A super VLAN corresponds to a logical subnet. No physical port can be added to a super VLAN and only a Layer 3 SVI can be configured for the super VLAN. IP addresses must be configured for Layer 3 SVIs. Sub VLANs of a super VLAN share the gateway address of the super VLAN and reside in the same network segment.

Proxy ARP must be enabled on the super VLAN and sub VLANs so that hosts in the sub VLANs can implement Layer 3 communication across broadcast domains through proxy ARP of the super VLAN.

1.4.2 Restrictions and Guidelines

- An SVI and an IP address must be configured for a super VLAN. The SVI is a virtual interface used for communication between sub VLANs. You cannot configure an SVI for a sub VLAN.
- In the **interface vlan** *interface-number* command, *interface-number* indicates the SVI number, which must be the ID of the VLAN to which the SVI belong. The value range is 1-4094.
- When you configure an SVI for a super VLAN, the system assigns a Layer 3 interface invisible to hosts to each sub VLAN in the super VLAN. The system may restore some sub VLANs to common VLANs due to resource deficiency. After an SVI is configured for a super VLAN, the attempt of adding sub VLANs may fail due to resource deficiency.
- Proxy ARP must be configured in VLAN configuration mode. Proxy ARP must be enabled on both the super VLAN and its sub VLANs so that Layer 3 communication can be implemented. Proxy ARP is enabled by default. To isolate communication between sub VLANs, you can use the **unset proxy-arp** command to disable proxy ARP.
- To configure an IPv4 address, see *Configuring IPv4*.
- To configure an IPv6 address, see *Configuring IPv6*.

1.4.3 Procedure

(1) Configure a gateway for a super VLAN.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create an SVI for the super VLAN and enter the SVI configuration mode.

interface vlan *interface-number*

No SVI exists in a VLAN by default.

d Configure an IP address and a subnet mask for the gateway of the super VLAN.

ip address { *ipv4-address mask* | *ipv4-address/mask-length* }

No IP address is configured for an SVI by default.

e Return to the global configuration mode.

exit

(2) (Optional) Enable proxy ARP.

a Enter the VLAN configuration mode of the super VLAN and sub VLAN.

vlan { *vlan-id* | **range** *vlan-range* }

b Enable proxy ARP for the super VLAN and the sub VLAN.

proxy-arp

Proxy ARP is enabled for the super VLAN and the sub VLAN by default.

1.5 Configuring an IP Address Range for a Sub VLAN

1.5.1 Overview

The gateway IP address and subnet mask of a super VLAN jointly determine the subnet range. A super VLAN sends a broadcast packet to all sub VLANs in the super VLAN by default. To limit the broadcast range, you can configure an IP address range for each sub VLAN so that the super VLAN can send a broadcast packet to a specific sub VLAN based on an IP address range. The IP addresses of hosts must be within the IP address range of the local sub VLAN. Otherwise, the hosts cannot make Layer 3 communication. Unless otherwise specified, you do not need to configure an IP address range for a sub VLAN.

1.5.2 Restrictions and Guidelines

- The IP address range of a sub VLAN must be within the gateway IP address range of the super VLAN that the sub VLAN belongs to.
- IP address ranges for different sub VLANs of a super VLAN cannot overlap.
- IP addresses (both static and dynamic IP addresses) of hosts in a sub VLAN must be within the IP address range of the sub VLAN. Otherwise, hosts in the sub VLAN cannot make Layer 3 communication.
- The IP addresses assigned by the DHCP function may fall out of the IP address range you specify for a sub VLAN. If the IP addresses assigned by the DHCP function fall out of the specified IP address range, hosts in

the sub VLAN cannot communicate with an external network. Therefore, you are not advised to configure an IP address range for a sub VLAN if DHCP is used to dynamically assign IP addresses to hosts.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the sub VLAN configuration mode.

vlan *vlan-id*

(4) Configure the IP address range for users in the sub VLAN.

subvlan-address-range *start-ipv4-address end-ipv4-address*

No IP address range is configured for users in the sub VLAN by default.

1.6 Configuring a Broadcast VLAN

1.6.1 Overview

A super VLAN sends a broadcast packet to all sub VLANs in the super VLAN by default. To limit the broadcast range, you can configure a sub VLAN as a broadcast VLAN so that broadcast packets of the super VLAN are sent in only this sub VLAN.

1.6.2 Restrictions and Guidelines

Only a sub VLAN of a super VLAN can be configured as a broadcast VLAN. Only one broadcast VLAN can be configured for a super VLAN. A later configured broadcast VLAN overwrites a previously configured broadcast VLAN.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the super VLAN configuration mode.

vlan { *vlan-id* | **range** *vlan-range* }

(4) Configure a broadcast VLAN.

bcast vlan *sub-vlan-id*

No broadcast VLAN exists by default. A super VLAN sends broadcast packets to all of its sub VLANs.

1.7 Monitoring

Run the **display** commands to check the running status of a configured function and verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging function immediately after use.

You can run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
display supervlan [**id** *super-vlan-id*]	Displays the configuration of super VLAN and sub VLAN.
debugging bridge svlan	Debugs the super VLAN function.

1.8 Configuration Examples

1.8.1 Configuring Super VLAN

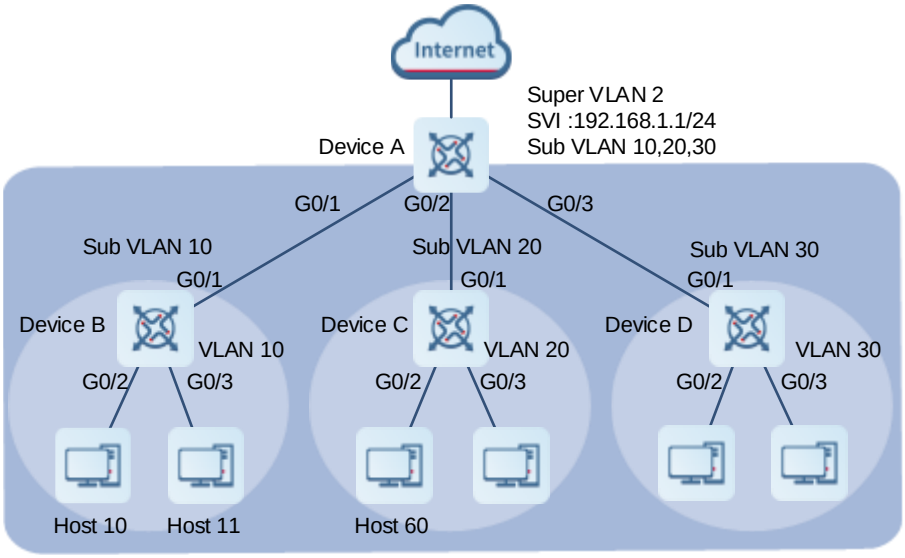
1. Requirements

A super VLAN 2 is configured and the network segment 192.168.1.0/24 is assigned to the super VLAN. Sub VLANs 10, 20 and 30 are added to the super VLAN. Hosts in the sub VLANs are in the same network segment and share a gateway IP address 192.168.1.1/24 to save IP addresses. Hosts are assigned static IP addresses. IP address ranges are configured for the three sub VLANs.

- Sub VLAN 10: 192.168.1.10/24 to 192.168.1.50/24
- Sub VLAN 20: 192.168.1.60/24 to 192.168.1.100/24
- Sub VLAN 30: 192.168.1.110/24 to 192.168.1.150/24

2. Topology

Figure 1-2 Configuring Super VLAN



3. Notes

- Configure a super VLAN and sub VLANs on Device A (core device).
- Configure the common VLANs on Devices B, C, and D (access devices) that correspond to the sub VLANs on Device A.

4. Procedure

(1) Create a common VLAN to be configured as a sub VLAN.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan range 10,20,30
DeviceA(cfg-vlan-range)# exit
```

(2) Create VLAN 2, configure VLAN 2 as a super VLAN, configure VLANs 10, 20 and 30 as sub VLANs of the super VLAN, and configure VLAN 10 as a broadcast VLAN of the super VLAN.

```
DeviceA(cfg)# vlan 2
DeviceA(cfg-vlan)# supervlan
DeviceA(cfg-vlan)# subvlan 10,20,30
DeviceA(cfg-vlan)# bcast vlan 10
DeviceA(cfg-vlan)# exit
```

(3) Configure a gateway for the super VLAN.

```
DeviceA(cfg)# interface vlan 2
DeviceA(cfg-if-VLAN 2)# ip address 192.168.1.1 255.255.255.0
DeviceA(cfg-if-VLAN 2)# exit
```

(4) Configure IP address ranges for the sub VLANs.

```
DeviceA(cfg)# vlan 10
DeviceA(cfg-vlan)# subvlan-address-range 192.168.1.10 192.168.1.50
DeviceA(cfg-vlan)# exit
DeviceA(cfg)# vlan 20
DeviceA(cfg-vlan)# subvlan-address-range 192.168.1.60 192.168.1.100
DeviceA(cfg-vlan)# exit
```

```
DeviceA(cfg)# vlan 30
DeviceA(cfg-vlan)# subvlan-address-range 192.168.1.110 192.168.1.150
DeviceA(cfg-vlan)# exit
```

(5) Configure GigabitEthernet 0/1, 0/2, and 0/3 as trunk ports, and add them to the sub VLANs.

```
DeviceA(cfg)# interface list gigabitethernet 0/1,0/2,0/3
DeviceA(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode trunk
DeviceA(cfg-if-list)# end
DeviceA(cfg-if-list)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-list)# switch-port trunk allow-pass vlan all
DeviceA# write
```

5. Verification

(1) Check the super VLAN configuration.

```
DeviceA(cfg)# display supervlan
supervlan id   supervlan arp-proxy   bcast vlan   subvlan id   subvlan arp-proxy   subvlan ip range
------------   -------------------   ----------   ----------   -----------------   ----------------
2                    ON                    10           10            ON       192.168.1.10 - 192.168.1.50
                                                        20            ON       192.168.1.60 - 192.168.1.100
                                                        30            ON       192.168.1.110 - 192.168.1.150
```

(2) Check ports that are added to VLANs.

```
DeviceA(cfg)# display vlan
VLAN          Name                                Status      Ports
---------- -------------------------------- --------- -----------------------------------
         2 VLAN0002                         SUPER
        10 VLAN0010                         SUB         Gi0/1, Gi0/2, Gi0/3
        20 VLAN0020                         SUB         Gi0/1, Gi0/2, Gi0/3
        30 VLAN0030                         SUB         Gi0/1, Gi0/3, Gi0/3
```

(3) Verify that hosts in a sub VLAN can make Layer 2 communication with each other.

On host 10 in sub VLAN 10, set the default gateway address to the gateway IP address of the super VLAN 192.168.1.1 and the IP address in the IP address range of the sub VLAN (that is, 192.168.1.10 to 192.168.1.50).

```
Host10> privilege
Host10# config-system terminal
Host10(cfg)# ip route-config 192.168.1.0 255.255.255.0 192.168.1.1
Host10(cfg)# interface gigabitethernet 0/1
Host10(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host10(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.10 255.255.255.0
```

On host 11 in sub VLAN 10, set the default gateway address to the gateway IP address of the super VLAN 192.168.1.1 and the IP address in the IP address range of the sub VLAN (that is, 192.168.1.10 to 192.168.1.50).

```
Host11> privilege
Host11# config-system terminal
Host11(cfg)# ip route-config 192.168.1.0 255.255.255.0 192.168.1.1
```

```
Host11(cfg)#interface gigabitethernet 0/1
Host11(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host11(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.11 255.255.255.0
```

Verify that host 10 can ping host 11 no matter whether the gateway IP address of the super VLAN is configured on Device A.

```
Host10# ping 192.168.1.11
Sending 5, 100-byte ICMP Echoes to 192.168.1.11, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/2/3 ms.
```

Set the IP address of host 11 to 192.168.1.51/24, which falls out of the IP address range of the sub VLAN, and verify that the communication of host 11 in the sub VLAN is not affected.

```
Host11(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.51 255.255.255.0
Host10# ping 192.168.1.51
Sending 5, 100-byte ICMP Echoes to 192.168.1.51, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/21/97 ms.
```

(4) Verify that hosts in different sub VLANs can make Layer 3 communication but are isolated at Layer 2.

On host 60 in sub VLAN 20, set the default gateway address to the gateway IP address of the super VLAN 192.168.1.1 and the IP address in the IP address range of the sub VLAN (that is, 192.168.1.60 to 192.168.1.100).

```
Host60> privilege
Host60# config-system terminal
Host60(cfg)# ip route-config 192.168.1.0 255.255.255.0 192.168.1.1
Host60(cfg)# interface gigabitethernet 0/1
Host60(cfg-if-GigabitEthernet 0/1)# unset switch-port
Host60(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.60 255.255.255.0
```

Configure an SVI and a gateway IP address for super VLAN 2, and verify that host 60 (192.168.1.60/24) can ping host 10 (192.168.1.10/24) and that host 60 cannot ping host 11 (192.168.1.51/24) because host 11 falls out of the IP address range of the sub VLAN.

```
Host60# ping 192.168.1.10
Sending 5, 100-byte ICMP Echoes to 192.168.1.10, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/19/95 ms.
Host60# ping 192.168.1.51
Sending 5, 100-byte ICMP Echoes to 192.168.1.51, timeout is 2 seconds:
  < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5).
```

Delete the gateway IP address of super VLAN 2, and verify that host 60 (192.168.1.60/24) cannot ping host 10 (192.168.1.10/24).

```
DeviceA(cfg)# interface vlan 2
DeviceA(cfg-if-VLAN 2)# unset ip address
%notice: all IPv4 addresses on this interface are removed.
Host60# ping 192.168.1.10
Sending 5, 100-byte ICMP Echoes to 192.168.1.10, timeout is 2 seconds:
   < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5).
```

6. Configuration Files

- Device A configuration file

```
vlan 2
 supervlan
 subvlan 10,20,30
 bcast vlan 10
!
vlan 10
 subvlan-address-range 192.168.1.10 192.168.1.50
!
vlan 20
 subvlan-address-range 192.168.1.60 192.168.1.100
!
vlan 30
 subvlan-address-range 192.168.1.110 192.168.1.150
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094

!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-passvlan only 1-4094
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface VLAN 2
 ip address 192.168.1.1 255.255.255.0
```

- Device B configuration file

```
vlan range 1,10
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode access
 switch-port access vlan 10
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode access
 switch-port access vlan 10
```

- Device C configuration file

```
vlan range 1,20
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode access
 switch-port access vlan 20
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode access
 switch-port access vlan 20
```

- Device D configuration file

```
vlan range 1,30
!
interface GigabitEthernet 0/1
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094!
interface GigabitEthernet 0/2
 switch-port
```

```
 switch-port mode access
 switch-port access vlan 30
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode access
 switch-port access vlan 30
```

7. Common Errors

- An SVI is configured for a common VLAN before the VLAN is configured as a sub VLAN. As a result, the common VLAN cannot be configured as a sub VLAN of a super VLAN.
- No SVI and gateway IP address are configured for a super VLAN. Consequently, communication between sub VLANs and between sub VLANs and other VLANs fails.
- Proxy ARP is disabled on the super VLAN or sub VLANs. Consequently, communication between the sub VLANs fails.
- An IP address range is configured for a sub VLAN. However, the static or dynamic IP addresses allocated to hosts in the sub VLAN are not in this range. Consequently, the hosts cannot make Layer 3 communication.

Contents

1 Configuring GVRP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Background and Function 1
1.1.3 Principles 1
1.1.4 Basic Concepts 3
1.1.5 Packet Format 5
1.1.6 Protocols and Standards 6
1.2 Configuration Task Summary 6
1.3 Configuring Basic Features 7
1.3.1 Overview 7
1.3.2 Restrictions and Guidelines 7
1.3.3 Procedure 8
1.4 Configuring GVRP BPDU Tunnels 11
1.4.1 Overview 11
1.4.2 Restrictions and Guidelines 11
1.4.3 Procedure 11
1.5 Configuring Transparent Transmission of GVRP BPDU 12
1.5.1 Overview 12
1.5.2 Restrictions and Guidelines 12
1.5.3 Procedure 12
1.6 Monitoring 12

1.7 Configuration Examples.....13

1.7.1 Configuring Basic Functions of GVRP.....13

1.7.2 Configuring GVRP BPDU Tunnels.....19

1 Configuring GVRP

Note

GVRP is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.1 Introduction

1.1.1 Overview

Generic Attribute Registration Protocol (GARP) VLAN registration protocol (GVRP) is a protocol for dynamically configuring virtual local area network (VLAN) attributes. It propagates VLAN attributes through protocol packets, and implements automatic registration and deregistration of VLANs on an IEEE 802.1Q trunk port.

1.1.2 Background and Function

A network is huge and complex. It is impossible to learn the network topology and VLAN conditions in a short period of time, and manual configuration brings heavy workload and easily leads to errors. The GARP provides a mechanism to help the members of the same switching network to mutually distribute, propagate, and register information including VLAN and multicast address.

An application entity that complies with GARP is called a GARP application. At present, major GARP applications include GVRP and GARP Multicast Registration Protocol (GMRP). An application entity diffuses the attribute information through protocol data packets. The destination MAC address of protocol packets is the specific multicast MAC address. If a device supporting GARP receives packets of GARP application entities, it will distinguish different applications according to their destination MAC address, and then deliver the packets to different applications (GVRP or GMRP) for processing.

- GVRP is a GARP application. It can dynamically configure and diffuse VLAN attributes, and implements dynamic and automatic registration/deregistration of VLANs on an IEEE 802.1Q trunk port.
- GMRP is another GARP application. It provides a restricted multicast diffusion function similar to the IGMP probing technology.

With the VLAN automatic registration function of GVRP, you can configure the VLAN, and dynamically maintain the process of creating a VLAN and adding/deleting ports to/from a VLAN, ensuring the VLAN connectivity in the topology.

GVRP simplifies the VLAN configuration and management. It reduces the workload of manually configuring VLANs and adding ports to VLANs, and reduces the possibility of network disconnection due to inconsistent configuration.

1.1.3 Principles

1. Working Process

(1) GVRP advertising

If the GVRP function is enabled for a device and the advertising mode of the trunk port is **Normal**, the trunk port will regularly collect the VLAN information on the local port, encapsulate the information in GVRP

packets and send them to the peer device, informing the peer device of the VLAN that the trunk port has joined or exited. The VLAN information type is GVRP Message Types, including Join message, Leave message, and LeaveAll message.

(2) GVRP registration (registration/deregistration)

The destination MAC address of a message packet is the specific multicast MAC address. After GVRP is enabled, the trunk port can recognize GVRP packets from other devices and obtain the VLAN registration or deregistration information by parsing the packets. If the registration mode is **Normal**, the port will create and join a VLAN dynamically, or the port will exit the VLAN and update the local VLAN registration information dynamically.

(3) GVRP advertisement diffusion

The device can also propagate the local VLAN registration information to other devices. The VLAN registration information includes the manually configured static registration information on the local device and the dynamic registration information from other devices. In this way, all devices in a VLAN maintain consistent VLAN information.

2. GVRP Advertising Modes

On a device with the GVRP function enabled, the trunk port regularly collects or does not collect VLAN information of the local port according to the advertising mode of the port (**Normal** or **Non-Applicant**).

- **Normal** mode: A device collects the VLAN information, and encapsulates it in GVRP packets, and sends the packets to the peer device to externally advertise its VLAN information, including dynamic and static VLAN information. On such a basis, the peer device determines the VLAN to be created.
- **Non-Applicant** mode: A device does not collect the VLAN information or advertise its VLAN information.

3. GVRP Registration Mode

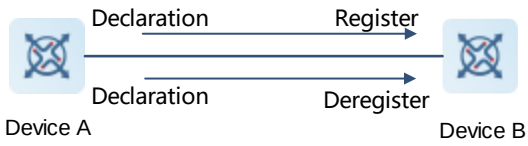
After receiving GVRP packets from the interconnected trunk port, the peer device parses the VLAN information and decides to process or not to process the VLAN information according to the port registration mode (**Normal** or **Disabled**).

- **Normal** mode: The device processes the VLAN information, dynamically creates a new VLAN and adds the port that receives the packet to the VLAN, or removes it from the VLAN.
- **Disabled** mode: The device does not process the VLAN information in the received GVRP packet.

4. VLAN Registration and Deregistration

Figure 1-1 shows the VLAN registration and deregistration process.

Figure 1-1 VLAN Registration and Deregistration



Device A sends a registration declaration to Device B through a GVRP advertisement, and the port of Device B receives the VLAN registration declaration and joins the VLAN contained in the declaration. This process is called VLAN registration.

Device A sends a deregistration declaration to Device B through a GVRP advertisement, and the port of Device B receives the VLAN deregistration declaration and exits the VLAN contained in the declaration. This process is called VLAN deregistration.

1.1.4 Basic Concepts

1. Static and Dynamic VLANs

- A manually created or deleted VLAN is a static VLAN.
- A VLAN dynamically created or deleted through GVRP is a dynamic VLAN.

The protocol state machine controls the process that a port joins or exits a dynamic VLAN. Only the trunk port that has received a GVRP registration or deregistration declaration can join or exit the dynamic VLAN created by GVRP; other trunk ports that have not received any GVRP registration or deregistration declaration cannot join or exit a dynamic VLAN, and even manual configuration cannot add or remove such a port to/from a dynamic VLAN.

The VLAN registration information propagated by GVRP includes the manually configured static registration information on the local device and the dynamic registration information from other devices. You can switch a dynamic VLAN to a static VLAN through configuration, but not the other way around.

2. Message Types

Terminals or bridge devices are GVRP members, and each port enabled with GVRP is called a GVRP application entity. The entities exchange information by transferring three types of messages.

- Join message

  The device sends a Join message in the following cases.

  - When a GVRP application entity wants other GVRP entities to register its own attribute information, it sends a Join message.
  - When a GVRP application entity receives a Join message from another GVRP entity, it sends a Join message for diffusion.
  - When a GVRP application entity is statically configured with attributes and requires other entities to register the attributes, it sends a Join message.

  Two types of Join messages are available: JoinEmpty and JoinIn. The differences between the two are as follows.

  - JoinEmpty message: Used to declare an unregistered attribute.
  - JoinIn message: Used to declare a registered attribute.

- Leave message

  The device sends a Leave message in the following cases.

  - When a GVRP application entity wants other GVRP entities to deregister its own attribute information, it sends a Leave message.
  - When a GVRP application entity receives a Leave message from another GVRP entity, it sends a Leave message.
  - When a GVRP application entity statically deregisters VLAN attributes and requires other entities to deregister the attributes, it sends a Leave message.

Two types of Leave messages are available: LeaveEmpty and LeaveIn. The differences between the two are as follows.

- LeaveEmpty message: Used to deregister an unregistered attribute.
- LeaveIn message: Used to deregister a registered attribute.

- LeaveAll message

  After GVRP is enabled, each GVRP application entity will enable its own LeaveAll timer. If the LeaveAll timer of a GVRP entity times out, it will send a LeaveAll message to deregister all attributes, which will cause other GVRP entities to re-register the attribute information. If a GVRP entity receives a LeaveAll message from another entity, it will send a LeaveAll message and restart the LeaveAll timer to initiate a new cycle.

3. Timer Types

GVRP defines four timers to control GVRP message sending.

- Hold timer

  The Hold timer defines the maximum delay interval *hold* before a port sends a Join message or Leave message. When a GVRP application entity has its attributes changed or receives a GVRP message from another entity, it enables the Hold timer. During the timeout period, the GVRP encapsulates all GVRP messages to be sent into packets as few as possible, and sends the packets when the timer times out. This reduces the quantity of sent packets and saves bandwidth resources. The interval is from 0 to *hold*.

- Join timer

  The Join timer defines the waiting time for Join message retransmission, conductive to ensuring the reliability of Join message transmission. After sending a Join message, the GVRP application entity will wait for a Join timer interval. If it receives a JoinIn message sent from another entity before the Join timer times out, it will not resend the Join message; otherwise, it will resend the Join message. Each GVRP application entity has one Join timer, and all attributes share one Join timer.

- Leave timer

  The Leave timer defines the waiting time from receiving a Leave message by the port to deleting the port from the VLAN, which helps to avoid frequent network changes. If the GVRP application entity wants other entities to deregister its VLAN attribute information, it will send a Leave message; the entity receiving the message will enable the Leave timer, and, if the port does not receive the Join message of the attribute information after the Leave timer times out, the state of the port becomes **Empty** and the port is removed from the VLAN member list; if the port receives the Join message of the attribute information before the Leave timer times out, the port will not be removed from the VLAN and the timer will time out.

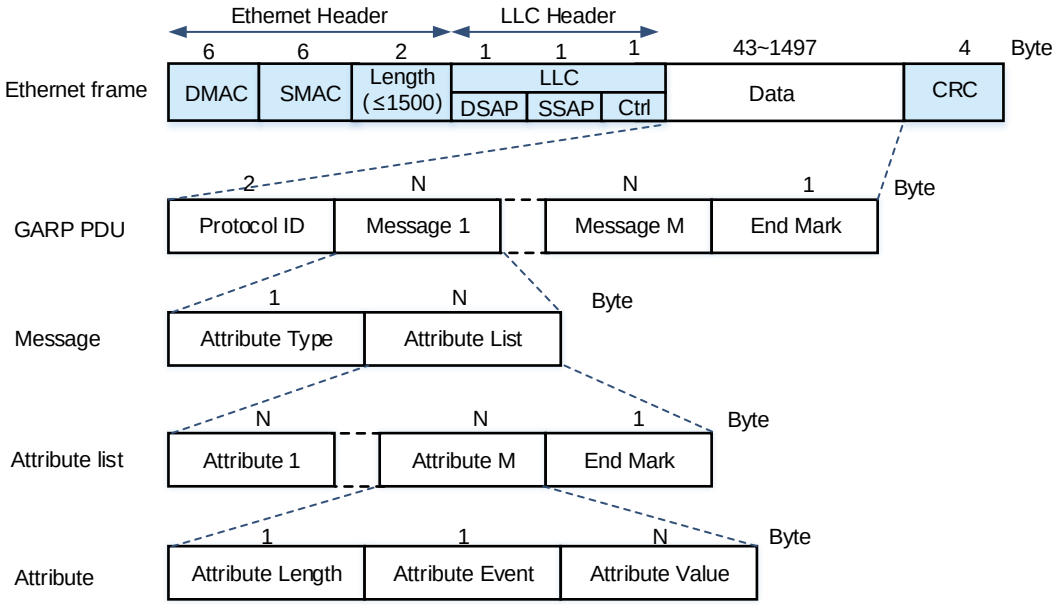
- LeaveAll timer

  The LeaveAll timer defines the minimum time interval *leaveall* for the port to send a LeaveAll message. Each GVRP application entity enables its own LeaveAll timer when the entity is enabled. If the LeaveAll timer times out, the GVRP application entity will send a LeaveAll message to instruct other entities to re-register the attribute information. The interval ranges from *leaveall* to the sum of *leaveall* and *hold*. The LeaveAll message is also sent to the local port to trigger the Leave timer to start timing again and initiate a new cycle; if the port receives the LeaveAll message before the timer times out, the timer restarts timing.

1.1.5 Packet Format

The GVRP protocol packets are encapsulated in Ethernet frames. Figure 1-2 shows the packet format.

Figure 1-2 GVRP Packet Format of Ethernet Frame



The following table describes the meaning of each GVRP Ethernet frame field.

Table 1-1 Meanings of GVRP Ethernet Frame Fields

Field	Number of Bytes	Description
Protocol ID	2	Indicates the protocol ID, 0x0001.
Message	N	Indicates 1 to M (M is a positive integer) messages. Each message consists of the attribute type and attribute list.
Attribute Type	1	Indicates the attribute type, defined by the specific GARP application. 0x01: Indicates GVRP application, and **Attribute Value** is **VLAN ID**.
Attribute List	N	Indicates the attribute list, composed of 1 to M (M is a positive integer) attributes and the end mark.
Attribute	N	Indicates an attribute. Each attribute consists of the attribute length, attribute event, and attribute value.
Attribute Length	1	Indicates the attribute length. The value range is from 0x02 to 0xFF, that is, from 2 to 255.

Field	Number of Bytes	Description
Attribute Event	1	Indicates the event type described by the attribute. 0x00: LeaveAll Event 0x01: JoinEmpty Event 0x02: JoinIn Event 0x03: LeaveEmpty Event 0x04: LeaveIn Event 0x05: Empty Event
Attribute Value	N	Indicates the attribute value. The attribute value of GVRP is **VLAN ID**, but the attribute value of **LeaveAll** is invalid.
End Mark	1	Indicates the end mark of GVRP PDU, 0x00.

1.1.6 Protocols and Standards

- IEEE 802.1D: IEEE Standard for Local and metropolitan area networks: Media Access Control (MAC) Bridges
- IEEE 802.1Q: IEEE Standard for Local and Metropolitan Area Networks-Virtual Bridged Local Area Networks

1.2 Restrictions and Guidelines

- GVRP function can only be enabled on Trunk ports. Only Layer 2 Ethernet interfaces and Layer 2 aggregate interface can be configured as Trunk ports.
- You can use the **display interface switch-port** command to view the interface type. if the **Switch-port** field corresponding to the interface is "enabled", it means that the interface is a Layer 2 interface. If the Switchport field corresponding to the interface is "disabled", it means that the interface is a Layer 3 interface.
- You can use the **switch-port** command to configure the interface as a Layer 2 interface. Use the **unset switch-port mode** command to restore the Layer 2 interface mode to the default value, and then use the **unset switch-port** command to configure the Layer 2 interface as a Layer 3 interface. When the Layer 2/Layer 3 interface mode is changed, all properties of the interface will be restored to the default values of the interface type.

1.3 Configuration Task Summary

- Configuring GVRP includes the following tasks:
- When the GVRP function is enabled, the configuration includes the following tasks:
  - a Configuring Basic Features
  - b (Optional) Configuring GVRP advertising mode
  - c (Optional) Confgiruing GVRP registration mode

d (Optional) Configuring GVRP BPDU Tunnels

- When the GVRP function is disabled, the configuration includes the following task:
  - (Optional) Configuring Transparent Transmission of GVRP BPDU

1.4 Configuring Basic Features

1.4.1 Overview

When the GVRP function is enabled, the device can send GVRP packets containing the VLAN information, dynamically create or delete VLANs and add or remove ports to/from VLANs. Devices synchronize respective VLAN information to maintain communication in the topology. This function can reduce the manual configuration workload and simplify VLAN management. When the GVRP function is disabled, the device will not send or handle GVRP packets containing the VLAN information.

Dynamic VLAN creation must be enabled. When the VLAN indicated in the packet received by the port (only JoinIn and JoinEmpty) does not exist on the local device, GVRP may create this VLAN. If this function is enabled, when the device receives the Join message from GVRP, a VLAN will be created dynamically.

1.4.2 Restrictions and Guidelines

- GVRP needs to be enabled on the two devices connected for communication. The GVRP function takes effect only on trunk ports. All the VLANs created by GVRP are tagged VLANs. GVRP information is transmitted only by trunk Links. The transmitted information contains the information on all VLANs on the current device, including dynamically learned VLANs and manually configured VLANs. The VLAN information that is dynamically learned by GVRP is saved in the memory temporarily, but not in the flash memory of the system. When the device is reset, all the information will be lost. Users cannot save the VLAN information that is dynamically learned.
- The relationship between GVRP and Spanning Tree Protocol (STP) is as follows.
  - If STP is not enabled, all available ports can participate in GVRP.
  - If STP is enabled, only ports in **Forwarding** state participate in GVRP (receiving and sending GVRP BPDU packets) and have their VLAN information diffused by GVRP.
  - If Multi Spanning Tree Protocol (MSTP) is enabled, GVRP can run in the spanning tree context to which VLAN1 belongs. You cannot specify other spanning tree context for GVRP.
  - If Single Spanning Tree Protocol (SSTP) is enabled, only ports in **Forwarding** state in the current spanning tree context participate in GVRP.
  - The parameters of a dynamic VLAN created through GVRP cannot be modified manually.
- Configure the timers.
  - GVRP uses the Hold, Leave, and LeaveAll timers to define the sending interval of various types of packets. The timers are controlled by the GVRP state machine and can be triggered by each other. The three timers are subject to this relationship: 3 x Hold ≤ Leave ≤ LeaveAll.
  - All devices that need to exchange GVRP information on the network must have consistent GVRP timer configuration; otherwise, the GVRP function will not take effect.
- Enable the GVRP function.
  - When GVRP is not enabled globally, GVRP configuration not take effect. Only when GVRP is enabled

globally, will GVRP configuration take effect.

- The information of super VLAN and sub VLAN cannot be synchronized to the interconnected devices through GVRP.

1.4.3 Prerequisites

- Configure the interface as a Layer 2 interface.

1.4.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of creating VLANs dynamically.

gvrp dynamic-vlan-creation enable

Creating VLANs dynamically is disabled by default.

(4) (Optional) Configure the GVRP timer.

gvrp timer-set { **join** *hold* | **leave** *leave* | **leaveall** *leaveall* }

By default, the maximum delay before a port sends a Join or Leave message is 200 ms, the waiting time from receiving a Leave message by the port to deleting the port from the VLAN is 600 ms, and the minimum time interval for the port to send a LeaveAll message is 10,000 ms.

a Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.
  - **interface** { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

b Configure the Layer 2 interface as a trunk port (global GVRP function only takes effect on a trunk port).

switch-port mode trunk

A Layer 2 interface works in access mode by default.

c Return to the global configuration mode.

exit

(5) Enable the GVRP function globally.

gvrp enable

The GVRP function is disabled by default.

1.5 Configure GVRP Advertising Mode

1.5.1 Overview

On the device with GVRP enabled, the Trunk port collects or does not collect VLAN information of the port regularly according to the port's notification mode (Normal or Non Application). You can use this function to configure the notification mode of the interface.

- Normal: Collect VLAN information, encapsulate these VLAN information in GVRP message, send notification to the peer device, and externally notify the VLAN information on the device, including dynamic and static VLAN information, according to which the peer device determines the VLAN to be created.
- Non-Applicant: Do not collect the VLAN information of local port, and do not notify the VLAN information on local device.

1.5.2 Restrictions and Guidelines

- Use the **gvrp applicant state normal** command to enable the configuration to collect port VLAN information and send notifications.
- Use the **gvrp applicant state non-applicant** command to configure not to collect port VLAN information and not send notifications.

1.5.3 Prerequisites

- The interface is configured as a Layer 2 interface.
- Basic functions of GVRP are configured, including: enabling the dynamic VLAN creation function, configuring the Layer 2 interface as a Trunk port, and enabling the GVRP function globally.

1.5.4 Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

a Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.
  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }
- Enter the Layer 2 aggregate interface configuration mode.
  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

b Configure the advertising mode on the interface.
gvrp applicant state { **normal** | **non-applicant** }

By default, the port advertising mode is Normal, the notification will be sent.

1.6 Configure GVRP Registration Mode

1.6.1 Overview

After receiving the GVRP message from the interconnected Trunk port, the peer device parses the VLAN information and chooses to process or not process the VLAN information according to the port registration mode (Normal or Disabled). You can use this function to configure the registration mode of the interface.

- Normal: Process VLAN information, dynamically create a non-existent VLAN and add the port receiving the message to the VLAN, or exit the VLAN.
- Disabled: The VLAN information in the message is not processed.

These two registration modes will not affect the static VLAN on the port. The static VLAN created by the user is always a Fixed Registrar.

1.6.2 Restrictions and Guidelines

- Use the **gvrp registration mode normal** command to allow dynamically create, register, or unregister VLANs on a port.
- Use the **gvrp register mode disable** command to disallow dynamically create, register, or unregister VLANs on a port.
- If enabled, it creates a dynamic VLAN after receiving the peer VLAN information. If disabled, when GVRP message is received, dynamic VLAN will not be added.

1.6.3 Prerequisites

- The interface is configured as a Layer 2 interface.
- Basic functions of GVRP are configured, including: enabling the dynamic VLAN creation function, configuring the Layer 2 interface as a Trunk port, and enabling the GVRP function globally.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

a Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.
  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }
- Enter the Layer 2 aggregate interface configuration mode.
  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

b Configure the registration mode of the port.
gvrp registration mode { **normal** | **disable** }

By default, the port registration mode is Normal, that is, the VLAN information is dynamically updated.

1.7 Configuring GVRP BPDU Tunnels

1.7.1 Overview

In the QinQ network, if GVRP of the user network and the SP network needs to be calculated separately without interfering with each other, you can configure the GVRP BPDU tunnel on the interface connected to the user network, so that the GVRP packets of the user network can be transmitted through the SP network tunnel.

The destination MAC address of GVRP BPDU packets is the dedicated multicast address 0180.c200.0006. To avoid impact on the SP network by GVRP packets of the user network, when GVRP packets of the user network enter an edge device of the SP network, the edge device changes the destination MAC address of the packets from the GVRP dedicated address (0180.c200.0006) to the tunnel address (01d0.f800.0006 by default) before forwarding on the SP network. After the packets are forwarded to an edge device at the other end, the destination MAC address is restored from the tunnel address (01d0.f800.0006 by default) to the GVRP dedicated address (0180.c200.0006), and the packets are forwarded to the user network at the other end. In this way, the GVRP packets of the user network are transmitted through the tunnel of the SP network.

1.7.2 Restrictions and Guidelines

- The tunnel function takes effect only when the GVRP BPDU tunnel is enabled in both global configuration mode and interface configuration mode. In the global configuration mode, run the **l2protocol-tunnel gvrp** command to enable the GVRP BPDU tunnel function globally. In Layer 2 Ethernet interface configuration mode or the Layer 2 aggregate interface configuration mode, run the **l2protocol-tunnel gvrp enable** command to enable the GVRP BPDU tunnel function on a port.
- The optional tunnel addresses of GVRP packets are 01d0.f800.0006 (default value) and 011a.a900.0006.

1.7.3 Prerequisites

- The interface connecting the user network is configured as a Layer 2 interface.
- The interface connecting the user network is configured as a Dot1q-tunnel interface.
- The QinQ encapsulation strategy is configured on the interface connecting to the user network.

1.7.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure an address for the GVRP tunnel.

l2protocol-tunnel gvrp tunnel-dmac *mac-address*

The GVRP tunnel address is 01d0.f800.0006 by default.

(4) Enable the GVRP BPDU tunnel function globally.

l2protocol-tunnel gvrp

The global GVRP BPDU tunnel function is disabled by default.

(5) Enter the interface configuration mode to configure the interface connecting to the user network.

- Enter the Layer 2 Ethernet interface configuration mode.

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(6) Enable GVRP BPDU tunnel function on the interface.

l2protocol-tunnel gvrp enable

By default, GVRP BPDU tunnel function of the interface is disabled.

1.8 Configuring Transparent Transmission of GVRP BPDU

1.8.1 Overview

In IEEE 802.1Q, the destination MAC address of GVRP BPDU packets is the dedicated multicast address 0180.c200.0006. Generally, devices complying with IEEE 802.1Q do not forward GVRP BPDU frames with the destination MAC address 0180.c200.0006, no matter whether GVRP is enabled. If the device has GVRP enabled, it will use GVRP BPDU frames for GVRP calculation; otherwise, it will discard GVRP packets.

However, in specific applications, devices may be required to transparently transmit GVRP packets. For example, a device does not have GVRP enabled, but the device interconnected to it has GVRP enabled. To enable the devices at both ends to perform GVRP calculation normally, the local device needs to transparently transmit GVRP BPDU frames. In this case, the GVRP BPDU transparent transmission function can be enabled on interface connecting two devices to allow the interface to forward the packets with the destination MAC address of 0180.c200.0006.

1.8.2 Restrictions and Guidelines

- The GVRP BPDU transparent transmission function takes effect only when GVRP is disabled.

When the device has GVRP enabled, this function cannot be used to transparently transmit GVRP BPDU frames. For example, in the QinQ network, both the SP network device and the user network device have GVRP enabled. If the SP network device needs to forward GVRP packets of the user network but does not forward GVRP packets of the SP network, the GVRP BPDU tunnel function is needed.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the transparent transmission function of GVRP BPDU packets.

bridge-frame forwarding protocol gvrp

The transparent transmission function of GVRP BPDU packets is disabled by default.

1.9 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging function immediately after use.

You can run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-2 Monitoring

Command	Purpose
cleanup gvrp statistics { *interface-type interface-number* | **all** }	Clears port statistics.
display gvrp statistics { *interface-type interface-number* | **all** }	Displays port statistics.
display gvrp status	Displays the current GVRP status.
display gvrp configuration	Displays the current GVRP configuration.
display l2protocol-tunnel gvrp	Displays the GVRP BPDU tunnel information.
debugging gvrp event	Enables the GVRP event debugging function.
debugging gvrp timer	Enables the GVRP timer debugging function.

1.10 Configuration Examples

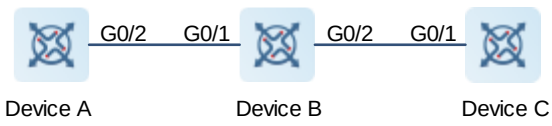
1.10.1 Configuring Basic Functions of GVRP

1. Requirements

As indicated in Figure 1-3, The GVRP function is enabled in a topology to dynamically maintain the VLANs and the VLAN-port relationship.

2. Topology

Figure 1-3 Configuring Basic Functions of GVRP



3. Notes

- On Devices A and C, configure VLANs used for communication on the user network.
- Enable GVRP on Devices A, B, and C, and enable the dynamic creation function of VLAN.
- Configure the ports connected between Devices A, B, and C as trunk ports, and ensure that the permitted VLAN lists of trunk ports include the communication VLANs.
- You are advised to enable STP to avoid loops.

4. Procedure

(1) Create VLANs for user network communication.

Create VLANs 2 to 10 for user network communication on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan range 2-10
DeviceA(cfg-vlan)# exit
```

Create VLANs 11 to 20 for user network communication on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# vlan range 11-20
DeviceC(cfg-vlan)# exit
```

(2) Enable the dynamic creation function of VLAN.

Enable the dynamic creation function of VLAN on Device A.

```
DeviceA(cfg)# gvrp dynamic-vlan-creation enable
```

Enable the dynamic creation function of VLAN on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# gvrp dynamic-vlan-creation enable
```

Enable the dynamic creation function of VLAN on Device C.

```
DeviceC(cfg)# gvrp dynamic-vlan-creation enable
```

(3) Configure the ports for connecting to the devices as trunk ports.

Configure the ports for the interconnected devices as trunk ports on Device A.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

Configure the ports for the interconnected devices as trunk ports on Device B.

```
DeviceB(cfg)# interface list gigabitethernet 0/1-2
DeviceB(cfg-if-list)# switch-port
DeviceB(cfg-if-list)# switch-port mode trunk
DeviceB(cfg-if-list)# switch-port trunk pvid vlan 1
```

```
DeviceB(cfg-if-list)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-list)# exit
```

Configure the ports for the interconnected devices as trunk ports on Device C.

```
DeviceC(cfg)# interface gigabitethernet 0/1
DeviceC(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceC(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
DeviceC(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
DeviceC(cfg-if-GigabitEthernet 0/1)# exit
```

(4) Enable the GVRP function.

Enable the GVRP function on Device A.

```
DeviceA(cfg)# gvrp enable
```

Enable the GVRP function on Device B.

```
DeviceB(cfg)# gvrp enable
```

Enable the GVRP function on Device C.

```
DeviceC(cfg)# gvrp enable
```

5. Verification

(1) Before the GVRP function is enabled globally, check the static VLAN configuration of the device; check whether the interconnection interface are trunk ports and allow all VLANs to pass.

Only static VLANs exist on Device A; the interconnection port GigabitEthernet 0/2 is a trunk port and allows all VLANs to pass.

```
DeviceA# display vlan
VLAN              Name                          Status             Ports
----------    ---------------------    ------------    --------------------------
           1    VLAN0001                      STATIC           Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                                 Gi0/5, Gi0/6, Gi0/7, Gi0/8
        2-10    VLAN0002-VLAN0010           STATIC           Gi0/2
DeviceA# display interface gigabitethernet 0/2 switch-port
Interface                       Switch-port Mode       Access Native Protected VLAN lists
-------------------------- ---------- --------- ------ ------ --------- --------------------
GigabitEthernet 0/2           enabled      TRUNK               1         Disabled   ALL
```

Only a static VLAN exists on Device B; the interconnection interface GigabitEthernet 0/1 and GigabitEthernet 0/2 are trunk ports and allow all VLANs to pass.

```
DeviceB# display vlan
VLAN              Name                          Status             Ports
----------    ---------------------    ------------    --------------------------
           1    VLAN0001                      STATIC           Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                                 Gi0/5, Gi0/6, Gi0/7, Gi0/8
DeviceB# display interface switch-port
Interface                       Switch-port Mode       Access Native Protected VLAN lists
-------------------------- ---------- --------- ------ ------ --------- --------------------
GigabitEthernet 0/1           enabled      TRUNK               1         Disabled   ALL
```

```
GigabitEthernet 0/2             enabled      TRUNK              1          Disabled   ALL
```

Only static VLANs exist on Device C; the interconnection port GigabitEthernet 0/1 is a trunk port and allows all VLANs to pass.

```
DeviceC# display vlan
VLAN              Name                          Status             Ports
----------    ---------------------    ------------    --------------------------
             1    VLAN0001                       STATIC           Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                                  Gi0/5, Gi0/6, Gi0/7, Gi0/8
         11-20    VLAN0011-VLAN0020          STATIC           Gi0/1
DeviceA# display interface gigabitethernet 0/1 switch-port
Interface                       Switch-port Mode        Access Native Protected VLAN lists
--------------------------- ---------- --------- ------ ------ --------- --------------------
GigabitEthernet 0/1             enabled      TRUNK              1          Disabled   ALL
```

(2) After the GVRP function is enabled globally, check the GVRP configuration of devices.

Check whether the GVRP function of Device A has been enabled, and the trunk port normally advertises the GVRP information and dynamically updates the VLAN information.

```
DeviceA# display gvrp configuration
Global GVRP Configuration:
GVRP Feature:enabled
GVRP dynamic VLAN creation:enabled
Join Timers(ms):200
Leave Timers(ms):600
Leaveall Timers(ms):1000
Port based GVRP Configuration:
            PORT                    Applicant Status            Registration Mode
----------------------  --------------------      ---------------------
GigabitEthernet 0/2               normal                        normal
```

Check whether the GVRP function of Device B has been enabled, and the trunk port normally advertises the GVRP information and dynamically updates the VLAN information.

```
DeviceB# display gvrp configuration
Global GVRP Configuration:
GVRP Feature:enabled
GVRP dynamic VLAN creation:enabled
Join Timers(ms):200
Leave Timers(ms):600
Leaveall Timers(ms):1000
Port based GVRP Configuration:
            PORT                    Applicant Status            Registration Mode
----------------------  --------------------      ---------------------
GigabitEthernet 0/1               normal                        normal
GigabitEthernet 0/2               normal                        normal
```

Check whether the GVRP function of Device C has been enabled, and the trunk port normally advertises the GVRP information and dynamically updates the VLAN information.

```
DeviceC# display gvrp configuration
```

```
Global GVRP Configuration:
GVRP Feature:enabled
GVRP dynamic VLAN creation:enabled
Join Timers(ms):200
Leave Timers(ms):600
Leaveall Timers(ms):1000
Port based GVRP Configuration:
          PORT                  Applicant Status          Registration Mode
---------------------- -------------------- ---------------------
GigabitEthernet 0/1               normal                    normal
```

(3) After the GVRP function is enabled globally, check the VLANs dynamically created by devices and check whether the trunk ports join dynamic VLANs.

Dynamic VLANs 11 to 20 are created on Device A, and the trunk port Gigabit Ethernet 0/2 joins the dynamic VLANs 11 to 20.

```
DeviceA(cfg)# display vlan
VLAN            Name                        Status          Ports
----------  ---------------------  ------------  --------------------------
         1  VLAN0001                    STATIC        Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                      Gi0/5, Gi0/6, Gi0/7, Gi0/8
      2-10  VLAN0002-VLAN0010         STATIC        Gi0/2
     11-20  DVLAN0011-DVLAN0020       DYNAMIC       Gi0/2
```

Dynamic VLANs 2 to 20 are created on Device B, and the trunk ports GigabitEthernet 0/1 and GigabitEthernet 0/2 join the dynamic VLANs 2 to 20.

```
DeviceB(cfg)# display vlan
VLAN            Name                        Status          Ports
----------  ---------------------  ------------  --------------------------
         1  VLAN0001                    STATIC        Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                      Gi0/5, Gi0/6, Gi0/7, Gi0/8
      2-20  DVLAN0002-DVLAN0020       DYNAMIC       Gi0/1, Gi0/2
```

Dynamic VLANs 2 to 10 are created on Device C, and the trunk port Gigabit Ethernet 0/1 joins the dynamic VLANs 2 to 10.

```
DeviceC(cfg)# display vlan
VLAN            Name                        Status          Ports
----------  ---------------------  ------------  --------------------------
         1  VLAN0001                    STATIC        Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                      Gi0/5, Gi0/6, Gi0/7, Gi0/8
      2-10  DVLAN0002-DVLAN0010       DYNAMIC       Gi0/1
     11-20  VLAN0011-VLAN0020         STATIC        Gi0/1
```

6. Configuration Files

- Device A configuration file

```
gvrp enable
gvrp dynamic-vlan-creation enable
!
```

```
vlan range 1-10
!
interface GigabitEthernet 0/2
   switch-port
  switch-port mode trunk
  switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

- Device B configuration file

```
gvrp enable
gvrp dynamic-vlan-creation enable
!
vlan range 1
!
interface GigabitEthernet 0/1
   switch-port
  switch-port mode trunk
  switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
   switch-port
  switch-port mode trunk
  switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

- Device C configuration file

```
gvrp enable
gvrp dynamic-vlan-creation enable
!
vlan range 1,11-20
!
interface GigabitEthernet 0/1
   switch-port
  switch-port mode trunk
  switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

7. Common Errors

- The interconnection interface of devices is not in trunk mode.
- The permitted VLAN list of the interconnection interface of devices does not contain the user communication VLANs.
- The GVRP advertising mode and registration mode of trunk ports are not set to **Normal**.

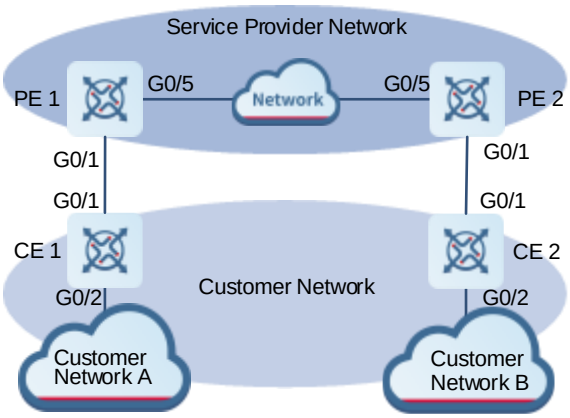
1.10.2 Configuring GVRP BPDU Tunnels

1. Requirements

As indicated in Figure 1-4, the GVRP packets of the user network are transmitted through the tunnel of the SP network.

2. Topology

Figure 1-4 Configuring GVRP BPDU Tunnel



3. Notes

- Configure the basic functions of GVRP on PE 1 and PE 2, including enabling the dynamic creation function of VLAN, configuring the interconnection port GigabitEthernet 0/5 as trunk port, and enabling the GVRP function globally.
- Enable the basic QinQ function on the GigabitEthernet 0/1 connected to the user network on PE 1 and PE 2 so that data packets of the user network are transmitted within VLAN 200 on the SP network.
- Enable the GVRP tunnel function on the GigabitEthernet 0/1 connected to the user network on PE 1 and PE 2 so that the SP network can transmit GVRP packets of the user network through the GVRP BPDU tunnel.
- Configure the basic functions of GVRP on CE 1 and CE 2, including: enabling the dynamic VLAN creation function, configuring the uplink interface Gigabit Ethernet 0/1 as the Trunk port, and enabling the GVRP function globally.

4. Procedure

(1) Configure the basic functions of GVRP on the interconnection interface of SP devices.

Enable the GVRP function globally on PE 1 and on Gigabit Ethernet 0/5 in turn. Configure static VLAN 251 to verify the GVRP function.

```
PE1> privilege
PE1# config-system terminal
PE1(cfg)# vlan 251
PE1(cfg-vlan)# exit
PE1(cfg)# gvrp dynamic-vlan-creation enable
PE1(cfg)# interface gigabitethernet 0/5
PE1(cfg-if-GigabitEthernet 0/5)# switch-port
PE1(cfg-if-GigabitEthernet 0/5)# switch-port mode trunk
```

```
PE1(cfg-if-GigabitEthernet 0/5)# switch-port trunk pvid vlan 1
PE1(cfg-if-GigabitEthernet 0/5)# switch-port trunk allow-pass vlan all
PE1(cfg-if-GigabitEthernet 0/5)# exit
PE1(cfg)# gvrp enable
```

Enable the GVRP function globally on PE 2 and Gigabit Ethernet 0/5 in turn. Configure static VLAN 252 to verify the GVRP function.

```
PE2> privilege
PE2# config-system terminal
PE2(cfg)# vlan 252
PE2(cfg-vlan)# exit
PE2(cfg)# gvrp dynamic-vlan-creation enable
PE2(cfg)# interface gigabitethernet 0/5
PE2(cfg-if-GigabitEthernet 0/5)# switch-port
PE2(cfg-if-GigabitEthernet 0/5)# switch-port mode trunk
PE2(cfg-if-GigabitEthernet 0/5)# switch-port trunk pvid vlan 1
PE2(cfg-if-GigabitEthernet 0/5)# switch-port trunk allow-pass vlan all
PE2(cfg-if-GigabitEthernet 0/5)# exit
PE2(cfg)# gvrp enable
```

(2) Configure the basic QinQ function on SP devices. Use VLAN 250 to encapsulate the packets of users at two ends.

Configure the basic QinQ encapsulation on the port Gigabit Ethernet 0/1 connected to the user on PE 1.

```
PE1(cfg)# vlan 250
PE1(cfg-vlan)# exit
PE1(cfg)# interface gigabitethernet 0/1
PE1(cfg-if-GigabitEthernet 0/1)# switch-port
PE1(cfg-if-GigabitEthernet 0/1)# switch-port mode dot1q-tunnel
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel pvid vlan 250
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel allow-pass vlan add untagged 250
```

Configure the basic QinQ encapsulation on the port Gigabit Ethernet 0/1 connected to the user on PE 2.

```
PE2(cfg)# vlan 250
PE2(cfg-vlan)# exit
PE2(cfg)# interface gigabitethernet 0/2
PE2(cfg-if-GigabitEthernet 0/1)# switch-port
PE2(cfg-if-GigabitEthernet 0/1)# switch-port mode dot1q-tunnel
PE2(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel pvid vlan 250
PE2(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel allow-pass vlan add untagged 250
```

(3) Enable the GVRP tunnel on SP devices.

Enable the GVRP tunnel on port Gigabit Ethernet 0/1 of PE 1 and globally in turn.

```
PE1(cfg-if-GigabitEthernet 0/1)# l2protocol-tunnel gvrp enable
PE1(cfg-if-GigabitEthernet 0/1)# exit
PE1(cfg)# l2protocol-tunnel gvrp
```

Enable the GVRP tunnel on port Gigabit Ethernet 0/1 of PE 2 and globally in turn.

```
PE2(cfg-if-GigabitEthernet 0/1)# l2protocol-tunnel gvrp enable
```

```
PE2(cfg-if-GigabitEthernet 0/1)# exit
PE2(cfg)# l2protocol-tunnel gvrp
```

(4) Configure the basic functions of GVRP on user network devices.

Enable the GVRP function globally on CE 1 and port Gigabit Ethernet 0/1 in turn.

```
CE1> privilege
CE1# config-system terminal
CE1(cfg)# vlan range 60-69
CE1(cfg-vlan)# exit
CE1(cfg)# gvrp dynamic-vlan-creation enable
CE1(cfg)# interface gigabitethernet 0/1
CE1(cfg-if-GigabitEthernet 0/1)# switch-port
CE1(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
CE1(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
CE1(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
CE1(cfg-if-GigabitEthernet 0/1)# exit
CE1(cfg)# gvrp enable
```

Enable the GVRP function globally on CE 2 and port Gigabit Ethernet 0/1 in turn.

```
CE2> privilege
CE2# config-system terminal
CE2(cfg)# vlan range 70-79
CE2(cfg-vlan)# exit
CE2(cfg)# gvrp dynamic-vlan-creation enable
CE2(cfg)# interface gigabitethernet 0/1
CE2(cfg-if-GigabitEthernet 0/1)# switch-port
CE2(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
CE2(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
CE2(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
CE2(cfg-if-GigabitEthernet 0/1)# exit
CE2(cfg)# gvrp enable
```

5. Verification

(1) Before the GVRP function is enabled globally, check the static VLAN configuration of the device; check whether the ports for advertising and processing GVRP packets are trunk ports and allow all VLANs to pass.

Only static VLANs exist on PE 1; the uplink port Gigabit Ethernet 0/5 is a trunk port and allows all VLANs to pass.

```
PE1# display vlan
VLAN        Name                                Status      Ports
---------- -------------------------------- --------- -----------------------------------
           1 VLAN0001                            STATIC      Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                             Gi0/5, Gi0/6, Gi0/7, Gi0/8
         250 VLAN0250                            STATIC      Gi0/1, Gi0/5
         251 VLAN0251                            STATIC      Gi0/5
PE1# display interface switch-port
Interface                    Switch-port Mode      Access Native Protected VLAN lists
```

```
------------------------ ---------- --------- ------ ------ --------- ----------------------
GigabitEthernet 0/1      enabled    TUNNEL             250    Disabled   1,250
GigabitEthernet 0/5      enabled    TRUNK              1      Disabled   ALL
```

Only static VLANs exist on PE 2; the uplink port Gigabit Ethernet 0/5 is a trunk port and allows all VLANs to pass.

```
PE2# display vlan
VLAN        Name                                Status     Ports
---------- -------------------------------- --------- -----------------------------------
           1 VLAN0001                            STATIC     Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                            Gi0/5, Gi0/6, Gi0/7, Gi0/8
         250 VLAN0250                            STATIC     Gi0/1, Gi0/5
         252 VLAN0252                            STATIC     Gi0/5
PE2# display interface switch-port
Interface                 Switch-port Mode      Access Native Protected VLAN lists
----------------------- ---------- --------- ------ ------ --------- ----------------------
GigabitEthernet 0/1       enabled    TUNNEL     1      250    Disabled   1,250
GigabitEthernet 0/5       enabled    TRUNK              1      Disabled   ALL
```

Only static VLANs exist on GE 1; the unlink port Gigabit Ethernet 0/1 is a trunk port and allows all VLANs to pass.

```
CE1# display vlan
VLAN        Name                                Status     Ports
---------- -------------------------------- --------- -----------------------------------
           1 VLAN0001                            STATIC     Gi0/1,Gi0/2, Gi0/3, Gi0/4
                                                            Gi0/5, Gi0/6, Gi0/7, Gi0/8
       60-69 VLAN0060-VLAN0069                   STATIC     Gi0/1
CE1# display interface switch-port
Interface                 Switch-port Mode      Access Native Protected VLAN lists
------------------------ ---------- --------- ------ ------ --------- ----------------------
GigabitEthernet 0/1       enabled    TRUNK              1      Disabled   ALL
```

Only static VLANs exist on CE 2; the uplink port Gigabit Ethernet 0/1 is a trunk port and allows all VLANs to pass.

```
CE2# display vlan
VLAN        Name                                Status     Ports
---------- -------------------------------- --------- -----------------------------------
           1 VLAN0001                            STATIC     Gi0/1,Gi0/2, Gi0/3, Gi0/4
                                                            Gi0/5, Gi0/6, Gi0/7, Gi0/8
       70-79 VLAN0070-VLAN0079                   STATIC     Gi0/1
CE2# display interface switch-port
Interface                 Switch-port Mode      Access Native Protected VLAN lists
------------------------ ---------- --------- ------ ------ --------- ----------------------
GigabitEthernet 0/1       enabled    TRUNK              1      Disabled   ALL
```

(2) Check whether the QinQ function is enabled, the SP device port connected to the user is the Dot1q-tunnel port, and the GVRP tunnel function is enabled.

Check the GVRP BPDU tunnel configuration on PE 1, and check whether the interface type is **Dot1q-tunnel**.

```
PE1# display interfaces dot1q-tunnel
========Interface Gi0/1========
Native vlan: 250
Allowed vlan list:1,250,
Tagged vlan list:

PE1# display l2protocol-tunnel gvrp
L2protocol-tunnel: Gvrp Enable
L2protocol-tunnel destination mac address: 01d0.f800.0006
GigabitEthernet 0/1 l2protocol-tunnel gvrp enable
```

Check the GVRP BPDU tunnel configuration on PE 2, and check whether the interface type is **Dot1q-tunnel**.

```
PE2# display interfaces dot1q-tunnel
========Interface Gi0/1========
Native vlan: 250
Allowed vlan list:1,250,
Tagged vlan list:

PE2# display l2protocol-tunnel gvrp
L2protocol-tunnel: Gvrp Enable
L2protocol-tunnel destination mac address:01d0.f800.0006
GigabitEthernet 0/1 l2protocol-tunnel gvrp enable
```

(3) After the GVRP function is enabled globally using the **gvrp enable** command, check the GVRP configuration of devices.

Check whether the GVRP function of PE 1 has been enabled, and the trunk port normally advertises the GVRP information and dynamically updates the VLAN information.

```
PE1# display gvrp configuration
Global GVRP Configuration:
GVRP Feature:enabled
GVRP Dynamic VLAN Creation:enabled
Join Timers(ms):200
Leave Timers(ms):600
Leaveall Timers(ms):10000
Port based GVRP Configuration:
          PORT                Applicant Status          Registration Mode
---------------------- ------------------- ---------------------
GigabitEthernet 0/5           normal                    normal
```

Check whether the GVRP function of PE 2 has been enabled, and the trunk port normally advertises the GVRP information and dynamically updates the VLAN information.

```
PE2# display gvrp configuration
Global GVRP Configuration:
GVRP Feature:enabled
```

```
GVRP Dynamic VLAN Creation:enabled
Join Timers(ms):200
Leave Timers(ms):600
Leaveall Timers(ms):10000
Port based GVRP Configuration:
          PORT                  Applicant Status          Registration Mode
----------------------  --------------------  --------------------
GigabitEthernet 0/5               normal                    normal
```

Check whether the GVRP function of CE 1 has been enabled, and the trunk port normally advertises the GVRP information and dynamically updates the VLAN information.

```
CE1# display gvrp configuration
Global GVRP Configuration:
GVRP Feature:enabled
GVRP Dynamic VLAN Creation:enabled
Join Timers(ms):200
Leave Timers(ms):600
Leaveall Timers(ms):10000
Port based GVRP Configuration:
          PORT                  Applicant Status          Registration Mode
----------------------  --------------------  --------------------
GigabitEthernet 0/1               normal                    normal
```

Check whether the GVRP function of CE 2 has been enabled, and the trunk port normally advertises the GVRP information and dynamically updates the VLAN information.

```
CE2# display gvrp configuration
Global GVRP Configuration:
GVRP Feature:enabled
GVRP Dynamic VLAN Creation:enabled
Join Timers(ms):200
Leave Timers(ms):600
Leaveall Timers(ms):10000
Port based GVRP Configuration:
          PORT                  Applicant Status          Registration Mode
----------------------  --------------------  --------------------
GigabitEthernet 0/1               normal                    normal
```

(4) After the GVRP function is enabled globally, check whether SP network devices can dynamically create VLANs on them; VLANs of user network will not be created dynamically. It indicates that the GVRP packets of user network do not affect the GVRP operation of SP devices.

Dynamic VLAN 252 has been created on PE 1, and the trunk port Gigabit Ethernet 0/5 joins the dynamic VLAN 252.

```
PE1(cfg)# display vlan
VLAN        Name                                  Status      Ports
---------- -------------------------------- --------- -----------------------------------
         1 VLAN0001                              STATIC      Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                             Gi0/5, Gi0/6, Gi0/7, Gi0/8
```

```
        250 VLAN0250                          STATIC      Gi0/1, Gi0/5
        251 VLAN0251                          STATIC      Gi0/5
        252 DVLAN0252                         DYNAMIC     Gi0/5
```

Dynamic VLAN 251 has been created on PE 2, and the trunk port Gigabit Ethernet 0/5 joins the dynamic VLAN 251.

```
PE2(cfg)# display vlan
VLAN          Name                            Status      Ports
---------- -------------------------------- --------- -----------------------------------
           1 VLAN0001                         STATIC      Gi0/1, Gi0/2, Gi0/3, Gi0/4
                                                          Gi0/5, Gi0/6, Gi0/7, Gi0/8
         250 VLAN0250                         STATIC      Gi0/1, Gi0/5
         251 DVLAN0251                        DYNAMIC     Gi0/5
         252 VLAN0252                         STATIC      Gi0/5
```

(5) After the GVRP function is enabled globally, check whether user network devices dynamically create VLANs of user network. It indicates that the GVRP packets of user network are transmitted through the tunnels of SP devices.

Dynamic VLANs 70 to 79 have been created on CE 1, and the trunk port Gigabit Ethernet 0/1 joins the dynamic VLANs 70 to 79.

```
CE1(cfg)# display vlan
VLAN          Name                            Status      Ports
---------- -------------------------------- --------- -----------------------------------
           1 VLAN0001                         STATIC      Gi0/1,Gi0/2, Gi0/3, Gi0/4
                                                          Gi0/5, Gi0/6, Gi0/7, Gi0/8
       60-69 VLAN0060-VLAN0069                STATIC      Gi0/1
       70-79 DVLAN0070-DVLAN0079              DYNAMIC     Gi0/1
```

Dynamic VLANs 60 to 69 have been created on CE 2, and the trunk port Gigabit Ethernet 0/1 joins the dynamic VLANs 60 to 69.

```
CE2(cfg)# display vlan
VLAN          Name                            Status      Ports
---------- -------------------------------- --------- -----------------------------------
           1 VLAN0001                         STATIC      Gi0/1,Gi0/2, Gi0/3, Gi0/4
                                                          Gi0/5, Gi0/6, Gi0/7, Gi0/8
       60-69 DVLAN0060-DVLAN0069              DYNAMIC     Gi0/1
       70-79 VLAN0070-VLAN0079                STATIC      Gi0/1
```

6. Configuration Files

- PE 1 configuration file

```
l2protocol-tunnel gvrp
gvrp enable
gvrp dynamic-vlan-creation enable
!
vlan range 1,250-251
!
interface GigabitEthernet 0/1
```

```
 switch-port mode dot1q-tunnel
 switch-port dot1q-tunnel allow-passvlan add untagged 250
 switch-port dot1q-tunnel pvid vlan 250
 l2protocol-tunnel gvrp enable
!
interface GigabitEthernet 0/5
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-passvlan only 1-4094
```

- PE 2 configuration file

```
l2protocol-tunnel gvrp
gvrp enable
gvrp dynamic-vlan-creation enable
!
vlan range 1,250,252
!
interface GigabitEthernet 0/1
 switch-port mode dot1q-tunnel
 switch-port dot1q-tunnel allow-passvlan add untagged 250
 switch-port dot1q-tunnel pvid vlan 250
 l2protocol-tunnel gvrp enable
!
interface GigabitEthernet 0/5
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

- CE 1 configuration file

```
gvrp enable
gvrp dynamic-vlan-creation enable
!
vlan range 1,60-69
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
```

- CE 2 configuration file

```
gvrp enable
gvrp dynamic-vlan-creation enable
!
vlan range 1,70-79
!
```

```
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
```

7. Common Errors

On the SP network, the configured GVRP BPDU tunnel addresses are inconsistent at two ends, causing that GVRP BPDU frames cannot be correctly transmitted.

Contents

1 Configuring QinQ....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Background and Function....1
1.1.3 QinQ Packet Format....1
1.1.4 QinQ Packet Forwarding....3
1.1.5 Basic QinQ Encapsulation....4
1.1.6 Selective QinQ Encapsulation....4
1.1.7 Modifying QinQ Tag....5
1.1.8 VLAN-Mapping....5
1.1.9 MAC Address Replication....5
1.1.10 Layer-2 Protocol Tunneling....6
1.1.11 Protocols and Standards....7
1.2 Restrictions and Guidelines....7
1.3 Configuration Task Summary....8
1.3.1 Configuring Tag Adding....8
1.3.2 Configuring Tag Modification....8
1.3.3 Configuring Layer-2 Protocol Tunneling....8
1.4 Notes....8
1.5 Configuring Basic QinQ Encapsulation....9
1.5.1 Overview....9
1.5.2 Restrictions and Guidelines....9

1.5.3 Prerequisites 9
1.5.4 Procedure 9
1.6 Configuring C-Tag-Based Selective QinQ Encapsulation 10
1.6.1 Overview 10
1.6.2 Restrictions and Guidelines 10
1.6.3 Prerequisites 11
1.6.4 Procedure 11
1.7 Configuring ACL-Based Selective QinQ Encapsulation 12
1.7.1 Overview 12
1.7.2 Restrictions and Guidelines 12
1.7.3 Prerequisites 12
1.7.4 Procedure 13
1.8 Configuring MAC Address Replication 14
1.8.1 Overview 14
1.8.2 Restrictions and Guidelines 14
1.8.3 Prerequisites 15
1.8.4 Procedure 15
1.9 Configuring Priority Replication or Priority Mapping 15
1.9.1 Overview 15
1.9.2 Restrictions and Guidelines 16
1.9.3 Configuration Tasks 16
1.9.4 Configuring Priority Replication 16
1.9.5 Configuring Priority Mapping 17
1.10 Configuring TPID 18

1.10.1 Overview 18
1.10.2 Restrictions and Guidelines 18
1.10.3 Prerequisites 18
1.10.4 Procedure 19
1.11 Configuring QinQ Tag Modification Policy 19
1.11.1 Overview 19
1.11.2 Restrictions and Guidelines 20
1.11.3 Prerequisites 20
1.12 Configuring VLAN-Mapping 21
1.12.1 Overview 21
1.12.2 Restrictions and Guidelines 23
1.12.3 Prerequisites 23
1.12.4 Procedure 23
1.13 Configuring Layer-2 Protocol Tunneling 24
1.13.1 Overview 24
1.13.2 Restrictions and Guidelines 24
1.13.3 Configuration Tasks 24
1.13.4 Configuring STP Packet Tunneling 25
1.13.5 Configuring GVRP Packet Tunneling 26
1.14 Monitoring 26
1.15 Configuration Examples 28
1.15.1 Configuring Basic QinQ to Implement Layer-2 VPN 28
1.15.2 Configuring C-Tag-Based Selective QinQ to Implement Layer-2 VPN and Service Flow Management 36

1.15.3 Configuring ACL-Based Selective QinQ to Implementing Layer-2 VPN and Service Flow Management 40

1.15.4 Configuring VLAN-Mapping to Implement Aggregation of Different Services 44

1.15.5 Configuring Layer-2 Protocol Tunneling Based on QinQ 53

1 Configuring QinQ

Note

QinQ is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.1 Introduction

1.1.1 Overview

When a client packet enters a provider edge (PE), the QinQ technology first encapsulates the packet with a public virtual local area network (VLAN) tag before transmitting the packet in the SP network. The private VLAN tag, if any, in the client packet is regarded as data. Then, the packet carries two VLAN tags to pass through the SP network.

1.1.2 Background and Function

1. Background

On a metropolitan area network (MAN), SPs need a large number of VLANs to isolate and identify clients so as to achieve refined management. As a variety of services develop, SPs also need to differentiate services to provide different transmission pipelines and QoS policies. However, IEEE 802.1Q supports 4094 VLANs only, far from meeting the needs of SPs.

2. Function

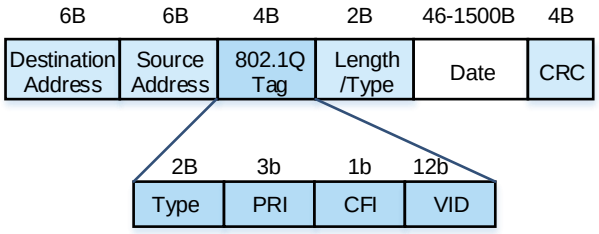
- QinQ is a kind of dual-tag encapsulation technology. It adds outer VLAN tags to packets in the SP network so that client VLANs in different private networks can be reused to increase the number of VLAN tags available to clients. The number of VLANs actually available to SPs is increased to 4094x4094.
- An outer tag is also equivalent to a tunnel, which enables transparent transmission of inner tags. This turns QinQ into a layer-2 VPN solution that is low in cost, easy to deploy, and suitable for small-sized networks.
- Inner and outer tags can also represent different information according to service requirements, helpful for SPs to segment and manage their services.

1.1.3 QinQ Packet Format

- Ethernet frame with a single VLAN tag

The IEEE 802.1Q standard modifies the Ethernet frame format by adding a 4-byte 802.1Q tag between the source MAC address field and protocol type field. Figure 1-1 shows the frame format. The tag contains a VID (VLAN ID) field, which is used to indicate the VLAN where the frame is. Not all devices can identify 802.1Q frames with tags, and therefore Ethernet frames on VLANs can be tagged or untagged.

Figure 1-1 Frame Format of 802.1Q Encapsulation

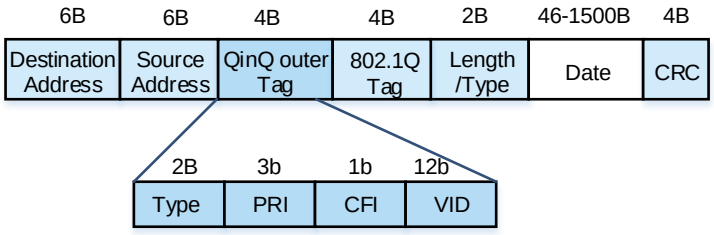


- **Type**: Consisting of 2 bytes; 0x8100 indicates 802.1Q frame. Devices that do not support IEEE 802.1Q will drop such frames received.
- **PRI**: Consisting of 3 bits, indicating layer-2 priority. The value range is 0 to 7, corresponding to CoS priorities of QoS. A larger value indicates a higher priority.
- **CFI**: Consisting of 1 bit, used to distinguish Ethernet frame, FDDI frame, and token ring network frame; 0 indicates Ethernet frame.
- **VID**: Consisting of 12 bits, indicating the client VLAN ID; values 0 and 4095 are reserved and unavailable. The value range is from 1 to 4094.

- Ethernet frame with QinQ dual VLAN tags

The QinQ function inserts a 4-byte QinQ outer tag (including four fields: **Type**, **PRI**, **CFI**, and **VID**) before the 802.1Q tag.

Figure 1-2 Frame Format of QinQ Encapsulation



- **Type**: Consisting of 2 bytes, also known as TPID; the default value is 0x8100, indicating 802.1Q frame. On the devices of some vendors, the TPID is set to 0x9100 or other values. TPID can be configured (Configuring TPID) to ensure that the TPIDs of packets to be forwarded are compatible with the TPIDs supported by third-party devices.
- **PRI**: Consisting of 3 bits, indicating priority; the default value 0 indicates a normal flow. The value of this field corresponds to the CoS value in the QoS policy. A larger value indicates a higher priority. The value of this field can be modified by configuring priority replication or priority mapping (Configuring Priority Replication or Priority Mapping) to define different transmission priority policies for packets.
- **CFI**: Consisting of 1 bit, used to distinguish Ethernet frame, FDDI frame, and token ring network frame; value 0 indicates Ethernet frame.
- **VID**: Consisting of 12 bits, indicating the SP VLAN ID. The value is from 1 to 4094.

1.1.4 QinQ Packet Forwarding

1. QinQ Encapsulation

The ingress of an SP's QinQ edge device is called dot1q-tunnel port. All the frames entering the edge device, no matter whether they carry the 802.1Q tag or not, will be packaged with the SP tag and then be forwarded. As shown in Figure 1-3, when a packet of CE-A1 is forwarded to PE 1, the dot1q-tunnel port encapsulates the packet with a VLAN 10 tag, and then forwards the packet from the uplink port. This process is called QinQ encapsulation. QinQ encapsulation is classified into Basic QinQ Encapsulation and Selective QinQ Encapsulation. Selective QinQ encapsulation is further classified into client tag-based encapsulation and ACL-based encapsulation.

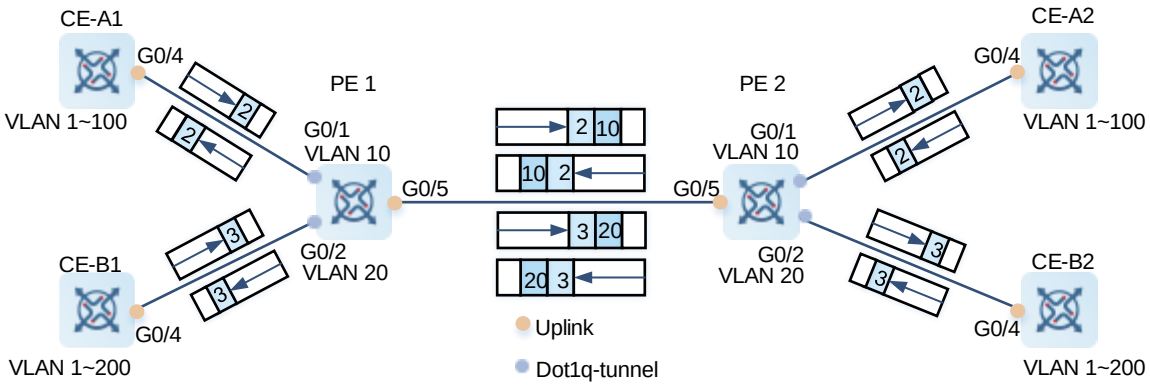
2. QinQ Forwarding

A QinQ-encapsulated packet will be transmitted in the SP network. Generally, devices directly forward the QinQ packet with dual tags. During the transmission, you may need to modify the VIDs in the inner and outer tags. In this case, you can configure a modification policy on the access, trunk, hybrid, or uplink port to modify QinQ tags (Modifying QinQ Tag).

3. QinQ Decapsulation

When a packet with an SP tag is forwarded to PE 2, PE 2 will forward the packet to the dot1q-tunnel port with native VLAN set to VLAN 10. The dot1q-tunnel port then strips the outer tag of the packet and sends the packet to CE-A2. This process is called QinQ decapsulation. The decapsulation mode must be consistent with the encapsulation mode.

Figure 1-3 QinQ Packet Forwarding Process



Note

- The QinQ encapsulation and decapsulation functions are usually used on client-oriented interfaces to implement classification, package management, and transmission of client data.
- QinQ termination means that, after receiving packets, the device identifies and strips the dual VLAN tags, and then sends the packets to other service modules for processing. The QinQ termination function is usually used on a server or network core device to terminate client VLAN tags and realize the functions such as client location or service identification. QinQ termination is not QinQ decapsulation. QinQ decapsulation strips only the outer tag, while QinQ termination strips the dual tags.

1.1.5 Basic QinQ Encapsulation

In the QinQ scenario, the PE edge port connected to the client network will first encapsulate the outer tag for the user packets, and then transmit the packets in the ISP network. The edge port must be configured as a dot1q tunnel port. One dot1q tunnel port can have one native VLAN and several allowed VLANs. Allowed VLANs are divided into Tagged VLANs and Untagged VLANs. The dot1q tunnel port can only be connected to trunk links.

The basic QinQ encapsulation mode is based on the process that the interface encapsulates the packets with tags. The VLAN ID in the tag is the native VLAN of dot1q-tunnel port. Therefore, the VLAN of the ISP network must be configured as the native VLAN of the dot1q tunnel port. To ensure the communication of upstream and downstream packets, the native VLAN must be added to the allowed untagged VLAN list of the port.

After receiving packets, the dot1q-tunnel port regards all the packets as untagged regardless of whether they carry tags, determines whether the packets belong to the native VLAN of the local interface, learns the corresponding relationship between the client-end MAC address and the native VLAN, and stores it in the MAC address table. After that, the interface adds a tag with VID set to the native VLAN to the packets and then forwards the packets.

If the original packet does not carry a VLAN tag, the packet will have only one tag with native VLAN ID after undergoing QinQ encapsulation. If the original packet carries a VLAN tag, the packet will carry dual tags after undergoing QinQ encapsulation and be transmitted in the SP network.

By default, the native VLAN of dot1q-tunnel port is VLAN 1 and the only one allowed VLAN is untagged VLAN 1. "Untagged" in the QinQ function means that the outer tag of QinQ is not carried, because the dot1q-tunnel port must strip the outer tag before sending the packet back to the client network.

Weaknesses: The basic QinQ encapsulation cannot distinguish the types of client packets received by the interface or select the VLAN used for packet encapsulation, and therefore it is not flexible enough.

1.1.6 Selective QinQ Encapsulation

Selective QinQ encapsulation flexibly adds different outer VLAN tags to data flows according to rules.

Selective QinQ implements two encapsulation modes:

- Add an outer VLAN tag based on the inner client VLAN tag.

  The selective QinQ based on C-Tag can distinguish different types of data flows according to the client VLAN tag, and encapsulate different outer tags to implement transparent transmission.

- Add an outer VLAN tag based on the packet content ACL.

  The selective QinQ based on ACL can distinguish the MAC address, IP protocol, source address, destination address, and priority of service flows, or the interface number of applications by using ACL policies, and encapsulate different service flows with different VLAN tags to implement transparent transmission. In this case, the VLAN ID in the VLAN tag is not necessarily the native VLAN of the interface.

When using this function, in addition to configuring the selective QinQ encapsulation policy, you also need to add the outer VLAN to the list of allowed VLANs of the dot1q-tunnel port.

Weaknesses: After the dot1q-tunnel port receives a packet, the learned MAC address entry is the corresponding relationship between the native VLAN and the client MAC address; the packet returned by the peer end belongs to the outer VLAN, and the corresponding entries of the outer VLAN and the client-end MAC address cannot be found in the MAC address table, which will cause the device to perform flooding so as to find the egress of packets.

To solve this problem, use the function in MAC Address Replication to replicate the MAC address of the native VLAN to the outer VLAN.

1.1.7 Modifying QinQ Tag

In the process of QinQ packet transmission, if the outer or inner tags of data packets in the SP network need to be modified due to the actual network topology requirement, the QinQ tag modifying function needs to be enabled. The device can distinguish the packets of different clients, different services, and different priorities based on policies, and modify the VLAN IDs for the inner or outer tags of packets.

1.1.8 VLAN-Mapping

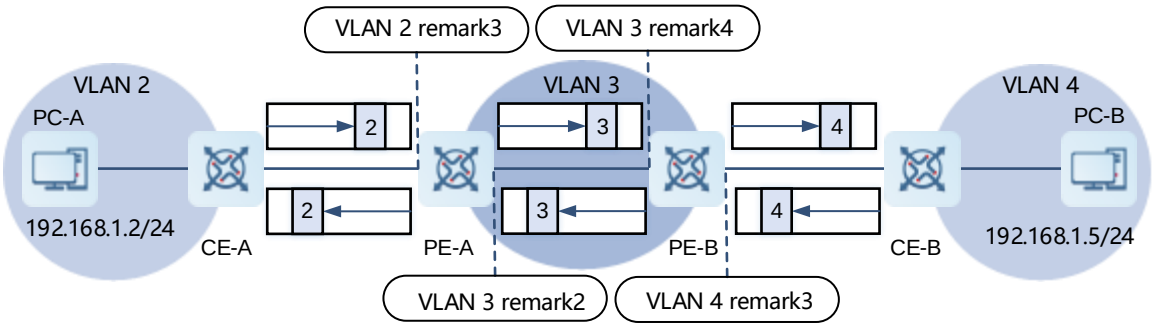
The VLAN mapping function can convert client VLAN IDs and SP VLAN IDs to implement inter-VLAN layer-2 communication. This function takes effect on the access, trunk, hybrid, or uplink ports. VLAN mapping is a function of modifying tags, not adding tags.

As shown in Figure 1-4, when receiving a client packet of PC-A in VLAN 2, PE-A replaces private VLAN 2 in the packet with public VLAN 3 for transmission according to the configured mapping relationship VLAN 2 remark 3; when receiving a packet from public VLAN 3, PE-B maps public VLAN 3 to client VLAN 4 according to the mapping relationship VLAN 3 remark 4 so that PC-B can receive the packet. When receiving a client packet of PC-B in VLAN 4, PE-B restores private VLAN 4 in the packet to public VLAN 3 for transmission according to the mapping relationship VLAN 4 remark 3; when receiving a packet from public VLAN 3, PE-A restores public VLAN 3 to client VLAN 2 according to the mapping relationship VLAN 3 remark 2 so that PC-A can receive the response packet. In this way, inter-VLAN layer-2 communication is achieved between PC-A and PC-B.

If the planning of two client networks is asymmetric or the VLAN IDs of the new and old networks conflict with each other, the VLAN mapping function can be used to achieve inter-VLAN layer-2 communication. Two prerequisites are required for this communication mode:

- The clients at two ends are in the same network segment. If the receiving and sending end clients are not in the same network segment, they will rely on layer-3 communication, and VLAN mapping makes no sense.
- The packet received by PE needs to carry a tag. Therefore, the uplink port of CE needs to send the packet carrying a tag.

Figure 1-4 Introduction to the VLAN Mapping Function



1.1.9 MAC Address Replication

- Source

After receiving a packet, the interface will learn the corresponding relationship between the source MAC address of the packet and the VLAN and write it to the MAC address table. If receiving a tagged packet, the access, trunk, hybrid, and uplink ports will learn the corresponding relationship between the VLAN ID and MAC address in the tag; if receiving an untagged packet, they will learn the corresponding relationship between the native VLAN and MAC address. After receiving packets, the dot1q-tunnel port regards all the packets as untagged regardless of whether they carry tags, and determines whether the packets belong to the native VLAN of the local interface. Therefore, the dot1q-tunnel port always learns the corresponding relationship between its native VLAN and client MAC address.

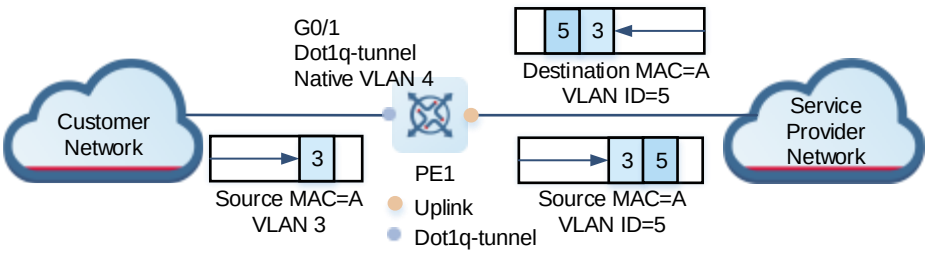
If basic QinQ encapsulation is used, the VLAN ID in the outer tag is native VLAN ID. When receiving a response packet from the peer end, the device can find the corresponding relationship between the client MAC address and the outer VLAN in the MAC address table. If selective QinQ encapsulation is used, the VLAN ID in the outer tag is not the native VLAN ID. When receiving a response packet from the peer end, the device cannot find the matched MAC address entry, and therefore it will perform packet flooding, occupying network resources and affecting the performance.

If the MAC address replication function is enabled, when learning a dynamic address, the interface can replicate the MAC address in the native VLAN to the VLAN where the outer tag is located, thus solving the problem of continuous flooding of the packets returned from the public network. Similarly, the MAC address in the VLAN where the outer tag is can also be replicated to the native VLAN to avoid continuous flooding of the packets from the client network to the public network.

- Principles

As shown in Figure 1-5, the interface connecting PE 1 and the client network is dot1q-tunnel, and native VLAN 4 is configured on this interface. Configure an ACL-based QinQ encapsulation policy, and encapsulate the packet with source MAC address A by using the outer tag VLAN 5. When receiving a packet of client network VLAN 3 and source MAC address A from the interface Gigabit Ethernet 0/1, PE 1 adds the outer tag of VLAN 5. At the same time, because the native VLAN of the receiving interface is VLAN 4, PE1 learns the entry <VLAN 4, MAC-A, GigabitEthernet 0/1>. When a response packet is returned, because the VLAN ID of its outer tag is 5, PE 1 will look for MAC-A in VLAN 5, but MAC-A is not learned into VLAN 5, and therefore this packet will be flooded. If MAC address replication is configured and MAC-A in VLAN 4 is replicated to VLAN 5, the dynamic MAC address entry <VLAN 5, MAC-A, GigabitEthernet 0/1> will be added to PE 1. When a response packet is returned, PE 1 can find that the packet should be forwarded to GigabitEthernet 0/1, thus avoiding flooding.

Figure 1-5 MAC Address Replication



1.1.10 Layer-2 Protocol Tunneling

- Source

Spanning Tree Protocol (STP) is an application that eliminates layer-2 loops by blocking redundant links on the network. It also provides the link backup function. It will transmit a special bridge protocol data unit (BPDU)

packet on the network. After recognizing such a packet, the device will calculate the spanning tree according to the packet information, which may lead to modification of the network topology.

The generic attribute registration protocol VLAN registration protocol (GVRP) is an application used to dynamically configure and diffuse VLAN memberships. It will transmit special VLAN registration and deregistration packets on the network. After recognizing a GVRP packet, the device will modify the VLAN relationship according to the packet information, which will also affect the network topology.

To unify the internal topology of the client network when the client networks are interconnected across the SP network, run STP and GVRP on the client network. However, such a special layer-2 protocol packet will be identified as a special packet by the SP device and participate in the SP topology calculation, affecting the SP network topology.

Therefore, the STP and GVRP packets of the client network must be transmitted through a tunnel in the SP network, so that the SP can forward the client's special packets as ordinary packets. The layer-2 packet tunnel function is needed in this case.

- Principles

When a layer-2 protocol packet of the client network enters a PE, the destination MAC address is changed to a private address for forwarding in the SP network. When the packet arrives at an edge device of the other end, the destination MAC address is changed to a public address before the packet is returned to the client network of the other end. This implements layer-2 protocol packet transmission through a tunnel in the SP network.

1.1.11 Protocols and Standards

- IEEE 802.1ad: IEEE Standard for Local and Metropolitan Area Networks---Virtual Bridged Local Area Networks---Amendment 4: Provider Bridges

1.2 Restrictions and Guidelines

- The QinQ feature only takes effect on a Layer 2 interface. A Layer 3 interface cannot be directly configured as a dot1q tunnel port, and must be configured as a Layer 2 interface first.
- You can use the **display interface switch-port** command to view the interface type. If the Switchport field corresponding to the interface is "enabled", it means that the interface is a Layer 2 interface. If the Switchport field corresponding to the interface is "disabled", it means that the interface is a Layer 3 interface.
- You can use the **switch-port** command to configure the interface as a Layer 2 interface. First, use the **unset switch-port mode** command to restore the mode of the Layer 2 interface to the default value, and then use the **unset switch-port** command to configure the Layer 2 interface as a Layer 3 interface. When the Layer 2/Layer 3 mode of the interface is modified, all properties of the interface are restored to the default values of the interface type.
- The IEEE 802.1x function cannot be enabled on an interface configured as a dot1q-tunnel port.
- The port security function cannot be enabled on an interface configured as a dot1q-tunnel port.
- When a dot1q-tunnel port is configured as the source port of the remote switched port analyzer (RSPAN), the packets whose outer tags contain VIDs consistent with the RSPAN VLAN IDs are monitored as a data flow.
- If you want to match the ACL applied to the dot1q-tunnel port with the VID in the client tag, use the **inner** keyword. See *Configuring ACL* in *ACL and QoS Configuration Guide*.

- After a switch port is enabled with QinQ, you must enable SVGL sharing before IGMP snooping. Otherwise, IGMP snooping will not work on the QinQ-enabled port. See *Configuring IGMP Snooping*.
- You are advised to configure the egress on the client network that is connected to the SP network as an uplink port. The SP TPID value for the uplink port of the client network must be the same as that for the downlink interface of the SP network.
- The default MTU value of an interface is 1,500 bytes. After added with an outer VLAN tag, a packet is four bytes longer. You are advised to increase the MTU value of each interface in the SP networks to at least 1,504 bytes.

1.3 Configuration Task Summary

1.3.1 Configuring Tag Adding

QinQ tag adding includes the following tasks:

(1) Configure the QinQ encapsulation mode of the dot1q-tunnel port. Select at least one of them to configure.

- Configuring Basic QinQ Encapsulation
- Configuring C-Tag-Based Selective QinQ Encapsulation
- Configuring ACL-Based Selective QinQ Encapsulation

(2) (Optional) Configuring MAC Address Replication

(3) (Optional) Configuring Priority Replication or Priority Mapping. The following configuration tasks are mutually exclusive.

- Configuring Priority Replication
- Configuring Priority Mapping

(4) (Optional) Configuring TPID

1.3.2 Configuring Tag Modification

Configure the tag modification mode. The following tasks are optional.

- Configuring QinQ Tag Modification Policy.
- Configuring VLAN-Mapping

1.3.3 Configuring Layer-2 Protocol Tunneling

Configuring Layer-2 Protocol Tunneling includes the following tasks:

- Configuring Layer-2 Protocol Tunneling. The following configuration tasks are optional.
  - Configuring STP Packet Tunneling
  - Configuring GVRP Packet Tunneling

1.4 Notes

This section describes precautions for configuring the QinQ function.

Whether the QinQ function is supported depends on devices.

Table 1-1 Products Supporting the QinQ Function

Function	Product Not Supporting the Function
Configuring ACL-Based Selective QinQ Encapsulation	S6980-64QC S6920-4C S6580-48CQ8QC S6580-24DC8QC
Configuring QinQ Tag Modification Policy	S6980-64QC S6920-4C S6580-48CQ8QC S6580-24DC8QC

1.5 Configuring Basic QinQ Encapsulation

1.5.1 Overview

Implement layer-2 VPN based on a port-based QinQ policy.

1.5.2 Restrictions and Guidelines

- Configure a VLAN of the SP network as the native VLAN of the dot1q-tunnel port. Because the trunk port will strip the tag of the packet whose VLAN ID is its native VLAN, do not configure the native VLAN of the trunk port in the SP network as the native VLAN of the dot1q-tunnel port.
- After the native VLAN of the dot1q-tunnel port is configured, the data of the VLAN cannot pass through the interface. The native VLAN must be added to the list of untagged VLANs allowed by the dot1q-tunnel port. After the configuration, the dot1q-tunnel port can output the packets of native VLAN without tags.
- If the native VLAN is added to the list of allowed tagged VLANs, packets of the native VLAN output from the dot1q-tunnel port will carry the tag of the SP VLAN, while the client device cannot identify the VLAN ID of SP network. Therefore, in order to ensure normal communication of the uplink and downlink packets in the application of QinQ, the SP VLAN (native VLAN) must be added to the allowed VLAN list of the interface in the untagged form.
- If port-based QinQ is enabled, you do not need to add the VLAN of the client network to the allowed VLAN list of the dot1q-tunnel port. If selective QinQ is enabled, add the VLAN of the client network to the allowed VLAN list of the interface in the tagged or untagged form based on the actual conditions.

1.5.3 Prerequisites

- Configure the interface as a Layer 2 interface.

1.5.4 Procedure

(1) Create an SP VLAN.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create an SP VLAN.

vlan *svid*

Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

(2) Enter the interface configuration mode, and configure the interface for connecting to the client network.

a Enter the Layer 2 Ethernet interface configuration mode.

interface { *10thernet-type interface-number* | **range** *10thernet-type interface-range* }

o Enter the Layer 2 aggregate interface configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(3) Configure the Layer 2 interface as a dot1q-tunnel port.

switch-port mode dot1q-tunnel

An interface works in access mode by default.

(4) Configure to add the SP VLAN to the untagged VLAN list of the dot1q-tunnel port.

switch-port dot1q-tunnel allow-pass vlan { [**add**] **untagged** *svid* | **remove** *other-vlan-id* }

By default, the allowed VLAN of dot1q-tunnel port is untagged VLAN 1.

(5) Configure the SP VLAN as the native VLAN of dot1q-tunnel port so that the client network data is transmitted through a tunnel on the SP VLAN.

switch-port dot1q-tunnel pvid vlan *svid*

The native VLAN of the dot1q-tunnel port is VLAN 1 by default.

1.6 Configuring C-Tag-Based Selective QinQ Encapsulation

1.6.1 Overview

Encapsulate outer VLAN tags (S-Tags) in packets based on inner tags to ensure preferential transmission and management of layer-2 VPN and service flows.

1.6.2 Restrictions and Guidelines

- This function must be configured based on the basic QinQ, and prevails over the basic QinQ policy.
- In the network environment, for packets received on the dot1q-tunnel port, the VID of the outer tag needs to be modified according to the VID of the inner tag. With this command, you can add an outer VID the same as the inner VID to the packet so as to output the original VID tag packet from the egress.
- To continue to adopt the VLAN tag priority specified by the client network, you can configure priority replication to configure an outer tag the same as the inner tag.
- If the SP network requires the transmission of packets based on the priority of the outer tag, you also need to configure priority replication to set the CoS of the outer tag to the specified value.

Caution

- The priority of the VID change policies in a descending order: ACL-based > C-Tag-based > port-based.
- When a member is added to/deleted from a Layer 2 aggregate interface, the policy of adding or modifying the VID configured on the aggregate interface is deleted and needs to be reconfigured. You are advised to configure the VIP policy on an aggregate interface after configuring the members.
- You must configure the dot1q-tunnel port and the interface connected to the public network to permit packets with specified VLAN IDs (including the native VLAN ID) in the outer tag to pass through.

1.6.3 Prerequisites

Configure the interface as a Layer 2 interface.

1.6.4 Procedure

(1) Create an SP VLAN.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create a group of SP VLANs.

vlan range *svidA,svidB*

Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

(2) Enter the interface configuration mode to configure the interface for connecting the client network.

a Enter the Layer 2 Ethernet interface configuration mode.

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

o Enter the Layer 2 aggregate interface configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(3) Configure the Layer 2 Ethernet interface as a dot1q-tunnel port.

switch-port mode dot1q-tunnel

A Layer 2 interface works in access mode by default.

(4) Configure to add the SP VLAN to the untagged VLAN list of the dot1q-tunnel port.

switch-port dot1q-tunnel allow-pass vlan { [**add**] **untagged** *svidA,svidB* | **remove** *other-vlan-id* }

By default, the allowed VLAN of dot1q-tunnel port is untagged VLAN 1.

(5) Configure the C-Tag-based QinQ encapsulation rule. The packets of different client VLANs are allocated to different SP VLANs for transmission through a tunnel.

a The packets of client VLAN *cvid-lista* are allocated to the SP VLAN *svidA*.

dot1q outer-vid *svidA* **register inner-vid** *cvid-lista*

No policy of adding an outer VID based on the inner VID of packets is configured by default.

b The packets of client VLAN *cvid-listb* are allocated to the SP VLAN *svidB*.

dot1q outer-vid *svidB* **register inner-vid** *cvid-listb*

No policy of adding the outer VID based on the inner VID of packets is configured by default.

1.7 Configuring ACL-Based Selective QinQ Encapsulation

Note

This function is not supported by the S6980-64QC, S6920-4C, S6580-48CQ8QC, or S6580-24DC8QC.

1.7.1 Overview

Encapsulate outer VLAN tags (S-Tags) in packets based on the ACL-based flow classification to allow the SP network to manage different services.

For input packets of dot1q-tunnel port on the network, if the outer tags with different VLAN IDs need to be added according to different packet contents, this function must be configured.

1.7.2 Restrictions and Guidelines

- To continue to adopt the VLAN tag priority specified by the client network, you can configure priority replication to configure an outer tag the same as the inner tag.
- If the SP network requires the transmission of packets based on the priority of the outer tag, you also need to configure priority replication to set the CoS of the outer tag to the specified value.
- The **expert access-control-list extended** command is used to configure the extended expert ACL and enter the extended expert ACL configuration mode. The command format is as follows: **expert access-control-list extended** { *acl-name* | *acl-number* }. *acl-name*: Extended expert ACL name. The value is a case-sensitive string of 1 to 99 characters. *acl-number*: Extended expert ACL number. The value range is from 2700 to 2899. For the sake of easy description, the **expert access-control-list extended** *acl-number* command is taken as an example in the configuration procedure.

  The **permit** command is used to add an ACL rule of the allowed type to an ACL. For the detailed syntax, see *Configuring ACL*. For the sake of easy description, the **permit** *ethernet-type any any* command is taken as an example in the configuration procedure.

Caution

- The policy of adding the VID based on ACL prevails over those based on C-Tag and port.
- When an ACL is deleted, the related policy will be automatically deleted.
- Upon receiving a packet with two or more tags, the dot1q-tunnel port cannot add an outer tag to the packet based on the flow-based matching rule.
- If a packet matches two or more data flow-based VID adding policies without priority, only one policy is executed. You are advised to specify the priority when configuring a policy.
- You must configure the dot1q-tunnel port and the interface connected to the public network to permit packets with specified VLAN IDs (including the native VLAN ID) in the outer tag to pass through.

1.7.3 Prerequisites

Configure the interface as a Layer 2 interface.

1.7.4 Procedure

(1) Create a traffic classification and define a packet matching rule.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create a professional extended ACL *acl-number-a* and enter the professional extended ACL configuration mode.

expert access-control-list extended *acl-number-a*

By default, there is no professional extended ACL rule in the ACL.

d Add an ACL rule of the allowed type to ACL *acl-number-a*.

permit *ethernet-type-a any any*

By default, there is no ACL rule of the allowed type in the ACL.

e Return to the global configuration mode.

exit

f Create a professional extended ACL *acl-number-b* and enter the professional extended ACL configuration mode.

expert access-control-list extended *acl-number-b*

By default, there is no professional extended ACL rule in the ACL.

g Add an ACL rule of the allowed type to ACL *acl-number-b*.

permit *ethernet-type-b any any*

By default, there is no ACL rule of the allowed type in the ACL.

h Return to the global configuration mode.

exit

(2) Create an SP VLAN.

a Create a group of SP VLANs.

vlan range *svidA,svidB*

Only VLAN 1 exists by default.

b Return to the global configuration mode.

exit

(3) Enter the interface configuration mode to configure the interface for connecting the client network.

a Enter the Layer 2 Ethernet interface configuration mode.

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

o Enter the Layer 2 aggregate interface configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure the Layer 2 Ethernet interface as a dot1q-tunnel port.

switch-port mode dot1q-tunnel

A Layer 2 interface works in access mode by default.

(5) Configure to add the SP VLAN to the untagged VLAN list of the dot1q-tunnel port.

switch-port dot1q-tunnel allow-pass vlan { [**add**] **untagged** *svidA,svidB* | **remove** *other-vlan-id* }

By default, the allowed VLAN of dot1q-tunnel port is untagged VLAN 1.

(6) Configure the ACL-based QinQ encapsulation rule. The packets matching different ACL rules are allocated to different SP VLANs for transmission through a tunnel.

a The packets matching ACL *acl-number-a* are allocated to the SP VLAN *svidA*.

traffic-redirect access-apply *acl-number-a* **nested-vlan** *svidA* **in**

The policy of adding an outer VLAN ID based on ACL is not configured by default.

b The packets matching ACL *acl-number-b* are allocated to the SP VLAN *svidB*.

traffic-redirect access-apply *acl-number-b* **nested-vlan** *svidB* **in**

The policy of adding an outer VLAN ID based on ACL is not configured by default.

1.8 Configuring MAC Address Replication

1.8.1 Overview

If selective QinQ encapsulation is used, the VLAN ID in the outer tag is not the native VLAN ID. When receiving a response packet from the peer end, the device cannot find the matched MAC address entry, and therefore it will perform packet flooding, occupying network resources and affecting the performance.

If the MAC address replication function is enabled, when learning a dynamic address, the dot1q-tunnel port can replicate the MAC address in the native VLAN to the VLAN where the outer tag is located, thus solving the problem of continuous flooding of the packets returned from the public network. Similarly, the MAC address in the VLAN where the outer tag is can also be replicated to the native VLAN to avoid continuous flooding of the packets from the client network to the public network.

1.8.2 Restrictions and Guidelines

- The MAC address replication function cannot be used together with the share VLAN.
- The MAC address replication function cannot be configured on the host and promiscuous interfaces, or the destination interface with port security and imaging enabled, or the interface with the 802.1x function enabled.
- After MAC address replication is disabled, the system will delete all the learned MAC address entries from the destination VLAN. The MAC address entries obtained through MAC address replication between VLANs cannot be deleted manually. When you need to delete these entries, disable MAC address replication.
- *mapping-id* indicates the index number of the MAC address replication policy. The value range is from 1 to 8, and therefore each port can be configured with up to eight destination VLANs.
- *source-vlan-list* indicates the source VLAN list. The value range is from 1 to 4094. One single command can contain one or more VLANs.
- *destination-vlan-id* indicates the destination VLAN. The value range is from 1 to 4094. Only one destination VLAN can be specified in one command. MAC address replication between VLANs can be configured on an interface for one single destination VLAN once only. To make modification, you need to delete the previous configuration before reconfiguration. MAC address replication takes effect even if the interface does not

belong to the specified destination VLAN. MAC address replication does not support address replication to dynamic VLANs.

- MAC address replication replicates only dynamic addresses. Address replication is disabled when the address table is full. If source addresses already exist before replication is enabled, corresponding MAC addresses will not be replicated. Replicated addresses have a higher priority than dynamic addresses but have a lower priority than other types of addresses.
- When a MAC address ages, the replicated MAC address will also age. When the MAC address is deleted, the replicated address will be deleted automatically.
- This function does not support hot backup. After primary/secondary switchover occurs, you are advised to disable MAC address replication and then enable it again.

1.8.3 Prerequisites

Configure the interface as a Layer 2 interface.

1.8.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 Ethernet interface configuration mode.

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

(4) Configure the MAC address replication policy, and enable the MAC address replication function.

mac-address-mapping *mapping-id* **source-vlan** *source-vlan-list* **destination-vlan** *destination-vlan-id*

The MAC address replication function is disabled by default.

1.9 Configuring Priority Replication or Priority Mapping

1.9.1 Overview

This function configures the CoS value of the outer tag based on the user priority value (CoS) of the client VLAN tag by replication or mapping.

- Priority replication

By default, the user priority (PRI) field of the Ethernet frame tag is 0, namely, the normal flow. To make the specified packet preferentially processed and transmitted, configure this field to change the priority of the packet. The value of this field corresponds to the CoS value in the QoS policy. The CoS-based QoS policy can be configured to ensure the processing priority of a service.

If the VLAN tag of a client packet is configured with the CoS value of the user priority, and the QoS priority policy is configured in the SP network for the CoS value, the CoS value for the VLAN tag of client packet can be replicated as the CoS value of the outer tag, so that the client packet can be transmitted transparently and the QoS priority policy provided by the SP network can be used.

- Priority mapping

When an outer VLAN tag is added, the user priority of the outer VLAN tag can be set according to the user priority of the inner VLAN tag through priority mapping so that the network QoS priority policy can be used after the packet is encapsulated with an outer VLAN tag.

The SP network sets different QoS policies for multiple CoS values, corresponding to different service flow services. To make some client services be preferentially transmitted in the SP network, the priority mapping function can be configured. The CoS value of the outer VLAN tag can be specified according to the CoS value of the client inner VLAN tag, so that the packet can enjoy the QoS policy of the SP network after being encapsulated with the outer VLAN tag.

1.9.2 Restrictions and Guidelines

- Priority replication and priority mapping cannot be enabled on the same interface at the same time.
- Priority replication or priority mapping of client tags can be configured only on the dot1q-tunnel port.
- The order of priority effectiveness from high to low is: priority mapping, QoS based on ACL data flow, priority replication, and QoS based in trust mode.
- If the interface trust mode is not configured (i.e. interface Trust None), the priority mapping will not take effect. If the configured trust mode does not match the configured mapping rule, priority mapping will also not take effect.

1.9.3 Configuration Tasks

The following configuration tasks are mutually exclusive. Select one task.

- Configuring Priority Replication
- Configuring Priority Mapping

1.9.4 Configuring Priority Replication

1. Overview

It is assumed that the SP network is configured with the QoS policy based on the user priority value (CoS) of client VLAN tag. If the CoS value of an inner tag is 3, the CoS value of the outer tag can be set to 3 by configuring priority replication, so that the packet can enjoy the QoS policy with a priority of 3 as the inner tag after being encapsulated with the outer VLAN tag.

Configure priority replication to apply the QoS policy provided by the SP network for the client VLAN tag.

2. Prerequisites

- Configure the interface as a Layer 2 interface.
- Configure the Layer 2 interface as a dot1q tunnel port.
- Configure the QinQ encapsulation mode of the dot1q tunnel port.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode
- **interface** { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }
- Enter the Layer 2 aggregate interface configuration mode.
- **interface** { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure the CoS priority for trusted packets of the interface.

mls qos trust cos

The default trust mode of an interface is untrust.

(5) Configure priority replication.

inner-priority-trust enable

The priority replication function is disabled by default.

1.9.5 Configuring Priority Mapping

1. Overview

It is assumed that the SP network is set with the QoS policy based on the user priority value (CoS) of client VLAN tag. Suppose the CoS value of an inner tag is 3. The CoS value of the outer tag can be set to 5 by configuring priority mapping (CoS-to-CoS), so that the packet can enjoy the QoS policy with a priority of 5 after being encapsulated with the outer VLAN tag.

If the user priority of the outer VLAN tag needs to be set according to the VLAN tag of the client to flexibly apply the QoS priority, configure the priority mapping policy.

2. Restrictions and Guidelines

- The value range of *inner-cos-value* and *outer-cos-value* are from 0 to 7.
- If priority mapping is not configured, the mapping rules shown in Table 1-1 are used by default.

Table 1-2 Default Priority Mapping

inner-cos-value	0	1	2	3	4	5	6	7
outer-cos-value	0	1	2	3	4	5	6	7

3. Prerequisites

- Configure the interface as a Layer 2 interface.
- Configure the Layer 2 interface as a dot1q tunnel port.
- Configure the QinQ encapsulation mode of the dot1q tunnel port.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure the CoS priority for trusted packets of the interface.

mls qos trust cos

The default trust mode of an interface is untrust.

(5) (Optional) Disable priority replication.

unset inner-priority-trust enable

The priority replication function is disabled by default.

(6) Configure priority mapping.

dot1q-tunnel cos *inner-cos-value* **remark-cos** *outer-cos-value*

No priority mapping rule is set by default.

1.10 Configuring TPID

1.10.1 Overview

An Ethernet frame tag consists of four fields: Tag Protocol Identifier (**TPID**, written as **Type** in the frame format map), User Priority (**PRI**), **CFI**, and **VLAN ID**.

The TPID value is 0x8100 by default, while the devices of some vendors set the TPID value for the outer tag of packets to 0x9100 or other values. To be compatible with these devices, this device supports port-based TPID modification for packets.

You can configure the TPID value of ports independently. When forwarding packets, these ports replace the TPID in the outer VLAN tag of the packets with the value set by the user, ensuring compatibility with the TPIDs of different vendors.

Purpose: The interface on the PE that connects to the public network needs to be configured as an uplink port. Configure the TPIDs in the tags on SP network devices to realize TPID compatibility with third-party devices. Perform the configuration in interface configuration mode of the uplink port.

1.10.2 Restrictions and Guidelines

- If a PE is connected to a third-party device on which the TPID is not the default value 0x8100 of IEEE 802.1Q, you need to configure the TPID on the interface connected to the third-party device.

⚠ Caution

Do not set the TPID to the following well-known types: 0x0806 (ARP), 0x0200 (PUP), 0x8035 (RARP), 0x0800 (IP), 0x86DD (IPv6), 0x8863/0x8864 (PPPoE), 0x8847/0x8848 (MPLS), 0x8137 (IPX/SPX), 0x8000 (IS-IS), 0x8809 (LACP), 0x888E (802.1x), 0x88A7 (cluster), and 0x0789 (reserved).

- TPIDs can be configured in interface configuration mode and global configuration mode.

1.10.3 Prerequisites

- Configure the interface as a Layer 2 interface.

- Configure the interface as an Uplink port.

1.10.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode, and configure the uplink port (the port where the SP's network edge device connects to the public network).

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure the TPID value.

frame-tag tpid *tpid*

The TPID value is 0x8100 by default.

1.11 Configuring QinQ Tag Modification Policy

Note

This function is not supported by the S6980-64QC, S6920-4C, S6580-48CQ8QC, or S6580-24DC8QC.

1.11.1 Overview

Packets encapsulated by QinQ carries double tags, and must be transmitted in the SP network. The Access, Trunk, Hybrid and Uplink ports of the device can receive and forward packets carrying double tags, while the dot1q tunnel port will discard packets carrying double tags.

In the SP network, the outer tag or inner tag of data packets of the SP network may need to be modified due to the actual networking topology requirements. In this case, the QinQ tag modifying function needs to be used. The device supports the following modification policies:

- Modify the outer VID based on the outer VID of packet

  For inbound packets on access, trunk, hybrid and uplink ports, this function can modify the VID of the outer tag to a different VID value according to the different VID in the inner tag of the packet.

- Modify the outer VID based on the inner VID of packet

  For inbound packets on access, trunk, hybrid and uplink ports, this function can modify the VID of the outer tag according to the inner and outer VID of the packet before the packet is forwarded.

- Modify the outer VID based on the outer and inner VIDs of packet

  For the inbound packets on access, trunk, hybrid and uplink ports, this function can modify the VID of the outer tag according to the VID in the outer tag of the packet before the packet is forwarded.

- Modify the outer VID based on the content of packet

For the outbound packets on access, trunk, hybrid and uplink ports, this function can modify the VID of the inner tag based on the packet content identified by ACL.

- Modify the inner VID based on the content of packet

  For the outbound packets on access, trunk, hybrid and uplink ports, this function can modify the VID of the outer tag based on the packet content identified by ACL.

1.11.2 Restrictions and Guidelines

- The policy of modifying the VID based on ACL prevails over those based on port and C-Tag.
- The function of modifying VID according to packet content requires configuring ACL to distinguish data flow. When an ACL is deleted, the related policy will be automatically deleted.
- The modification policy takes effect only on the access, trunk, hybrid, and uplink ports.
- If a packet matches two or more VID modification policies without priority, only one policy is executed.
- *new-svid* indicates the VID of the outer tag after modification, *old-svid* indicates the VID of the outer tag before modification, *old-svid-list* indicates the VID list of the outer tag before modification, which are VLAN IDs of the SP network.
- *cvid-list* indicates the VID list of the inner tag after modification, *new-cvid* indicates the VID of the inner tag after modification, which are VLAN IDs of the client network.
- *acl-name* indicates the name of the ACL policy for matching data flow. The value is a case-sensitive string of 1-99 characters. *acl-number* indicates the number of the ACL policy for matching data flow. The value range of IP standard ACL is 1-99 and 1300-1999. The value range of IP extended ACL is 100-199 and 2000-2699. The value range of MAC standard ACL is 700-799. The value range of extended expert ACL is 2700-2899. See *Configuring ACL* in *ACL and QoS Configuration Guide*.

1.11.3 Prerequisites

- Configure the interface as a Layer 2 interface.
- Configure the Layer 2 interface as an access, trunk, hybrid or uplink port.

1.11.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure the tag modification policy of double-tagged inbound packets. Please select at least one of them to configure.

(5) Configure the policy of modifying the outer VID based on the inner VID of packet.

dot1q relay-vid *new-svid* **translate inner-vid** *cvid-list*

No policy of modifying the outer VID based on the inner VID of packets is configured by default.

(6) Configure the policy of modifying the outer VID based on the outer VID of the packet.

dot1q relay-vid *new-svid* **translate local-vid** *old-svid-list*

No policy of modifying the outer VID based on the outer VID of input packets is configured by default.

(7) Configure the policy of modifying the outer VID based on the outer and inner VIDs of packet.

dot1q new-outer-vlan *new-svid* **translate old-outer-vlan** *old-svid* **inner-vlan** *cvid-list*

No policy of modifying the outer VID based on the outer and inner VIDs of packet is configured by default.

(8) Configure the tag modification policy of double-tagged outbound packets. Please select at least one of them to configure.

(9) Configure the policy of modifying the inner VID based on the packet content.

traffic-redirect access-apply { *acl-name* | *acl-number* } **inner-vlan** *new-cvid* **out**

No policy of modifying the inner VID based on packet content is configured by default.

(10) Configure the policy of modifying the outer VID based on the packet content.

traffic-redirect access-apply { *acl-name* | *acl-number* } **outer-vlan** *new-svid* **in**

No policy of modifying the outer VID based on the packet content is configured by default.

1.12 Configuring VLAN-Mapping

Note

This function is not supported by the S6920-4C.

1.12.1 Overview

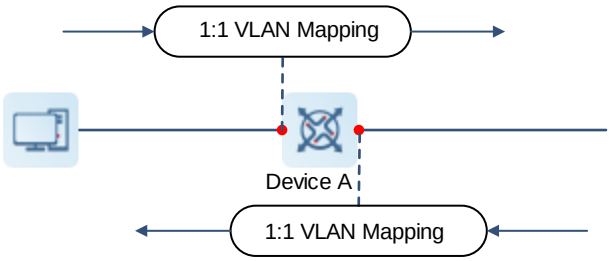
The function is used to replace the VLAN ID of a packet of the client network with a public VLAN ID so that the packet can be transmitted according to the VLAN rules of the public network; or replace the VLAN ID of a packet of the public network with a VLAN ID of the client network so that the packet can be transmitted according to VLAN planning of the client network.

1. 1:1 VLAN Mapping In

On the input interface of a device, the VID of the packet on one VLAN is modified to a VID of another VLAN. This mapping method is called one-to-one VLAN mapping in the input direction. This function is used for floor devices to use different VLANs to carry the same service from different clients and differentiate clients.

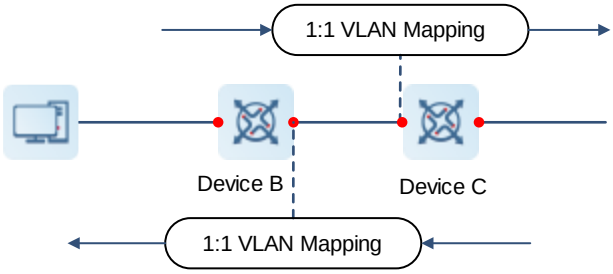
As shown in Figure 1-6, VLAN mapping in the input direction is configured on a single device (Device A). For uplink data flow, the VLAN mapping policy in the input direction is configured on the downlink port, and the original VLAN ID is mapped to a new VLAN ID before packet output from the uplink port. For downlink data flow, the VLAN mapping policy in the input direction is configured on the uplink port, and the packet VLAN ID is mapped to the original VLAN ID before packet output from the downlink port.

Figure 1-6 Configuring 1:1 VLAN Mapping in Input Direction on Single Device



As shown in Figure 1-7, VLAN mapping in the input direction is configured on two devices (Device B and Device C). For uplink data flow, the VLAN mapping policy in the input direction is configured on the downlink port of Device C, and the original VLAN ID is mapped to a new VLAN ID before packet output from the uplink port. For downlink data flow, the VLAN mapping policy in the input direction is configured on the uplink port of Device B, and the packet VLAN ID is mapped to the original VLAN ID before packet output from the downlink port.

Figure 1-7 Configuring 1:1 VLAN Mapping in Input Direction on Two Devices



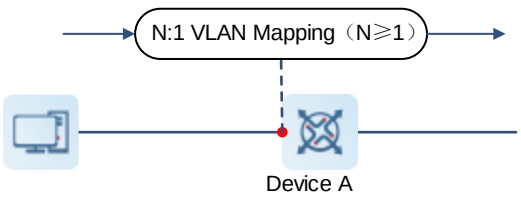
2. N:1 VLAN Mapping In

On the input interface of a device, the VIDs of the packets on multiple VLANs are modified to a VID of another VLAN. This mapping method is called many-to-one VLAN mapping in the input direction. This function is used for campus devices to combine the same services of different users originally carried by multiple VLANs for bearing on one VLAN and save VLAN resources.

Many-to-one VLAN mapping needs to be configured only for uplink data flow, and not for downlink data flow. Because the device first learns the source MAC address, source VLAN ID, and destination VLAN ID through the uplink packets, when receiving downlink packets, the device automatically maps and restores the VLAN ID of the downlink unicast packet according to the source table entry, but cannot restore the VLAN ID for downlink broadcast or multicast packets.

As shown in Figure 1-8, VLAN mapping in the input direction is configured on a single device (Device A).

Figure 1-8 Configuring VLAN Mapping in Input Direction on Single Device



1.12.2 Restrictions and Guidelines

- 1:1 VLAN Mapping In:

  For uplink data flow, the source VLAN is a client VLAN; the destination VLAN is an SP VLAN.

  For downlink data flow, the source VLAN is an SP VLAN; the destination VLAN is a client VLAN.

- N:1 VLAN Mapping In:

  For uplink data flow, the source VLAN list can contain one or more client VLANs; the destination VLAN is an SP VLAN.

  For downlink data flow, the device automatically maps and restores the VLAN IDs of downlink unicast packets according to the source table entries, but cannot restore downlink broadcast or multicast packets.

- When the VLAN mapping function in input direction is enabled, the VLAN ID of the packet sent to the CPU is the VLAN ID after modification. The interface only needs to allow the VLAN after modification to pass. If the VLAN data before modification does not need to be forwarded on the device, you do not need to create a VLAN before modification on the device.
- The VLAN tags of public network packets whose destination MAC address is multicast, broadcast, or unknown unicast address cannot be replaced with the VLAN tag of the client private network.
- VLAN mapping can be configured only on the access, trunk, hybrid, or uplink port. Do not configure VLAN mapping and selective QinQ on the same interface.

1.12.3 Prerequisites

- Configure the interface as a Layer 2 interface.
- Configure the Layer 2 interface as an Access, Trunk, Hybrid or Uplink port.

1.12.4 Procedure

(1) Create a VLAN after mapping.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create a VLAN after mapping.

vlan *destinationdest-vlan-id*Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

(2) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(3) Configure the VLAN mapping policy in the inbound direction on the interface.

vlan-mapping-in vlan *source-vlan-id* **remark** *destination-vlan-id*

No VLAN mapping policy in the inbound direction is configured by default.

Note

This section describes only the procedure of configuring one mapping policy on an interface, without reflecting the mapping logic. The mapping relationship needs to be configured on multiple interfaces to reflect the mapping logic.

1.13 Configuring Layer-2 Protocol Tunneling

1.13.1 Overview

BPDU tunneling is a layer-2 tunneling function based on the Layer 2 Tunnel Protocol (L2TP).

Spanning tree protocol packets use the dedicated multicast address 0180.c200.0000 as the destination MAC address. GVRP protocol packets use the dedicated multicast address 0180.c200.0006 as the destination MAC address. When the dot1q-tunnel port of the SP edge device receives these two types of packets, it will not treat them as ordinary client packets. The edge device will identify them as BPDU packets and will not forward them. The spanning tree calculation is performed according to the contents of the spanning tree BPDU, and VLAN information will be updated according to GVRP protocol packets by adding or deleting VLANs, which may affect the topology of the SP network.

When the STP or GVRP function is also enabled for the SP device, the BPDU tunnel function needs to be used to transmit the STP and GVRP packets of the client network in the SP network without affecting the SP network.

When the protocol packets on the client network enter the dot1q-tunnel of the PE, the destination MAC address is changed to the private address and then the packets are forwarded in the SP network. When the packets reach the PE at the other end, the destination MAC address is changed to the public address and the packets are transmitted on the client network at the other end. This achieves transmission of the protocol packets through the tunnel in the SP network, preventing the client network and the SP network from interfering with each other.

1.13.2 Restrictions and Guidelines

- If STP or GVRP is not enabled for the SP device, you can configure **bridge-frame forwarding protocol bpdu** to enable transparent transmission of BPDU packets.
- To make protocol tunneling effective on an interface, you need to first enable protocol tunneling in global configuration mode. When the protocol tunneling function on an interface takes effect, the interface does not participate in protocol calculation. If the dot1q-tunnel receives a packet whose destination MAC address is a special multicast address, there is an error in the networking and the packet will be dropped directly.

1.13.3 Configuration Tasks

The configuration includes the following tasks:

- Configuring STP Packet Tunneling
- Configuring GVRP Packet Tunneling

1.13.4 Configuring STP Packet Tunneling

1. Overview

If the STP function is enabled for the SP device and STP BPDU packets of the client network need to be transmitted, this function must be configured.

- When the spanning tree protocol packets in the client network enters the dot1q tunnel port of the edge device, its destination MAC address is changed from the public address 0180.c200.0000 to the private address (default is 01d0. f800.0005) before forwarding the packets in the SP network.
- After packets are forwarded to the edge device at the other end, the destination MAC address is changed from the private address (the default is 01d0. f800.0005) to the public address 0180.c200.0000 before and transmitting the packets in the client network at the other end.

2. Restrictions and Guidelines

- The STP tunneling function takes effect only when it is enabled in global configuration mode and interface configuration mode respectively.
- The optional tunnel addresses of STP packets are 01d0.f800.0005 (default value), 011a.a900.0005, 010f.e200.0003, 0100.0ccd.cdd0, 0100.0ccd.cdd1, and 0100.0ccd.cdd2.

3. Prerequisites

- Configure the interface as a Layer 2 interface.
- Configure the Layer 2 interface as a dot1q tunnel port.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the STP tunneling function globally.

l2protocol-tunnel stp

The STP packet tunneling function is disabled by default.

(4) (Optional) Configure a tunnel address in the corresponding protocol.

l2protocol-tunnel stp tunnel-dmac *mac-address*

The STP packet tunnel address is 01d0.f800.0005 by default.

(5) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode

interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(6) Enable STP packet tunnelling function on the dot1q-tunnel port.

l2protocol-tunnel stp enable

The STP packet tunneling function is disabled on the dot1q-tunnel port by default.

1.13.5 Configuring GVRP Packet Tunneling

1. Overview

If the GVRP function is enabled for the SP device and GVRP BPDU packets of the client network need to be transmitted, this function must be configured.

- When the spanning tree protocol packets in the client network enters the dot1q tunnel port of the edge device, its destination MAC address is changed from the public address 0180.c200.0000 to the private address (default is 01d0. f800.0006) before forwarding the packets in the SP network.
- After packets are forwarded to the edge device at the other end, the destination MAC address is changed from the private address (the default is 01d0. f800.0006) to the public address 0180.c200.0000 before and transmitting the packets in the client network at the other end.

2. Restrictions and Guidelines

- The GVRP tunneling function takes effect only when it is enabled in global configuration mode and interface configuration mode respectively.
- The optional tunnel addresses of GVRP packets are 01d0.f800.0006 (default value) and 011a.a900.0006.

3. Prerequisites

- Configure the interface as a Layer 2 interface.
- Configure the Layer 2 interface as a dot1q tunnel port.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the GVRP tunneling function globally.

l2protocol-tunnel gvrp

The GVRP packet tunneling function is disabled by default.

(4) (Optional) Configure a tunnel address in the corresponding protocol.

l2protocol-tunnel gvrp tunnel-dmac *mac-address*

The GVRP packet tunnel address is 01d0.f800.0006 by default.

(5) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode
  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }
- Enter the Layer 2 aggregate interface configuration mode.
  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(6) Enable GVRP packet tunneling function on dot1q-tunnel port.
l2protocol-tunnel gvrp enable

The GVRP packet tunneling function is disabled for the dot1q-tunnel port by default.

1.14 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

This section also describes the **debugging** command used for outputting debugging information.

Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging function immediately after use.

You can run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-3 Monitoring

Command	Purpose
display dot1q-tunnel [**interfaces** *interface-type interface-number*]	Displays whether the dot1q-tunnel function is enabled on the interface.
display interfaces dot1q-tunnel	Displays configuration of the dot1q-tunnel port.
display registration-table [**interfaces** *interface-type interface-number*]	Displays the policy of adding the VID based on the protocol on the dot1q-tunnel port.
display translation-table [**interfaces** *interface-type interface-number*]	Displays the policy of modifying the VID based on the access, trunk, and hybrid ports.
display interfaces [*interface-type interface-number*] **vlan-mapping**	Displays the VLAN mapping policy on the interface.
display traffic-redirect [**interfaces** *interface-type interface-number*]	Displays the policy of modifying or adding the VID based on data flow.
display frame-tag tpid interfaces [*interface-type interface-number*]	Displays the TPID configuration of the interface.
display inner-priority-trust [**interfaces** *interface-type interface-number*]	Displays the configuration of priority replication.
display interfaces [*interface-type interface-number*] **remark**	Displays the configuration of priority mapping.
display interfaces mac-address-mapping	Displays the configuration of MAC address replication.
display l2protocol-tunnel { **gvrp** | **stp** }	Displays the layer-2 tunnel configuration.
debugging bridge qinq	Enables the QinQ debugging function.

1.15 Configuration Examples

1.15.1 Configuring Basic QinQ to Implement Layer-2 VPN

1. Requirements

An SP provides a VPN for Company A and Company B, as shown in Figure 1-9. Basic QinQ is enabled on PEs to meet the layer-2 VPN requirement. On the public network, Company A and Company B belong to different VLANs, and client data packets are transmitted on different native VLANs to implement the simple layer-2 VPN.

- Company A uses public VLAN 10. In Area 1, clients are connected through CE-A1 and access the public network through PE 1; in Area 2, clients are connected through CE-A2, and access the public network through PE 2. The dot1q-tunnel port first encapsulates the client data packets of Company B with one VLAN 10 tag, and then forwards them on the public network.
- Company B uses public VLAN 20. In Area 1, clients are connected through CE-B1 and access the public network through PE 1; in Area 2, clients are connected through CE-B2, and access the public network through PE 2. The dot1q-tunnel first encapsulates the client data packets of Company B with one VLAN 20 tag, and then forwards them on the public network.

The client VLANs in Company A and Company B are transparent to the public network. The client VLANs can be reused without conflicts. Therefore, client VLANs do not need to be configured on the SP device, but need to be configured only on the CE device.

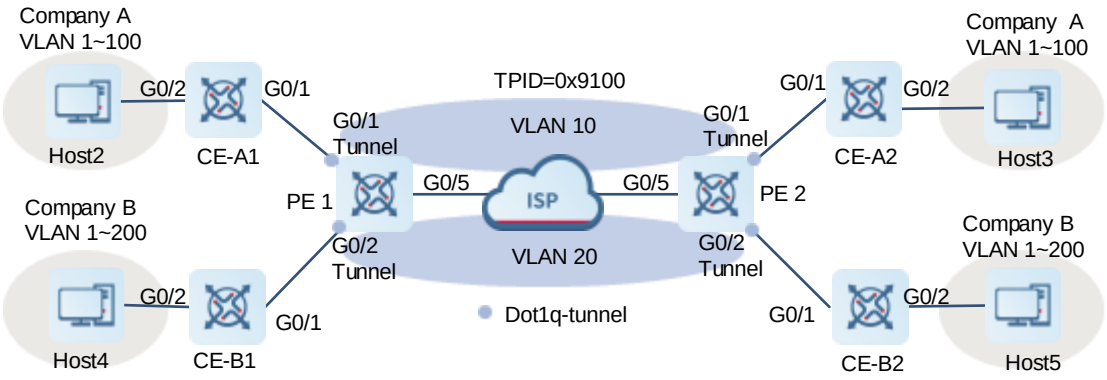
- The office network VLAN of Company A ranges from 1 to 100.
- The office network VLAN of Company B ranges from 1 to 200.

The priority replication and priority mapping functions of CoS can be configured on the dot1q-tunnel of the PE, and the QoS policy of CoS is configured to enable the client data packets to enjoy different QoS policies.

Generally, the TPID of the device is 0x8100 specified in IEEE 802.1Q. Because there are some devices in the SP network whose TPID is 0x9100, the TPID value needs to be modified on the uplink port of PE to ensure compatibility with these devices.

2. Topology

Figure 1-9 Configuring Basic QinQ for Layer 2 VPN



3. Notes

- Configure the interfaces on PE 1 and PE 2 connected to CE-A1, CE-A2, CE-B1 and CE-B2 as dot1q-tunnels.

The public network transmission VLAN is the native VLAN of the dot1q-tunnel, and belongs to the allowed untagged VLAN list of the port.

- Configure the interface on the PE that connects to the public network as a trunk, hybrid, or uplink port. When the connection interface between SP devices is a trunk or hybrid port, do not configure the native VLAN of the trunk or hybrid port as the native VLAN of dot1q-tunnel port. The reason is that, when the packets of the native VLAN is output from the trunk or hybrid port, the interface will strip the tag from the packet.
- The client VLAN is configured on the CE according to the user's needs, and the uplink port on the CE is configured as trunk, hybrid, or uplink port.

4. Procedure

(1) Configure the basic QinQ transmission channel VLAN 10 of Company A. VLAN 10 is the native VLAN of the dot1q-tunnel, and belongs to the allowed untagged VLAN list of the dot1q-tunnel.

On PE 1, configure VLAN 10 to implement tunnel transmission for the data of Company A's network.

```
PE1> privilege
PE1# config-system terminal
PE1(cfg)# vlan 10
PE1(cfg-vlan)# exit
PE1(cfg)# interface gigabitethernet 0/1
PE1(cfg-if-GigabitEthernet 0/1)# switch-port
PE1(cfg-if-GigabitEthernet 0/1)# switch-port mode dot1q-tunnel
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel pvid vlan 10
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel allow-pass vlan add untagged 10
PE1(cfg-if-GigabitEthernet 0/1)# exit
```

On PE 2, configure VLAN 10 to implement tunnel transmission for the data of Company A's network.

```
PE2> privilege
PE2# config-system terminal
PE2(cfg)# vlan 10
PE2(cfg-vlan)# exit
PE2(cfg)# interface gigabitethernet 0/1
PE2(cfg-if-GigabitEthernet 0/1)# switch-port
PE2(cfg-if-GigabitEthernet 0/1)# switch-port mode dot1q-tunnel
PE2(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel pvid vlan 10
PE2(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel allow-pass vlan add untagged 10
PE2(cfg-if-GigabitEthernet 0/1)# exit
```

(2) Configure the basic QinQ transmission channel VLAN 20 of Company B. VLAN 20 is the native VLAN of the dot1q-tunnel, and belongs to the allowed untagged VLAN list of the dot1q-tunnel.

On PE 1, configure VLAN 20 to implement tunnel transmission for the data of Company B's network.

```
PE1(cfg)# vlan 20
PE1(cfg-vlan)# exit
PE1(cfg)# interface gigabitethernet 0/2
PE1(cfg-if-GigabitEthernet 0/2)# switch-port
PE1(cfg-if-GigabitEthernet 0/2)# switch-port mode dot1q-tunnel
PE1(cfg-if-GigabitEthernet 0/2)# switch-port dot1q-tunnel pvid vlan 20
PE1(cfg-if-GigabitEthernet 0/2)# switch-port dot1q-tunnel allow-pass vlan add untagged 20
```

```
PE1(cfg-if-GigabitEthernet 0/2)# exit
```

On PE 2, configure VLAN 20 to implement tunnel transmission for the data of Company B's network.

```
PE2(cfg)# vlan 20
PE2(cfg-vlan)# exit
PE2(cfg)# interface gigabitethernet 0/2
PE2(cfg-if-GigabitEthernet 0/2)# switch-port
PE2(cfg-if-GigabitEthernet 0/2)# switch-port mode dot1q-tunnel
PE2(cfg-if-GigabitEthernet 0/2)# switch-port dot1q-tunnel pvid vlan 20
PE2(cfg-if-GigabitEthernet 0/2)# switch-port dot1q-tunnel allow-pass vlan add untagged 20
PE2(cfg-if-GigabitEthernet 0/2)# exit
```

(3) Configure the interface on the PE for connecting to the public network as a trunk port. VLAN 10 and VLAN 20 need to be added to the allowed VLAN list of the trunk port, but cannot be the native VLAN of the trunk port.

Configure the interface on PE 1 for connecting to the public network as a trunk port, and modify the TPID value of output packets to 0x9100.

```
PE1(cfg)# interface gigabitethernet 0/5
PE1(cfg-if-GigabitEthernet 0/5)# switch-port
PE1(cfg-if-GigabitEthernet 0/5)# switch-port mode trunk
PE1(cfg-if-GigabitEthernet 0/5)# switch-port trunk pvid vlan 1
PE1(cfg-if-GigabitEthernet 0/5)# switch-port trunk allow-pass vlan all
PE1(cfg-if-GigabitEthernet 0/5)# frame-tag tpid 9100
PE1(cfg-if-GigabitEthernet 0/5)# end
PE1# write
```

Configure the interface on PE 2 for connecting to the public network as a trunk port, and modify the TPID value of output packets to 0x9100.

```
PE2(cfg)# interface gigabitethernet 0/5
PE2(cfg-if-GigabitEthernet 0/5)# switch-port
PE2(cfg-if-GigabitEthernet 0/5)# switch-port mode trunk
PE2(cfg-if-GigabitEthernet 0/5)# switch-port trunk pvid vlan 1
PE2(cfg-if-GigabitEthernet 0/5)# switch-port trunk allow-pass vlan all PE2(cfg-if-GigabitEthernet 0/5)#
frame-tag tpid 9100
PE2(cfg-if-GigabitEthernet 0/5)# end
PE2# write
```

(4) On the CE, configure the interface for connecting to the PE as a trunk port, and configure the interface for connecting to the user terminal as an access port and add it to the client VLAN.

Create client VLANs 1–100 on CE-A1; configure the interface GigabitEthernet 0/1 for connecting to the PE as a trunk port; configure one client interface GigabitEthernet 0/2 to join VLAN 60 (the configuration of other client interfaces is omitted here).

```
CE-A1> privilege
CE-A1# config-system terminal
CE-A1(cfg)# vlan range 1-100
CE-A1(cfg-vlan)# exit
CE-A1(cfg)# interface gigabitethernet 0/1
```

```
CE-A1(cfg-if-GigabitEthernet 0/1)# switch-port
CE-A1(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
CE-A1(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
CE-A1(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
CE-A1(cfg-if-GigabitEthernet 0/1)# exit
CE-A1(cfg)# interface gigabitethernet 0/2
CE-A1(cfg-if-GigabitEthernet 0/2)# switch-port
CE-A1(cfg-if-GigabitEthernet 0/2)# switch-port mode access
CE-A1(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 60
```

Create client VLANs 1–100 on CE-A2; configure the interface GigabitEthernet 0/1 for connecting to the PE as a trunk port; configure one client interface GigabitEthernet 0/2 to join VLAN 60 (the configuration of other client interfaces is omitted here).

```
CE-A2> privilege
CE-A2# config-system terminal
CE-A2(cfg)# vlan range 1-100
CE-A2(cfg-vlan)# exit
CE-A2(cfg)# interface gigabitethernet 0/1
CE-A2(cfg-if-GigabitEthernet 0/1)# switch-port
CE-A2(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
CE-A2(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
CE-A2(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
CE-A2(cfg-if-GigabitEthernet 0/1)#exit
CE-A2(cfg)# interface gigabitethernet 0/2
CE-A2(cfg-if-GigabitEthernet 0/2)# switch-port
CE-A2(cfg-if-GigabitEthernet 0/2)# switch-port mode access
CE-A2(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 60
```

Create client VLANs 1–200 on CE-B1; configure the interface GigabitEthernet 0/1 for connecting to the PE as a trunk port; configure one client interface GigabitEthernet 0/2 to join VLAN 60 (the configuration of other client interfaces is omitted here).

```
CE-B1> privilege
CE-B1# config-system terminal
CE-B1(cfg)# vlan range 1-200
CE-B1(cfg-vlan)# exit
CE-B1(cfg)# interface gigabitethernet 0/1
CE-B1(cfg-if-GigabitEthernet 0/1)# switch-port
CE-B1(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
CE-B1(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
CE-B1(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
CE-B1(cfg-if-GigabitEthernet 0/1)# exit
CE-B1(cfg)# interface gigabitethernet 0/2
CE-B1(cfg-if-GigabitEthernet 0/2)# switch-port
CE-B1(cfg-if-GigabitEthernet 0/2)# switch-port mode access
CE-B1(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 60
```

Create client VLANs 1–200 on CE-B2; configure the interface GigabitEthernet 0/1 for connecting to the PE as a trunk port; configure the client interface GigabitEthernet 0/2 to join VLAN 60 (the configuration of other client interfaces is omitted here).

```
CE-B2> privilege
CE-B2# config-system terminal
CE-B2(cfg)# vlan range 1-200
CE-B2(cfg-vlan)# exit
CE-B2(cfg)# interface gigabitethernet 0/1
CE-B2(cfg-if-GigabitEthernet 0/1)# switch-port
CE-B2(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
CE-B2(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
CE-B2(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
CE-B2(cfg-if-GigabitEthernet 0/1)# exit
CE-B2(cfg)# interface gigabitethernet 0/2
CE-B2(cfg-if-GigabitEthernet 0/2)# switch-port
CE-B2(cfg-if-GigabitEthernet 0/2)# switch-port mode access
CE-B2(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 60
```

5. Verification

(1) Display the interface configuration of PEs.

Display the interface configuration of PE 1.

```
PE1# display interfaces switch-port
Interface                 Switch-port Mode      Access Native Protected VLAN lists
--------------------------------- ---------- --------- ------ ------ --------- -----------------
GigabitEthernet 0/1       enabled    TUNNEL    10     Disabled  1,10
GigabitEthernet 0/2       enabled    TUNNEL    20     Disabled  1,20
GigabitEthernet 0/5       enabled    TRUNK     1      Disabled  ALL
```

Display the interface configuration of PE 2.

```
PE2# display interfaces switch-port
Interface                 Switch-port Mode      Access Native Protected VLAN lists
--------------------------------- ---------- --------- ------ ------ --------- -----------------
GigabitEthernet 0/1       enabled    TUNNEL    10     Disabled  1,10
GigabitEthernet 0/2       enabled    TUNNEL    20     Disabled  1,20
GigabitEthernet 0/5       enabled    TRUNK     1      Disabled  ALL
```

(2) Display the dot1q-tunnel configuration of PEs.

Display the dot1q-tunnel configuration of PE 1.

```
PE1# display interfaces dot1q-tunnel
========Interface Gi0/1========
Native vlan: 10
Allowed vlan list:1,10,
Tagged vlan list:
========Interface Gi0/2========
Native vlan: 20
Allowed vlan list:1,20,
```

```
Tagged vlan list:
```

Display the dot1q-tunnel configuration of PE 2.

```
PE2# display interfaces dot1q-tunnel
========Interface Gi0/1========
Native vlan: 10
Allowed vlan list:1,10,
Tagged vlan list:
========Interface Gi0/2========
Native vlan: 20
Allowed vlan list:1,20,
Tagged vlan list:
```

(3) Display the uplink port configuration of PEs.

Display the uplink port configuration of PE 1.

```
PE1# display frame-tag tpid
Ports              Tpid
------------    ----------
Gi0/5              0x9100
```

Display the uplink port configuration of PE 2.

```
PE2# display frame-tag tpid
Ports              Tpid
------------    ----------
Gi0/5              0x9100
```

(4) Display the CE configuration of Company A.

Display the configuration of CE-A1.

```
CE-A1(cfg)# display interface switch-port
Interface                          Switch-port Mode         Access Native Protected VLAN lists
---------------------------------- ---------- --------- ------ ------ --------- ----------------
GigabitEthernet 0/1                enabled      TRUNK       1        1         Disabled    ALL
GigabitEthernet 0/2                enabled      ACCESS      60       1         Disabled    ALL
```

Display the configuration of CE-A2.

```
CE-A2(cfg)#display interface switch-port
Interface                          Switch-port Mode         Access Native Protected VLAN lists
---------------------------------- ---------- --------- ------ ------ --------- ----------------
GigabitEthernet 0/1                enabled      TRUNK       1        1         Disabled    ALL
GigabitEthernet 0/2                enabled      ACCESS      60       1         Disabled    ALL
```

(5) Take Hosts 2 and 3 as examples to verify that clients in Company A can implement layer-2 interworking.

Configure an IP address for Host 2.

```
Host2> privilege
Host2# config-system terminal
Host2(cfg)#interface gigabitethernet 0/1
Host2(cfg-if-GigabitEthernet 0/1)#unset switch-port
Host2(cfg-if-GigabitEthernet 0/1)#ip address 192.168.60.2/24
```

Configure an IP address for Host 3.

```
Host3> privilege
Host3# config-system terminal
Host3(cfg)#interface gigabitethernet 0/1
Host3(cfg-if-GigabitEthernet 0/1)#unset switch-port
Host3(cfg-if-GigabitEthernet 0/1)#ip address 192.168.60.3/24
```

Verify that Host 2 can ping through Host 3.

```
Host2# ping 192.168.60.3
Sending 5, 100-byte ICMP Echoes to 192.168.60.3, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 3/4/6 ms.
```

(6) Take Hosts 4 and 5 as examples to verify that clients in Company B can implement layer-2 interworking.

Configure an IP address for Host 4.

```
Host4> privilege
Host4# config-system terminal
Host4(cfg)#interface gigabitethernet 0/1
Host4(cfg-if-GigabitEthernet 0/1)#unset switch-port
Host4(cfg-if-GigabitEthernet 0/1)#ip address 192.168.60.4/24
```

Configure an IP address for Host 5.

```
Host5> privilege
Host5# config-system terminal
Host5(cfg)#interface gigabitethernet 0/1
Host5(cfg-if-GigabitEthernet 0/1)#unset switch-port
Host5(cfg-if-GigabitEthernet 0/1)#ip address 192.168.60.5/24
```

Verify that Host 4 can ping through Host 5.

```
Host4# ping 192.168.60.5
Sending 5, 100-byte ICMP Echoes to 192.168.60.5, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 3/4/6 ms.
```

(7) Host 2 in Company A cannot ping through Host 5 in Company B, which indicates that clients in different companies cannot implement layer-2 interworking.

```
Host2# ping 192.168.60.5
Sending 5, 100-byte ICMP Echoes to 192.168.60.5, timeout is 2 seconds:
  < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5).
```

6. Configuration Files

- Configuration files of PE 1 and PE 2

```
vlan range 1,10,20
!
```

```
interface GigabitEthernet 0/1
switch-port
 switch-port mode dot1q-tunnel
 switch-port dot1q-tunnel allow-pass vlan add untagged 10
 switch-port dot1q-tunnel pvid vlan 10
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode dot1q-tunnel
 switch-port dot1q-tunnel allow-pass vlan add untagged 20
 switch-port dot1q-tunnel pvid vlan 20
!
interface GigabitEthernet 0/5
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094 frame-tag tpid 0x9100
```

- Configuration files of CE-A1 and CE-A2 (Only GigabitEthernet 0/2 is taken as an example of interfaces connected to user terminals; the configuration files of the other interfaces are omitted.)

```
vlan range 1-100
!
interface gigabitethernet 0/1
switch-port
switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface gigabitethernet 0/2switch-port access vlan 60
```

- Configuration files of CE-B1 and CE-B2 (Only GigabitEthernet 0/2 is taken as an example of interfaces connected to user terminals; the configuration files of the other interfaces are omitted.)

```
vlan range 1-200

interface gigabitethernet 0/1
switch-port
switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface gigabitethernet 0/2
 switch-port
 switch-port mode access
 switch-port access vlan 60
```

7. Common Errors

- The same company has different native VLANs on the dot1q-tunnel ports of PE 1 and PE 2, which leads to

the failure of interworking.

- The native VLAN of the dot1q-tunnel is not added to the allowed VLAN list of the interface in the untagged form.
- The TPID value of a third-party device is not the default value 0x8100, but the TPID is not configured on the egress connected to the third-party device, and therefore the packet cannot be recognized by the third-party device.

1.15.2 Configuring C-Tag-Based Selective QinQ to Implement Layer 2 VPN and Service Flow Management

1. Requirements

Broadband Internet access and IPTV are important services carried by the metropolitan area network (MAN). MAN SPs allocate different service flows to different SP VLANs to achieve differentiated management, and provide QoS policy services for these VLANs or CoSs.

The basic QinQ function can encapsulate the client data packets received by the tunnel port only with the outer tag with the same VLAN ID, and the VLAN ID of outer tags is the native VLAN of the dot1q-tunnel; the basic QinQ cannot meet the requirement for differentiating SP VLANs in multi-service scenarios.

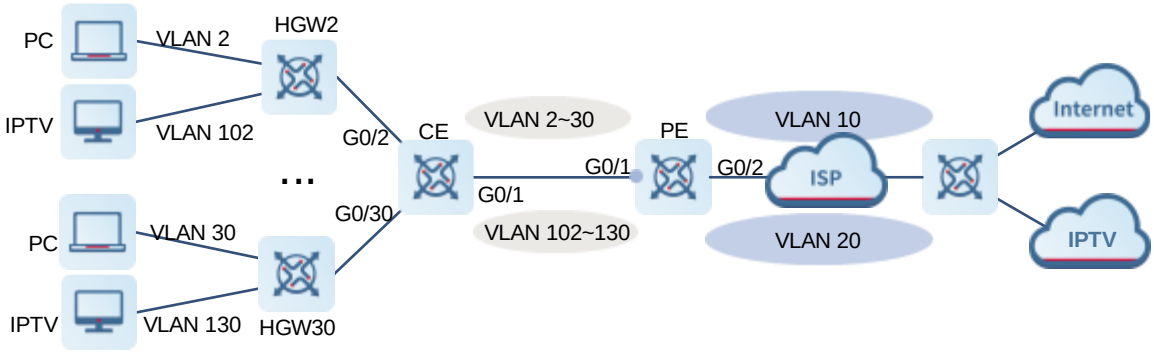
Selective QinQ can flexibly add the SP outer tag (S-Tag) according to the tag (C-Tag) of client packets, encapsulate the client service flows received by the same dot1q-tunnel into different SP VLANs, and use the QoS policy of SP network to guarantee reliability while transparently transmitting client data.

As shown in Figure 1-10, devices involved include provider edges (PEs), customer edges (CEs), and home gateways (HGWs) 1–30.

- The broadband Internet access data and IPTV service data are allocated to different VLANs. VLANs 2–30 are for client broadband Internet access service flows, and VLANs 102–130 are for client IPTV service flows.
- On the public network, the broadband Internet access data is transmitted through SP VLAN 10 to implement data communication with the Internet; the IPTV service flow is transmitted through VLAN 20 to implement data communication with the IPTV network. The dot1q-tunnel is configured on PE, and different VLANs are encapsulated to distinguish different service data. Client VLANs 2–30 are encapsulated as SP VLAN 10, while client VLANs 102–130 are encapsulated as SP VLAN 20.
- Client network packets carry tags and contain the CoS priority. The SP network enables the priority replication function on PEs, and uses the CoS value for the inner tag of packets in the outer tag so that service flows of the client network can use the QoS policy of the SP network to ensure transmission.

2. Topology

Figure 1-10 Selective QinQ Implements Layer-2 VPN



3. Notes

- Configure the interface GigabitEthernet 0/1 of PE for connecting to CE as dot1q-tunnel. Add SP VLANs 10 and 20 to the untagged VLAN list of the dot1q-tunnel connected to the client network so that the packets from the peer end return to the dot1q-tunnel and the SP tag is stripped when packet output.
- On the downlink interface of PE, configure the selective QinQ function of adding outer VLAN tags based on C-Tag. If the dot1q-tunnel receives the frames of client VLANs 2–30 (C-Tag), the frames will be encapsulated with the tag (S-Tag) of SP VLAN 10; if the dot1q-tunnel receives the frames of client VLANs 102–130 (C-Tag), the frames will be encapsulated with the tag (S-Tag) of SP VLAN 20.
- The SP network provides the CoS-based QoS priority policy. You can configure priority replication to replicate the CoS value of the client VLAN tag to the CoS of the outer VLAN tag so that the packet encapsulated with the outer tag is transmitted based on the priority policy for the client VLAN tag.
- The interface of PE for connecting to the public network is configured as an uplink port to transmit the data of VLAN 10 and VLAN 20.
- The interface of CE for connecting to the PE is configured as an uplink port to transmit the data of VLANs 2–30 and VLANs 102–130.
- The interface of CE for connecting to HGW is configured as a trunk port, and the VLAN transmitted by HGW device is added to the allowed VLAN list of the trunk port. The clients adopting HGW B use VLAN 2 to transmit broadband Internet access data and VLAN 102 to transmit IPTV data. The clients adopting HGW 30 use VLAN 30 to transmit broadband Internet access data and VLAN 130 to transmit IPTV data. The configurations of other clients are similar.

4. Procedure

(1) Configure the C-Tag-based QinQ encapsulation rule.

Add SP VLAN 10 and VLAN 20 to the untagged VLAN list of the dot1q-tunnel GigabitEthernet 0/1.

```
PE> privilege
PE# config-system terminal
PE(cfg)# vlan range 10,20
PE(cfg-vlan-range)# exit
PE(cfg)# interface gigabitethernet 0/1
```

```
PE(cfg-if-GigabitEthernet 0/1)# switch-port
PE(cfg-if-GigabitEthernet 0/1)# switch-port mode dot1q-tunnel
PE(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel pvid vlan 1
PE(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel allow-pass vlan add untagged 1,10,20
```

If the dot1q-tunnel receives the frames of client VLANs 2–30 (C-Tag), the frames will be encapsulated with the tag (S-Tag) of SP VLAN 10.

```
PE(cfg-if-GigabitEthernet 0/1)# dot1q outer-vid 10 register inner-vid 2-30
```

If the dot1q-tunnel receives the frames of client VLANs 102–130 (C-Tag), the frames will be encapsulated with the tag (S-tag) of SP VLAN 20.

```
PE(cfg-if-GigabitEthernet 0/1)# dot1q outer-vid 20 register inner-vid 102-130
```

(2) On the dot1q-tunnel, replicate the priority of inner tag in the client packet to the outer tag.

```
PE(cfg-if-GigabitEthernet 0/1)# mls qos trust cos
PE(cfg-if-GigabitEthernet 0/1)# inner-priority-trust enable
PE(cfg-if-GigabitEthernet 0/1)# exit
```

(3) Configure the interface on the PE for connecting to the public network as an uplink port.

```
PE(cfg)# interface gigabitethernet 0/2
PE(cfg-if-GigabitEthernet 0/2)# switch-port
PE(cfg-if-GigabitEthernet 0/2)# switch-port mode uplink
PE(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
PE(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan all PE(cfg-if-GigabitEthernet 0/2)# end
PE# write
```

(4) Create a client VLAN on CE, and configure the interface for connecting PE as an uplink port.

```
CE> privilege
CE# config-system terminal
CE(cfg)# vlan range 2-30,102-130
CE(cfg-vlan)# exit
CE(cfg)# interface gigabitethernet 0/1
CE(cfg-if-GigabitEthernet 0/1)# switch-port
CE(cfg-if-GigabitEthernet 0/1)# switch-port mode uplink
CE(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
CE(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
CE(cfg-if-GigabitEthernet 0/1)# exit
```

(5) Configure the interface on the CE for connecting to the HGW as a trunk port. Gigabit Ethernet 0/2 and Gigabit Ethernet 0/30 are taken as examples below. The configurations of the other interface are similar.

```
CE(cfg)# interface gigabitethernet 0/2
CE(cfg-if-GigabitEthernet 0/2)# switch-port
CE(cfg-if-GigabitEthernet 0/2)# switch-port mode trunk
CE(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
CE(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan only 2,102
CE(cfg-if-GigabitEthernet 0/2)# exit
CE(cfg)# interface gigabitethernet 0/30
CE(cfg-if-GigabitEthernet 0/30)# switch-port
CE(cfg-if-GigabitEthernet 0/30)# switch-port mode trunk
```

```
CE(cfg-if-GigabitEthernet 0/30)# switch-port trunk pvid vlan 1
CE(cfg-if-GigabitEthernet 0/30)# switch-port trunk allow-pass vlan only 30,130
CE(cfg-if-GigabitEthernet 0/30)# end
CE# write
```

5. Verification

(1) Display the PE configuration.

Display the interface configuration of the PE.

```
PE# display interfaces switch-port
Interface                          Switch-port Mode         Access Native Protected VLAN lists
---------------------------------- ---------- --------- ------ ------ --------- ----------------
GigabitEthernet 0/1                enabled    TUNNEL      1        Disabled   1,10,20
GigabitEthernet 0/2                enabled    UPLINK      1        Disabled   ALL
```

Display the dot1q-tunnel configuration of the PE. The outer tag VLAN has been added to the allowed VLAN list of the interface.

```
PE# display interfaces dot1q-tunnel
========Interface Gi0/1========
Native vlan: 1
Allowed vlan list:1,10,20
Tagged vlan list:
```

Verify that the encapsulation policy relationship for adding S-tags based on C-tags is correct.

```
PE# display registration-table
Ports      Type              Outer-VID      Inner-VID-list
------     ----------        ----------     --------------
Gi0/1      Add-outer         10             2-30
Gi0/1      Add-outer         20             102-130
```

Display the priority replication relationship of the interface.

```
PE# display inner-priority-trust interfaces gigabitethernet 0/1
Port      Inner-priority-trust
-----   ------------------
Gi0/1   Enable
```

(2) Display the CE configuration.

Display the interface configuration of the CE.

```
CE# display interfaces switch-port
Interface                          Switch-port Mode         Access Native Protected VLAN lists
---------------------------------- ---------- --------- ------ ------ --------- ----------------
GigabitEthernet 0/1                enabled    Uplink      1        Disabled   ALL
GigabitEthernet 0/2                enabled    Trunk       1        Disabled   2,102
GigabitEthernet 0/3                enabled    Trunk       1        Disabled   3,103
……
GigabitEthernet 0/30               enabled    Trunk       1        Disabled   30,130
```

6. Configuration Files

- Configuration file of the PE

```
vlan range 1,10,20
!
interface GigabitEthernet 0/1
switch-port
 switch-port mode dot1q-tunnel
 switch-port dot1q-tunnel allow-pass vlan add untagged 1, 10,20
 switch-port dot1q-tunnel pvid vlan 1
 dot1q outer-vid 100 register inner-vid 2-30
 dot1q outer-vid 200 register inner-vid 102-130
 mls qos trust cos
 inner-priority-trust enable
!
interface GigabitEthernet 0/2
 switch-port
switch-port mode uplink
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
```

- Configuration file of the CE (Only Gigabit Ethernet 0/2 and Gigabit Ethernet 0/30 are taken as examples of interfaces connected to the HGW. The configurations of the other interfaces are similar.)

```
vlan range 1-30,102-130
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode uplink
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 2,102
!
interface GigabitEthernet 0/30
 switch-port
switch-port mode trunk
switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 30,130
```

1.15.3 Configuring ACL-Based Selective QinQ to Implement Layer 2 VPN and Service Flow Management

Note

This function is not supported by the S6980-64QC, S6920-4C, S6580-48CQ8QC, or S6580-24DC8QC.

1. Requirements

If the service flows of client network are not classified according to VLAN, but according to the MAC address, IP address, or protocol type, or there are a large number of old low-end network access devices on the client network and the service flows cannot be effectively distinguished using VLAN ID, the packets from the client network cannot be encapsulated with outer tags based on their C-tags to realize transparent transmission and implement QoS policies.

ACL can classify service flows according to the MAC address, IP address, or protocol type, and selective QinQ can distinguish different service flows using ACL, add and modify outer tags, and assign different service data to different VLANs and implement the VPN channel for transparent transmission of client services. If the SP network provides different QoS service policies for different services, it can guarantee first transmission of high-priority services.

As shown in Figure 1-11, internal networks of the company's branches in two cities are connected to CE 1 and CE 2, and then connected to the SP network through PE 1 and PE 2.

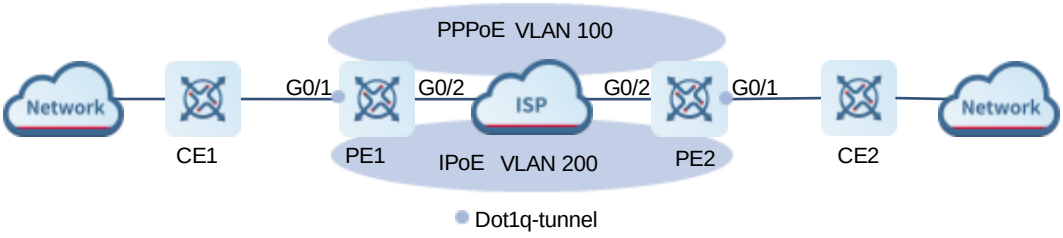
The selective QinQ function based on ACL is used on PE 1 and PE 2 to encapsulate different service data of the company with different outer VLAN tags so that the data can be transmitted through the SP network transparently and the branches can access each other's network.

- ACL 1 identifies service flows of PPPoE type and allocates the service flows to VLAN 100 for transmission.
- ACL 2 identifies service flows of IPTV (IPoE) type and allocates the service flows to VLAN 200 for transmission.

The SP network sets the QoS policy based on CoS. You can modify the CoS values of packets through priority mapping on PE 1 and PE 2 to ensure preferential transmission of service flows of the client network using the QoS policy of the SP network.

2. Topology

Figure 1-11 Configuring ACL-Based Selective QinQ for Layer 2 VPN



3. Notes

- Configure the ACL policy on PE 1 and PE 2 to distinguish different service flows of the client network: the protocol type of PPPoE packets is 0x8863 or 0x8864, and that of IPoE packets is 0x0800. Configure the interface for connecting to the CE as dot1q-tunnel, and configure the selective QinQ function based on ACL. The configuration of PE 2 is similar to that of PE 1. Only PE 1 is taken as an example in the configuration procedure.
- The SP network provides a CoS-based QoS priority policy, and the packets of client network are attached with tags. You can configure priority mapping to modify the CoS value of the outer tag. In this way, the packet is transmitted based on the corresponding QoS priority policy of the SP after being encapsulated with

an outer tag, and flexible control is implemented on the packet priority on the dot1q-tunnel.

4. Procedure

(1) On the PE, create an ACL for distinguishing service flows.

```
PE1> privilege
PE1# config-system terminal
PE1(cfg)# expert access-control-list extended acl1
PE1(cfg-exp-nacl)# permit 0x8863 any any
PE1(cfg-exp-nacl)# permit 0x8864 any any
PE1(cfg-exp-nacl)# exit
PE1(cfg)# expert access-control-list extended acl2
PE1(cfg-exp-nacl)# permit 0x0800 any any
PE1(cfg-exp-nacl)# exit
```

(2) Configure the ACL-based QinQ encapsulation rule.

Add the SP VLAN 100 and VLAN 200 to the untagged VLAN list of the dot1q-tunnel GigabitEthernet 0/1.

```
PE1(cfg)# vlan rang 100,200
PE1(cfg-vlan-range)# exit
PE1(cfg)# interface gigabitethernet 0/1
PE1(cfg-if-GigabitEthernet 0/1)# switch-port
PE1(cfg-if-GigabitEthernet 0/1)# switch-port mode dot1q-tunnel
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel pvid vlan 1
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel allow-pass vlan add untagged 1,100,200
```

If the dot1q-tunnel receives frames matching ACL1, the frames will be encapsulated with the tag of SP VLAN 100.

```
PE1(cfg-if-GigabitEthernet 0/1)# traffic-redirect access-apply acl1 nested-vlan 100 in
```

If the dot1q-tunnel receives frames matching ACL2, the frames will be encapsulated with the tag of SP VLAN 200.

```
PE1(cfg-if-GigabitEthernet 0/1)# traffic-redirect access-apply acl2 nested-vlan 200 in
```

Configure the MAC address replication function to replicate the learned MAC address from the default native VLAN 1 of the dot1q-tunnel to VLAN 100 and VLAN 200 of outer tag.

```
PE1(cfg-if-GigabitEthernet 0/1)# mac-address-mapping 1 source-vlan 1 destination-vlan 100,200
```

(3) Configure the outer tag priority of packets according to the inner tag priority of packets.

```
PE1(cfg-if-GigabitEthernet 0/1)# mls qos trust cos
PE1(cfg-if-GigabitEthernet 0/1)# unset inner-priority-trust enable
PE1(cfg-if-GigabitEthernet 0/1)# dot1q-tunnel cos 3 remark-cos 5
PE1(cfg-if)# exit
```

(4) Configure the interface connected to the public network as an uplink port.

```
PE1(cfg)# interface gigabitethernet 0/2
PE1(cfg-if-GigabitEthernet 0/2)# switch-port
PE1(cfg-if-GigabitEthernet 0/2)# switch-port mode uplink
PE1(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
PE1(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan allPE1(cfg-if-GigabitEthernet 0/2)# end
```

```
PE1# write
```

5. Verification

(1) Display the interface configuration of the PE.

```
PE1# display interfaces switch-port
Interface                          Switch-port Mode         Access Native Protected VLAN lists
-------------------------------- ---------- --------- ------ ------ --------- ----------------
GigabitEthernet 0/1                enabled    TUNNEL        1      Disabled  1,100,200
GigabitEthernet 0/2                enabled    UPLINK        1      Disabled  ALL
```

(2) Verify the encapsulation policy for adding tags based ACL.

Note

This command is not supported by the S6920-4C.

```
PE1# display traffic-redirect
Ports          Type             VID        Match-filter
-------      ----------      -----    ------------
Gi0/1          Nested-vid     100          acl1
Gi0/1          Nested-vid     200          acl2
```

(3) Display the priority mapping relationship on the interface.

```
PE1# display interfaces gigabitethernet 0/1 remark
Ports            From COS   To COS
------------ --------- --------
Gi0/1            3            5
```

(4) The clients of the same service in the company's branch can realize layer-2 VPN interworking. Specific service data can enjoy a higher transmission priority.

6. Configuration Files

Configuration files of PE 1 and PE 2

```
expert access-control-list extended acl1
  10 permit 0x8863 any any
  20 permit 0x8864 any any
!
expert access-control-list extended acl2
  10 permit 0x800 any any
!
vlan range 1,100,200
!
interface GigabitEthernet 0/1
switch-port
  switch-port mode dot1q-tunnel
  switch-port dot1q-tunnel allow-pass vlan add untagged 1,100,200
switch-port dot1q-tunnel pvid vlan 1
dot1q-tunnel cos 3 remark-cos 5
  traffic-redirect access-apply acl1 nested-vlan 100 in
```

```
 traffic-redirect access-apply acl2 nested-vlan 200 in
 mls qos trust cos
!
 interface GigabitEthernet 0/2
 switch-port
switch-port mode uplink
switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

7. Common Errors

- The ACL policy is configured incorrectly. The ACL policy number is not in the parameter range. The ACL name is already taken.
- Priority replication is not disabled before priority mapping is configured.
- The MAC address replication function is not configured, which may lead to packet flooding.

1.15.4 Configuring VLAN Mapping to Implement Aggregation of Different Services

1. Requirements

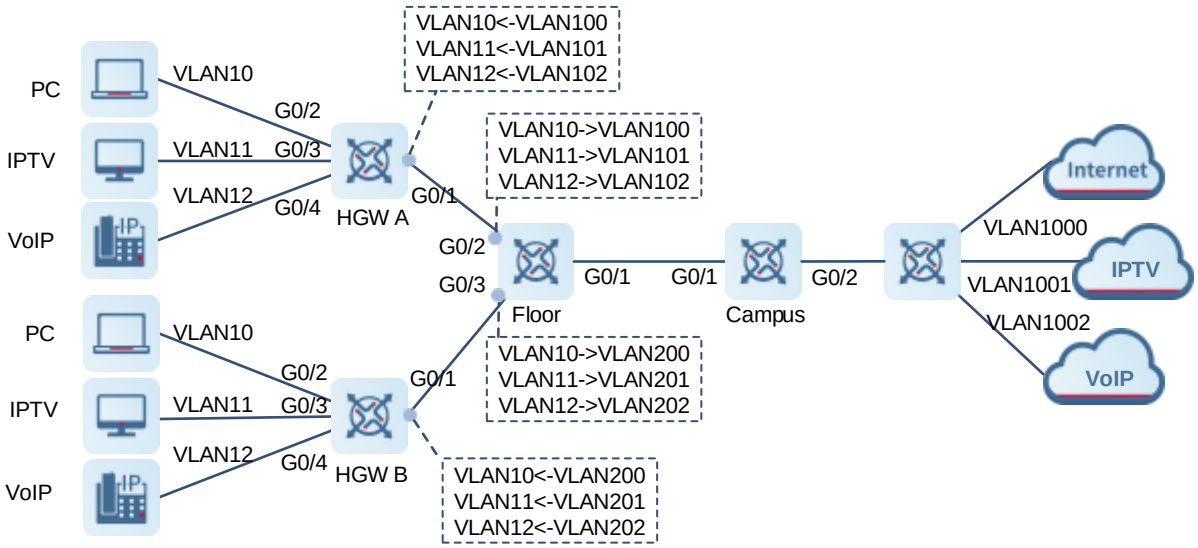
As shown in Figure 1-12, in a campus network environment, different services on the HGW are differentiated using different VLANs. VLAN 10 is used for PC services, VLAN 11 for IPTV services, and VLAN 12 for VoIP services.

For the sake of easy deployment, the same VLAN is used for the same service of HGW A and HGW B, and the packets of the same service from different homes are isolated using the VLAN mapping function on the floor devices, thus avoiding confusion of the packets from different homes.

Packets of the same service from different homes are sent from campus devices through the same VLAN to save VLAN resources.

2. Topology

Figure 1-12 Configuring VLAN Mapping for Aggregation of Different Services



3. Notes

- HGW A

  Create a VLAN after mapping, configure GigabitEthernet 0/1 as an uplink port, and configure the VLAN mapping policy for the input direction on GigabitEthernet 0/1 for the downlink data: Map VLAN 100 as VLAN 10 for forwarding to Gigabit Ethernet 0/2; map VLAN 101 as VLAN 11 for forwarding to Gigabit Ethernet 0/3; map VLAN 102 as VLAN 12 for forwarding to Gigabit Ethernet 0/4. If VLAN 200, VLAN 201, and VLAN 202 are not created, their data will be dropped even if the data is forwarded to Gigabit Ethernet 0/1.

  Configure Gigabit Ethernet 0/2, Gigabit Ethernet 0/3, and Gigabit Ethernet 0/4 as an access port, and add them to the VLAN where the corresponding service is.

- HGW B

  Create a VLAN after mapping, configure GigabitEthernet 0/1 as an uplink port, and configure the VLAN mapping policy for the input direction on GigabitEthernet 0/1 for the downlink data: Map VLAN 200 as VLAN 10 for forwarding to Gigabit Ethernet 0/2; map VLAN 201 as VLAN 11 for forwarding to Gigabit Ethernet 0/3; map VLAN 202 as VLAN 12 for forwarding to Gigabit Ethernet 0/4. If VLAN 100, VLAN 101, and VLAN 102 are not created, their data will be dropped even if the data is forwarded to Gigabit Ethernet 0/1.

  Configure Gigabit Ethernet 0/2, Gigabit Ethernet 0/3, and Gigabit Ethernet 0/4 as an access port, and add them to the VLAN where the corresponding service is.

- Floor device

  Create a VLAN after mapping, and configure GigabitEthernet 0/1 of the floor device as an uplink port.

  Configure the downlink interface GigabitEthernet 0/2 as an uplink port, and configure the VLAN mapping policy for the input direction for the uplink data on HGW A: Map VLAN 10 as VLAN 100; map VLAN 11 as VLAN 101; map VLAN 12 as VLAN 102. To avoid waste of the bandwidth, configure GigabitEthernet 0/2 to allow only the data of VLANs 100–102 to pass through.

  Configure the downlink interface GigabitEthernet 0/3 as an uplink port, and configure the VLAN mapping policy for the input direction for the uplink data on HGW B: Map VLAN 10 as VLAN 200; map VLAN 11 as VLAN 201; map VLAN 12 as VLAN 202. To avoid waste of the bandwidth, configure GigabitEthernet 0/3 to allow only the data of VLANs 200 to 202 to pass through.

- Campus device

  On Gigabit Ethernet 0/1, configure the C-tag-based QinQ encapsulation policy to encapsulate the packets of the same service with the same outer tag for transmission an access port: the data frames of VLAN 100 and VLAN 200 are encapsulated as frames of VLAN 1000 and forwarded to the Internet for communication; the data frames of VLAN 101 and VLAN 201 are encapsulated as frames of VLAN 1001 and forwarded to the IPTV network; the data frames of VLAN 102 and VLAN 202 are encapsulated as frames of VLAN 1002 and forwarded to the VoIP network.

  Configure GigabitEthernet 0/2 as an uplink port.

4. Procedure

(1) Configure HGW A.

Create a VLAN after mapping, configure the uplink port, and configure the VLAN mapping policy for input direction for the downlink data.

```
HGWA> privilege
HGWA# config-system terminal
HGWA(cfg)# vlan range 10-12
HGWA(cfg-vlan-range)# exit
HGWA(cfg)# interface gigabitethernet 0/1
HGWA(cfg-if-GigabitEthernet 0/1)# switch-port
HGWA(cfg-if-GigabitEthernet 0/1)# switch-port mode uplink
HGWA(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
HGWA(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
HGWA(cfg-if-GigabitEthernet 0/1)# vlan-mapping-in vlan 100 remark 10
HGWA(cfg-if-GigabitEthernet 0/1)# vlan-mapping-in vlan 101 remark 11
HGWA(cfg-if-GigabitEthernet 0/1)# vlan-mapping-in vlan 102 remark 12
HGWA(cfg-if-GigabitEthernet 0/1)# exit
```

Configure the downlink interface as an access port, and add it to the VLAN where the corresponding service is.

```
HGWA(cfg)# interface list gigabitethernet 0/2-4
HGWA(cfg-if-list)# switch-port
HGWA(cfg-if-list)# switch-port mode access
HGWA(cfg-if-list)# exit
HGWA(cfg)# interface gigabitethernet 0/2
HGWA(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 10
HGWA(cfg-if-GigabitEthernet 0/2)# exit
HGWA(cfg)# interface gigabitethernet 0/3
HGWA(cfg-if-GigabitEthernet 0/3)# switch-port access vlan 11
HGWA(cfg-if-GigabitEthernet 0/3)# exit
HGWA(cfg)# interface gigabitethernet 0/4
HGWA(cfg-if-GigabitEthernet 0/4)# switch-port access vlan 12
HGWA(cfg-if-GigabitEthernet 0/4)# end
HGWA# write
```

(2) Configure HGW B.

Create a VLAN after mapping, configure the uplink port, and configure the VLAN mapping policy for input direction for the downlink data.

```
HGWB> privilege
HGWB# config-system terminal
HGWB(cfg)# vlan range 10-12
HGWB(cfg-vlan-range)# exit
HGWB(cfg)# interface gigabitethernet 0/1
HGWB(cfg-if-GigabitEthernet 0/1)# switch-port
HGWB(cfg-if-GigabitEthernet 0/1)# switch-port mode uplink
HGWB(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
HGWB(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
HGWB(cfg-if-GigabitEthernet 0/1)# vlan-mapping-in vlan 200 remark 10
HGWB(cfg-if-GigabitEthernet 0/1)# vlan-mapping-in vlan 201 remark 11
HGWB(cfg-if-GigabitEthernet 0/1)# vlan-mapping-in vlan 202 remark 12
```

Configure the downlink interface as an access port, and add it to the VLAN where the corresponding service is.

```
HGWB(cfg)# interface list gigabitethernet 0/2-4
HGWB(cfg-if-list)# switch-port
HGWB(cfg-if-list)# switch-port mode access
HGWB(cfg-if-list)# exit
HGWB(cfg)# interface gigabitethernet 0/2
HGWB(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 10
HGWB(cfg-if-GigabitEthernet 0/2)# exit
HGWB(cfg)# interface gigabitethernet 0/3
HGWB(cfg-if-GigabitEthernet 0/3)# switch-port access vlan 11
HGWB(cfg-if-GigabitEthernet 0/3)# exit
HGWB(cfg)# interface gigabitethernet 0/4
HGWB(cfg-if-GigabitEthernet 0/4)# switch-port access vlan 12
HGWB(cfg-if-GigabitEthernet 0/4)# end
HGWB# write
```

(3) Configure the floor device.

Create a VLAN after mapping, and for the uplink data of HGW A, create the VLAN mapping policy for input direction on the downlink interface.

```
Floor> privilege
Floor# config-system terminal
Floor(cfg)# vlan range 100-102
Floor(cfg-vlan-range)# exit
Floor(cfg)# interface gigabitethernet 0/2
Floor(cfg-if-GigabitEthernet 0/2)# switch-port
Floor(cfg-if-GigabitEthernet 0/2)# switch-port mode uplink
Floor(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
Floor(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan only 100-102
Floor(cfg-if-GigabitEthernet 0/2)# vlan-mapping-in vlan 10 remark 100
Floor(cfg-if-GigabitEthernet 0/2)# vlan-mapping-in vlan 11 remark 101
Floor(cfg-if-GigabitEthernet 0/2)# vlan-mapping-in vlan 12 remark 102
Floor(cfg-if-GigabitEthernet 0/2)# exit
```

Create a VLAN after mapping, and for the uplink data of HGW B, create the VLAN mapping policy for input direction on the downlink interface.

```
Floor(cfg)# vlan range 200-202
Floor(cfg-vlan-range)# exit
Floor(cfg)# interface gigabitethernet 0/3
Floor(cfg-if-GigabitEthernet 0/3)# switch-port
Floor(cfg-if-GigabitEthernet 0/3)# switch-port mode uplink
Floor(cfg-if-GigabitEthernet 0/3)# switch-port trunk pvid vlan 1
Floor(cfg-if-GigabitEthernet 0/3)# switch-port trunk allow-pass vlan only 200-202
Floor(cfg-if-GigabitEthernet 0/3)# vlan-mapping-in vlan 10 remark 200
Floor(cfg-if-GigabitEthernet 0/3)# vlan-mapping-in vlan 11 remark 201
Floor(cfg-if-GigabitEthernet 0/3)# vlan-mapping-in vlan 12 remark 202
```

```
Floor(cfg-if-GigabitEthernet 0/3)# exit
```

Configure the uplink port.

```
Floor(cfg)# interface gigabitethernet 0/1
Floor(cfg-if-GigabitEthernet 0/1)# switch-port
Floor(cfg-if-GigabitEthernet 0/1)# switch-port mode uplink
Floor(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
Floor(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
Floor(cfg-if-GigabitEthernet 0/1)# end
Floor# write
```

(4) Configure the campus device.

Configure the C-Tag-based QinQ encapsulation rule.

```
Campus> privilege
Campus# config-system terminal
Campus(cfg)# vlan range 1000-1002
Campus(cfg-vlan-range)# exit
Campus(cfg)# interface gigabitethernet 0/1
Campus(cfg-if-GigabitEthernet 0/1)# switch-port
Campus(cfg-if-GigabitEthernet 0/1)# switch-port mode dot1q-tunnel
Campus(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel pvid vlan 1
Campus(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel allow-pass vlan add untagged 1,1000-1002
Campus(cfg-if-GigabitEthernet 0/1)# dot1q outer-vid 1000 register inner-vid 100,200
Campus(cfg-if-GigabitEthernet 0/1)# dot1q outer-vid 1001 register inner-vid 101,201
Campus(cfg-if-GigabitEthernet 0/1)# dot1q outer-vid 1002 register inner-vid 102,202
```

Configure the uplink port.

```
Campus(cfg)# interface gigabitethernet 0/2
Campus(cfg-if-GigabitEthernet 0/2)# switch-port
Campus(cfg-if-GigabitEthernet 0/2)# switch-port mode uplink
Campus(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
Campus(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan all

Campus(cfg-if-GigabitEthernet 0/2)# end
Campus# write
```

5. Verification

(1) Verify the configuration of HGW A.

Display the VLAN configuration of HGW A.

```
HGWA# display vlan
VLAN        Name                              Status      Ports
---------- -------------------------------- --------- -----------------------------------
         1 VLAN0001                          STATIC      Gi0/1, Gi0/5, Gi0/6, Gi0/7
                                                         Gi0/8
        10 VLAN0010                          STATIC      Gi0/1, Gi0/2
        11 VLAN0011                          STATIC      Gi0/1, Gi0/3
        12 VLAN0012                          STATIC      Gi0/1, Gi0/4
```

Display the interface configuration of HGW A.

```
HGWA# display interface switch-port
Interface                 Switch-port Mode      Access Native Protected VLAN lists
------------------------- ---------- --------- ------ ------ --------- ----------------------
GigabitEthernet 0/1       enabled    UPLINK             1      Disabled  ALL
GigabitEthernet 0/2       enabled    ACCESS     10             Disabled  ALL
GigabitEthernet 0/3       enabled    ACCESS     11             Disabled  ALL
GigabitEthernet 0/4       enabled    ACCESS     12             Disabled  ALL
```

Display the VLAN mapping rule of HGW A.

```
HGWA# display interfaces vlan-mapping
Ports           type    Status      Destination-Vlan    Source-Vlan-list
-----------     -----   ---------   ----------------    ----------------
Gi0/1           in      active      10                  100
Gi0/1           in      active      11                  101
Gi0/1           in      active      12                  102
```

(2) Verify the configuration of HGW B.

Verify that the VLAN configuration of HGW B is the same as that of HGW A.

```
HGWB# display vlan
VLAN        Name                                 Status      Ports
---------- -------------------------------- --------- -----------------------------------
         1 VLAN0001                             STATIC      Gi0/1, Gi0/5, Gi0/6, Gi0/7
                                                            Gi0/8
        10 VLAN0010                             STATIC      Gi0/1, Gi0/2
        11 VLAN0011                             STATIC      Gi0/1, Gi0/3
        12 VLAN0012                             STATIC      Gi0/1, Gi0/4
```

Verify that the interface configuration of HGW B is the same as that of HGW A.

```
HGWB# display interface switch-port
Interface                 Switch-port Mode      Access Native Protected VLAN lists
------------------------- ---------- --------- ------ ------ --------- ----------------------
GigabitEthernet 0/1       enabled    UPLINK             1      Disabled  ALL
GigabitEthernet 0/2       enabled    ACCESS     10             Disabled  ALL
GigabitEthernet 0/3       enabled    ACCESS     11             Disabled  ALL
GigabitEthernet 0/4       enabled    ACCESS     12             Disabled  ALL
```

Display the VLAN mapping rule of HGW B. The VLAN mapping rule is different from that of HGW A.

```
HGWB# display interfaces vlan-mapping
Ports           type    Status      Destination-Vlan    Source-Vlan-list
-----------     -----   ---------   ----------------    ----------------
Gi0/1           in      active      10                  200
Gi0/1           in      active      11                  201
Gi0/1           in      active      12                  202
```

(3) Verify the configuration of the floor device.

Display the VLAN configuration of the floor device.

```
Floor# display vlan
```

```
VLAN        Name                                Status      Ports
---------- -------------------------------- --------- -----------------------------------
           1 VLAN0001                          STATIC      Gi0/1, Gi0/4, Gi0/5, Gi0/6
                                                           Gi0/7, Gi0/8
   100-102 VLAN0100-VLAN0102                   STATIC      Gi0/1, Gi0/2
   200-202 VLAN0100-VLAN0102                   STATIC      Gi0/1, Gi0/3
```

Display the interface configuration of the floor device.

```
Floor# display interface switch-port
Interface                 Switch-port Mode      Access Native Protected VLAN lists
--------------------- ---------- --------- ------ ------ --------- ----------------------
GigabitEthernet 0/1    enabled    UPLINK      1        Disabled    ALL
GigabitEthernet 0/2    enabled    UPLINK      1        Disabled    100-102
GigabitEthernet 0/3    enabled    UPLINK      1        Disabled    200-202
```

Display the VLAN mapping rule of the floor device.

```
Floor# display interfaces vlan-mapping
Ports          type    Status      Destination-Vlan    Source-Vlan-list
-----------   -----   ---------   ----------------    ----------------
Gi0/2          in      active      100                  10
Gi0/2          in      active      101                  11
Gi0/2          in      active      102                  12
Gi0/3          in      active      200                  10
Gi0/3          in      active      201                  11
Gi0/3          in      active      202                  12
```

(4) Verify the configuration of the campus device.

Display the interface configuration.

```
Campus# display interfaces switch-port
Interface                          Switch-port Mode      Access Native Protected VLAN lists
--------------------------------- ---------- --------- ------ ------ --------- ----------------
GigabitEthernet 0/1                enabled    TUNNEL      1        Disabled    1,1000-1002
GigabitEthernet 0/2                enabled    UPLINK      1        Disabled    ALL
```

Display the dot1q-tunnel configuration. The outer tag VLAN has been added to the allowed VLAN list of the interface.

```
Campus# display interfaces dot1q-tunnel
========Interface Gi0/1========
Native vlan: 1
Allowed vlan list:1,1000-1002
Tagged vlan list:
```

Display the mapping policy relationship for adding S-tags based on C-tags.

```
Campus# display registration-table
Ports       Type              Outer-VID      Inner-VID-list
------      ----------        ----------     --------------
Gi0/1       Add-outer         1000           100,200
Gi0/1       Add-outer         1001           101,201
```

```
Gi0/1       Add-outer       1002            102,202
```

6. Configuration Files

- Configuration file of HGW A

```
vlan range 1,10-12
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode uplink
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094vlan-mapping-in vlan 100 remark 10
 vlan-mapping-in vlan 101 remark 11
 vlan-mapping-in vlan 102 remark 12
!
interface GigabitEthernet 0/2
switch-port
   switch-port mode access
 switch-port access vlan 10
!
interface GigabitEthernet 0/3
switch-port
 switch-port mode access
 switch-port access vlan 11
!
interface GigabitEthernet 0/4
switch-port
 switch-port mode access
 switch-port access vlan 12
```

- Configuration file of HGW B

```
vlan range 1,10-12
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode uplink
   switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 vlan-mapping-in vlan 200 remark 10
 vlan-mapping-in vlan 201 remark 11
 vlan-mapping-in vlan 202 remark 12
!
 interface GigabitEthernet 0/2
 switch-port
 switch-port mode access
 switch-port access vlan 10
!
```

```
 interface GigabitEthernet 0/3
   switch-port
 switch-port mode access
 switch-port access vlan 11
!
 interface GigabitEthernet 0/4
   switch-port
 switch-port mode access
 switch-port access vlan 12
```

- Configuration file of the floor device

```
vlan range 1,100-102,200-202
!
interface GigabitEthernet 0/1
switch-port
    switch-port mode uplink
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode uplink
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 100-102
 vlan-mapping-in vlan 10 remark 100
 vlan-mapping-in vlan 11 remark 101
 vlan-mapping-in vlan 12 remark 102
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode uplink
switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 200-202
 vlan-mapping-in vlan 10 remark 200
 vlan-mapping-in vlan 11 remark 201
 vlan-mapping-in vlan 12 remark 202
```

- Configuration file of the campus device

```
vlan range 1,1000-1002
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode dot1q-tunnel
 switch-port dot1q-tunnel allow-pass vlan add untagged 1,1000-1002
 switch-port dot1q-tunnel pvid vlan 1dot1q outer-vid 1000 register inner-vid 100,200
 dot1q outer-vid 1001 register inner-vid 101,201
 dot1q outer-vid 1002 register inner-vid 102,202
```

```
!
interface GigabitEthernet 0/2
 switch-port
switch-port mode uplink
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

1.15.5 Configuring Layer 2 Protocol Tunneling Based on QinQ

1. Requirements

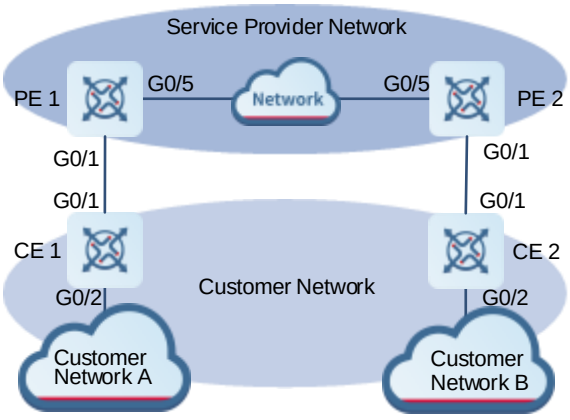
As shown in Figure 1-13, the upper part of the typical topology of QinQ network is the SP network, and the lower part is the client network. The SP network consists of PE 1 and PE 2. Customer Network A and Customer Network B are a user's two sites in different regions. CE 1 and CE 2 are access devices from the client network to the SP network, and access the SP network through PE 1 and PE 2 respectively.

The layer-2 protocol tunneling is configured to transparently transmit layer-2 protocol packets from one end to the other end of the client network so that the layer-2 protocols of the client network and the SP network are calculated separately without interference.

- Configure the STP protocol tunneling function based on user requirements to realize tunnel transmission of BPDU packets of the client network in the SP network and to perform unified spanning tree calculation for the cross-regional client network across the SP network.
- Configure the GVRP protocol tunneling function based on user requirements to realize tunnel transmission of GVRP packets of the client network and perform dynamic VLAN configuration of the client network across the SP network.

2. Topology

Figure 1-13 Configuring QinQ for Layer 2 Protocol Tunneling



3. Notes

- On PE 1 and PE 2, configure the STP and GVRP tunneling functions globally.
- Configure the QinQ encapsulation mode. Basic QinQ is taken as an example (for details, see "1.15.1 Configuring Basic QinQ to Implement Layer-2 VPN") to configure the native VLAN of dot1q-tunnel as VLAN 10, and VLAN 10 is used as the public network transmission channel. A client data packet is first encapsulated with one VLAN 10 tag, and then forwarded on the public network.

- Enable the STP and GVRP tunneling functions on the interface GigabitEthernet 0/1 of PE 1 and PE 2.
- Configure the uplink port of PE 1 and PE 2.

4. Procedure

Both PE 1 and PE 2 need to be configured. The configuration method on the two devices is the same. PE 1 is taken as an example below.

(1) Enable the STP and GVRP tunneling functions globally.

```
PE1> privilege
PE1# config-system terminal
PE1(cfg)# l2protocol-tunnel stp
PE1(cfg)# l2protocol-tunnel gvrp
```

(2) Configure the basic QinQ transmission channel VLAN 10. VLAN 10 is the native VLAN of the dot1q-tunnel, and belongs to the allowed untagged VLAN list of the dot1q-tunnel.

```
PE1(cfg)# vlan 10
PE1(cfg-vlan)# exit
PE1(cfg)# interface gigabitethernet 0/1
PE1(cfg-if-GigabitEthernet 0/1)# switch-port
PE1(cfg-if-GigabitEthernet 0/1)# switch-port mode dot1q-tunnel
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel pvid vlan 10
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel allow-pass vlan add untagged 10
```

(3) Enable the STP and GVRP tunneling functions on the dot1q-tunnel.

```
PE1(cfg-if-GigabitEthernet 0/1)# l2protocol-tunnel stp enable
PE1(cfg-if-GigabitEthernet 0/1)# l2protocol-tunnel gvrp enable
PE1(cfg-if-GigabitEthernet 0/1)# exit
```

(4) Configure the uplink port.

```
PE1(cfg)# interface gigabitethernet 0/5
PE1(cfg-if-GigabitEthernet 0/5)# switch-port
PE1(cfg-if-GigabitEthernet 0/5)# switch-port mode uplink
PE1(cfg-if-GigabitEthernet 0/5)# switch-port trunk pvid vlan 1
PE1(cfg-if-GigabitEthernet 0/5)# switch-port trunk allow-pass vlan all
PE1(cfg-if-GigabitEthernet 0/5)# end
PE1# write
```

5. Verification

Verify that the STP tunneling function is enabled globally and on the interface at the same time. c

```
PE1# display l2protocol-tunnel stp
L2protocol-tunnel: Stp Enable
L2protocol-tunnel destination mac address: 01d0.f800.0005
GigabitEthernet 0/1 l2protocol-tunnel stp enable
```

Verify that the GVRP tunneling function is enabled globally and on the interface at the same time. The verification methods on PE1 and PE2 are the same. PE1 is taken as an example.

Verify PE 1.

```
PE1# display l2protocol-tunnel gvrp
```

```
L2protocol-tunnel: Gvrp Enable
L2protocol-tunnel destination mac address: 01d0.f800.0006
GigabitEthernet 0/1 l2protocol-tunnel gvrp enable
```

6. Configuration Files

Configuration files of PE 1 and PE 2

```
l2protocol-tunnel stp
l2protocol-tunnel gvrp
!
vlan range 1,10
!
interface GigabitEthernet 0/1
   switch-port
  switch-port mode dot1q-tunnel
  switch-port dot1q-tunnel allow-pass vlan add untagged 10
  switch-port dot1q-tunnel pvid vlan 10
  l2protocol-tunnel stp enable
  l2protocol-tunnel gvrp enable
!
interface GigabitEthernet 0/5
  switch-port
switch-port mode uplink
switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

7. Common Errors

- The layer-2 tunneling function is not enabled globally and on the interface at the same time, and therefore it fails to take effect.
- In the SP network, the configured BPDU tunnel addresses are inconsistent, and therefore BPDU frames of the client network cannot be correctly transmitted.

Contents

1 Configuring MSTP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Background and Functions 1
1.1.3 STP 2
1.1.4 RSTP 12
1.1.5 MSTP 17
1.1.6 Protocols and Standards 26
1.2 Restrictions and Guidelines 27
1.3 Configuration Task Summary 27
1.3.1 Configuring STP 27
1.3.2 Configuring RSTP 27
1.3.3 Configuring MSTP 27
1.3.4 Configuring Advanced Functions for Spanning Trees 28
1.3.5 Configuring BPDU Transparent Transmission 28
1.4 Notes 29
1.5 Configuring the Spanning Tree Mode 29
1.5.1 Overview 29
1.5.2 Restrictions and Guidelines 29
1.5.3 Procedure 29
1.6 Configuring an MST Region 30
1.6.1 Overview 30

1.6.2 Restrictions and Guidelines 30
1.6.3 Prerequisites 30
1.6.4 Procedure 30
1.7 Configuring the Bridge Priority and Port Priority 31
1.7.1 Overview 31
1.7.2 Restrictions and Guidelines 31
1.7.3 Prerequisites 32
1.7.4 Procedure 32
1.8 Configuring the Port Path Cost 32
1.8.1 Overview 32
1.8.2 Configuration Tasks 32
1.8.3 Configuring the Path Cost Calculation Method 33
1.8.4 Configuring the Path Cost Value 34
1.9 Configuring Spanning Tree Time Parameters 35
1.9.1 Overview 35
1.9.2 Restrictions and Guidelines 35
1.9.3 Procedure 36
1.10 Configuring the Maximum Hop Count for BPDUs 36
1.10.1 Overview 36
1.10.2 Restrictions and Guidelines 36
1.10.3 Procedure 37
1.11 Configuring Fast Convergence of Spanning Trees 37
1.11.1 Overview 37
1.11.2 Restrictions and Guidelines 37

1.11.3 Prerequisites 37
1.11.4 Procedure 37
1.12 Enabling Protocol Migration 38
1.12.1 Overview 38
1.12.2 Procedure 39
1.13 Configuring Spanning Tree Compatibility for Interfaces 39
1.13.1 Overview 39
1.13.2 Prerequisites 39
1.13.3 Procedure 39
1.14 Enabling the STP Function 39
1.14.1 Overview 39
1.14.2 Restrictions and Guidelines 40
1.14.3 Procedure 40
1.15 Enabling the V-STP Function 40
1.15.1 Overview 40
1.15.2 Restrictions and Guidelines 42
1.15.3 aggregate interfaceProcedure 42
1.16 Configuring Port Protection 42
1.16.1 Overview 42
1.16.2 Restrictions and Guidelines 43
1.16.3 Configuration Tasks 43
1.16.4 Configuring Root Guard 43
1.16.5 Configuring Loop Guard 44
1.17 Configuring BPDU Source MAC Address Check 45

1.17.1 Overview 45
1.17.2 Restrictions and Guidelines 46
1.17.3 Prerequisites 46
1.17.4 Procedure 46
1.18 Configuring an Edge Port 47
1.18.1 Overview 47
1.18.2 Restrictions and Guidelines 47
1.18.3 Configuration Tasks 47
1.18.4 Configuring Autoedge 47
1.18.5 Configuring the Port Fast Attribute 48
1.19 Configuring BPDU Guard or BPDU Filter 50
1.19.1 Overview 50
1.19.2 Restrictions and Guidelines 50
1.19.3 Configuration Tasks 50
1.19.4 Configuring BPDU Guard 50
1.19.5 Configuring BPDU Filter 51
1.20 Configuring TC Attack Defense Functions 52
1.20.1 Overview 52
1.20.2 Restrictions and Guidelines 53
1.20.3 Configuration Tasks 53
1.20.4 Configuring TC Protection 53
1.20.5 Configuring TC Guard 54
1.20.6 Configuring TC Filter 55
1.21 Configuring BPDU Tunnel 56

1.21.1 Overview 56
1.21.2 Restrictions and Guidelines 57
1.21.3 Prerequisites 57
1.21.4 Procedure 57
1.22 Configuring BPDU Transparent Transmission 58
1.22.1 Overview 58
1.22.2 Restrictions and Guidelines 58
1.22.3 Procedure 58
1.23 Monitoring 58
1.24 Configuration Examples 60
1.24.1 Configuring MSTP 60
1.24.2 Configuring MSTP+VRRP 76
1.24.3 Configuring Spanning Tree Compatibility for MSTP Interfaces 91
1.24.4 Configuring BPDU Tunnel 98
1.24.5 Configuring BPDU Transparent Transmission 106
1.24.6 Configuring V-STP 110

1 Configuring MSTP

1.1 Introduction

1.1.1 Overview

The Spanning Tree Protocol (STP) is a Layer 2 management protocol used to eliminate Layer 2 loops by blocking redundant links in the network. It enables a backup link in the case of a link failure.

With the network development and update, multiple STP versions become available. This device supports STP, Rapid Spanning Tree Protocol (RSTP), and Multiple Spanning Tree Protocol (MSTP).

1.1.2 Background and Functions

1. 802.1D STP

To prevent single point of failures (SPOFs), Layer 2 devices enhance line reliability by adding redundant lines. One of the most economical ways is to add loops. However, loops may cause broadcast storms and Media Access Control (MAC) address flapping. Therefore, the STP protocol arises.

STP can generate the best loop-free tree topology for a local area network (LAN) via computation. If detecting a fault, STP automatically eliminates the fault and updates the topology.

STP cannot enter forwarding state quickly but needs 50 seconds to complete topology convergence. All virtual local area networks (VLANs) in an LAN share one spanning tree. Packets of all the VLANs are forwarded along this spanning tree, and the bandwidth of redundant lines is wasted. VLAN data may be isolated at Layer 2, and Layer 2 loops do not exist in some circumstances. However, STP cannot recognize VLANs, and packets from interfaces in the same VLAN may be intercepted by blocking ports, resulting in a communication failure.

2. 802.1w RSTP

RSTP is compatible with STP backward. Apart from the conventional STP functions of eliminating loops and enabling backup links, RSTP solves the problem of entering forwarding state rapidly. If all bridges in a LAN support RSTP and are properly configured, once the network topology changes, RSTP takes only less than 1s to regenerate a topology tree.

Although RSTP can complete convergence quickly, it cannot identify VLANs just as STP does not.

3. 802.1s MSTP

MSTP resolves defects of STP and RSTP. It not only supports fast convergence but also forwards traffic of different VLANs along their respective paths, thereby providing a better load balancing mechanism for redundant links. STP/RSTP works based on ports while MSTP works based on instances. An instance is a collection of multiple VLANs. Binding multiple VLANs to one instance can reduce communication costs and improve resource usage. MSTP can implement the following functions:

- When lines are normal, MSTP can eliminate loops and prevent broadcast storms and MAC address flapping, to ensure normal Layer 2 communication.
- When a line is faulty, MSTP enables a backup link to enhance the link reliability. It can complete convergence rapidly to enhance the link availability.

- MSTP can generate multiple spanning trees by VLAN instance to implement traffic balancing.

1.1.3 STP

1. Overview

STP, defined by the Institute of Electrical and Electronics Engineers (IEEE) 802.1D standard, is used to break physical loops at the data link layer in a LAN and prevent broadcast storms, so as to implement link redundancy.

Layer 2 devices connected using single lines have no redundant lines and devices. Any SPOF can cause device disconnection and the line reliability is low. When a redundant line is added, a Layer 2 loop will form and the SPOF can be eliminated.

The adding of a redundant line brings two new problems:

- Once a broadcast frame appears in a Layer 2 loop, it will be flooded continuously by Layer 2 devices and a broadcast storm will be incurred, which will heavily consume device resources and network bandwidth.
- In a loop topology, a Layer 2 device may receive packets from the same source MAC address through two ports of two paths, causing MAC address flapping and the failure of the device to work properly.

Therefore, for Layer 2 Ethernet, only one valid channel is required to forward data between two LANs. Ideally, when a Layer 2 loop exists on a network, a specific port can be blocked to invalidate the redundant link so as to break the loop; when a network fault occurs, the specific port can be enabled to restore a redundant link to the valid state. It is difficult to attain this ideal effect via manual control. Therefore, the STP protocol emerges to automatically complete this work:

- Devices with STP enabled exchange specific STP Protocol Packets, that is, bridge protocol data units (BPDUs), and elect Device Roles (root bridge and non-root bridge) and Port Roles (root port, designated port, and blocking port) based on Key Information of BPDUs. Different port roles are in different Port States and have different functions. They form a Generating a Spanning Tree Topology in a LAN.
- Protocol packets are used to monitor the line status to Maintaining a Spanning Tree Topology. Once a fault is found, STP rectifies the fault and automatically Updating a Spanning Tree Topology.

2. STP Protocol Packets

STP uses two types of protocol packets: configuration BPDU and topology change notification (TCN) BPDU.

- Configuration BPDU: A configuration BPDU contains 35 bytes, which provide information required for calculating the spanning tree topology (root bridge ID, root path cost, designated bridge ID, and designated port ID) and information required for maintaining the spanning tree (**Message Age**, **Max Age**, **Hello Time**, and **Forward Delay**). When STP is enabled on a designated port, the port sends configuration BPDUs at an interval of **Hello Time**. When a device receives a configuration BPDU from a root port, it sends the configuration BPDU to each designated port on the device, and the designated ports generate their own BPDUs. If the configuration BPDU received by a designated port from the line has a lower priority than the BPDU generated by the designated port itself (see 1.1.3 3. Key Information of BPDUs), the designated port sends its BPDU to the downstream device.
- TCN BPDU: A TCN BPDU contains four bytes. On a bridge with at least one designated port, when a port on the bridge transitions to forwarding or blocking state, the network topology has changed. The bridge will send a TCN BPDU to inform the upstream device of the network topology change. When the upstream device receives the TCN BPDU through a designated port, it sets TCA in the configuration BPDU to 1 and instructs the downstream device to stop sending the TCN BPDU. The upstream device also sends a copy of

the TCN BPDU to its upstream device until the TCN BPDU reaches the root bridge. By setting the TC flag bit in the configuration BPDU to 1, the root bridge informs all bridges in the spanning tree of the network topology change, and asks them to clear dynamic MAC address entries of ports and relearn MAC addresses. A shorter MAC address aging time facilitates fast convergence of the topology.

BPDUs are encapsulated in IEEE 802.3 logical link control (LLC) Ethernet frames. The format differences between BPDUs and other Ethernet frames are as follows.

Figure 1-1 Formats of Ethernet Frames

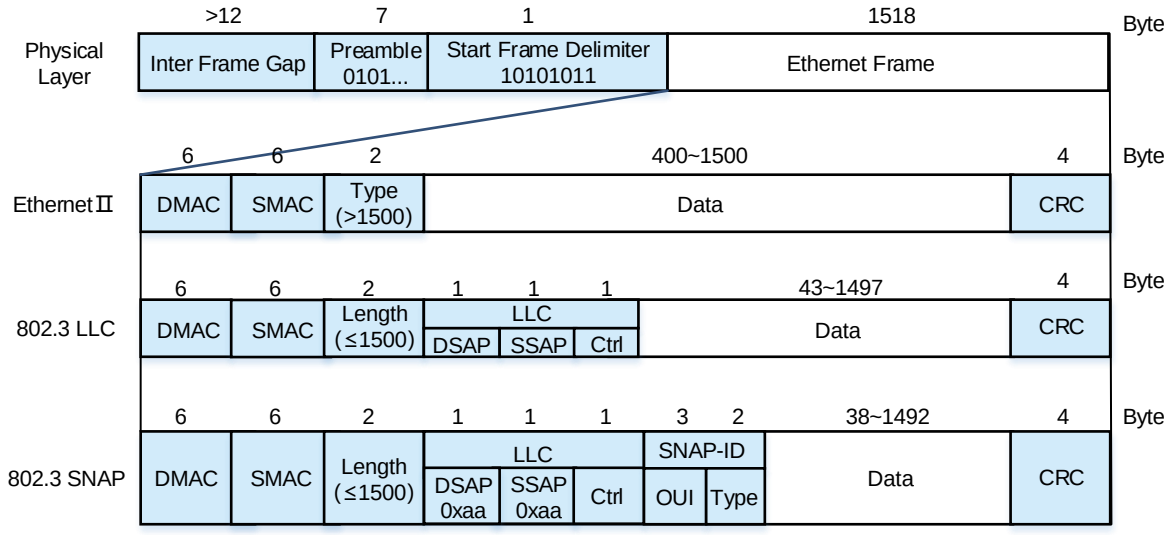


Figure 1-2 is an actual STP configuration BPDU taken as an example.

Figure 1-2 STP Configuration BPDU

```
0000  01 80 c2 00 00 00 00 d0  f8 22 35 4a 00 26 42 42
0010  03 00 00 00 00 00 80 00  00 d0 f8 22 35 4a 00 00
0020  00 00 80 00 00 d0 f8 22  35 4a 80 1c 00 00 14 00
0030  02 00 0f 00 00 00 00 00  00 00 00 00
```

The header of an IEEE 802.3 Ethernet frame contains 14 bytes in three fields in the header, as described in Table 1-1.

Table 1-1 Description of Fields in the Header of an IEEE 802.3 Ethernet Frame

Field	Bytes	Description
DMAC	6	Destination MAC address. According to the protocol, the destination address of the Ethernet frame used for encapsulating a BPDU is fixed to the multicast address 0180.c200.0000. A device receiving a frame with such a destination address will not forward the frame.
SMAC	6	Source MAC address of the device sending the BPDU. The value in the example is 00-d0-f8-22-35-4a.

Field	Bytes	Description
Length	2	The value is 0x0026, indicating that the packet length is 38 bytes (3 bytes of the LLC header and 35 bytes of the BPDU).

The LLC header occupies three bytes in three fields, as described in Table 1-2.

Table 1-2 Description of Fields in the LLC Header of an Ethernet Frame

Field	Bytes	Description
DSAP	1	Destination service access point. The value is 0x42.
SSAP	1	Source service access point. The value is 0x42.
Control	1	This field stores various control information and identifies the type of an LLC frame. The value is 0x03.

The bytes behind the LLC header is a BPDU. There are four types of BPDUs for STP, RSTP, and MSTP, which are differentiated using different PVIDs and BPDU type values. A configuration BPDU of STP occupies 35 bytes in 12 fields, as described in Table 1-3. The last part is an 8-byte filling field and a 4-byte CRC field. They form a 64-byte Ethernet frame.

Table 1-3 Format of an STP Configuration BPDU

Field	Bytes	Description
PID	2	Protocol identifier. The value is 0x0000.
PVID	1	Protocol version identifier. 0x00 indicates an STP packet.
BPDU Type	1	BPDU packet type. 0x00 indicates a configuration BPDU of STP.
Flags	1	The value is 0x00 in the example. Configuration BPDUs of STP use only bit 7 and bit 0 and other bits are not used. ● Bit 7: Topology change acknowledgment (TCA). An upstream device sets bit 7 in a configuration BPDU to 1 and sends the BPDU to notify the downstream device of the topology change and requests the downstream device to stop sending the TCN BPDU. ● Bit 0: Topology change (TC). The root bridge sets bit 0 in a configuration BPDU to 1 and sends the BPDU to instruct the downstream device to directly delete MAC address entries of the bridges.
Root ID	8	Identifier of the root bridge recognized by the device. It consists of a 2-byte bridge priority and a 6-byte bridge MAC address. In the example, the priority value is 0x8000, which indicates 32768, and the bridge MAC address is 00-d0-f8-22-35-4a.
RPC	4	Root path cost, which is the cumulative cost from the port to the root bridge. The value is 0x00-00-00-00 in the example.

Field	Bytes	Description
Bridge ID	8	Bridge identifier of the device. It consists of a 2-byte bridge priority and a 6-byte bridge MAC address. In the example, the priority value is 0x8000, which indicates 32768, and the bridge MAC address is 00-d0-f8-22-35-4a.
Port ID	2	Identifier of the port that sends the BPDU. It consists of a 1-byte priority and 1-byte port ID. In the example, the priority value is 0x80, which indicates 128, and the port ID is 1c.
Message Age	2	Age of the BPDU, which represents the time that the BPDU has been alive in the network. The value is 0x0000 in the example.
Max Age	2	Aging time of the BPDU. If a non-root bridge fails to receive a BPDU from the root bridge within the time of **Max Age**, the root bridge or the link to the root bridge is deemed as faulty. The value in the example is 0x1400 (low-order bytes on the left), indicating 20 seconds.
Hello Time	2	Interval for sending two adjacent BPDUs. The root bridge sends configuration BPDUs at an interval of **Hello Timer**. A non-root switch sends a configuration BPDU only after receiving the configuration BPDU from the upstream device. The value in the example is 0x0200 (low-order bytes on the left), indicating 2 seconds.
FWD Delay	2	Forwarding delay. This time parameter controls the duration of listening and learning states. The value in the example is 0x0f00 (low-order bytes on the left), indicating 15 seconds.

A TCN BPDU consists of four bytes and contains only the PID, PVID, and BPDU Type fields. The packet format is described in Table 1-4.

Table 1-4 Format of an STP TCN BPDU

Field	Bytes	Description
PID	2	Protocol identifier. The value is 0x0000.
PVID	1	Protocol version identifier. 0x00 indicates an STP packet.
BPDU Type	1	BPDU packet type. 0x80 indicates a TCN BPDU of STP. When the downstream network topology changes, it is used to notify the upstream device of the network topology change. After receiving a TCN BPDU, a port copies the BPDU and forwards it to upstream devices till the root device receives the BPDU.

3. Key Information of BPDUs

- Priority vector

Bridges exchange BPDUs to obtain information and generate a stable optimal tree topology. The most basic elements in the information are listed as follows:

- Root ID: ID of the root bridge recognized by this local bridge. It consists of the bridge priority and MAC address of the root. The root bridge ID is the bridge ID of the root.
- Root Path Cost: Path cost from this bridge to the root bridge.
- Bridge ID: Each bridge has a unique bridge ID consisting of the bridge priority and MAC address.
- Port ID: ID of each port, which consists of the port priority and port ID.

These key elements constitute a priority vector <**Root ID**, **Root Path Cost**, **Bridge ID**, **Port ID**>. Bridges compare elements in the priority vector in sequence to determine the priority of BPDUs. A smaller priority element value indicates a higher priority. If a port of a bridge receives a BPDU with the priority higher than the BPDU priority of the port, the port saves and transmits the BPDU. If it receives a BPDU with the priority lower than the BPDU priority of the port, in case of STP, it discards the BPDU, but in case of RSTP or MSTP, it will send a Proposal BPDU to request the peer to recognize the local device as the upstream device. This mechanism ensures that high-priority BPDUs are transmitted throughout the network.

- Bridge ID

  According to IEEE 802.1W, a root ID is used to identify a root bridge in a spanning tree while a bridge ID is used to identify each bridge. When a device just runs STP, it uses its bridge ID as the root ID and STP compares root IDs to elect the root bridge according to algorithms. A bridge ID consists of eight bytes, with the first two bytes of the bridge priority and the last six bytes of the bridge MAC address. A device compares the bridge priority first and then the MAC address during root bridge selection. A smaller value indicates a higher priority.

  Bridge priority: To ensure network reliability, you are advised to specify bridge priorities during network planning, to prevent devices from electing the root bridge by comparing MAC addresses. In the 2-byte bridge priority, bit 1 to bit 12 indicate the system ID used for protocol extension and the value is **0** in RSTP. Bit 13 to bit 16 indicate the priority, which can be configured. If the configured priority is 0001 0000 0000 0000 (**4096** in decimal notation), the bridge priority is a multiple of 4096. Bit 13 to bit 16 can indicate 15 numbers from 0000 to 1111. Therefore, the priority is 4096 × 0 to 4096 × 15 in decimal notation.

Table 1-5 Bridge Priority Field (Two Bytes)

	Priority				**System ID**											
Bit	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
Value	32768	16384	8192	4096	2048	1024	512	256	128	64	32	16	8	4	2	1
-	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0

- Port ID

  Each port participating in the spanning tree calculation must have a port ID, which is the port ID of a device.

  Format: A port ID consists of two bytes, with the first byte of the port priority and the second byte of the port ID.

  The default port priority is **128** and the port priority range is from 0 to 255. The port ID range is from 0 to 255, for example, the ID (0) of port GigabitEthernet 0/0 is smaller than that (1) of port GigabitEthernet 0/1.

- Root path cost

  Root path cost is the cumulative path cost from the port to the root bridge. A link between two directly connected ports is called a "segment". A segment with a higher network bandwidth has a lower path cost. The total cost calculated based on the path costs of all ports in the necessary lines from a port to the root is the cumulative path cost from the port to the root bridge. Table 1-6 shows the cost values supported by different protocols. Path cost values defined for the device are provided in Table 1-12.

Table 1-6 Path Costs Supported by Protocols

Bandwidth	RPC Supported by IEEE 802.1t	RPC Supported by IEEE 802.1d
10 Mbps	2000000	100
100 Mbps	200000	19
1 Gbps	20000	4
10 Gbps	2000	2

4. Device Roles

Devices with STP enabled exchange BPDUs and compare root IDs to elect device roles.

- Root bridge: Each spanning tree has a unique root.
- Designated bridge: A device in the shortest path between the local device and the root bridge, that is, the upstream device of the local device.
- Bridge: Generally the local device, that is, downstream device of a designated bridge. Designated bridges and bridges can be collectively referred to as non-root bridges.

5. Port Roles

The root bridge and non-root bridges elect different port roles by exchanging BPDUs.

- Root port: A port with the shortest path from a non-root bridge to the root bridge.
- Designated port: Each port on the root bridge is a designated port. The root bridge connects to a root port of a non-root bridge through a designated port. The port that connects a designated bridge to a downstream device is also a designated port.
- Disable port: Also called a blocking port. A disable port is a port in blocked or down state on a non-root bridge.

6. Port States

Ports work in different states and perform different functions according to the protocol. Table 1-7 lists functions corresponding to port states.

STP defines five port states: disabled, blocking, listening, learning, and forwarding. RSTP reduces port states to only three states: discarding, learning, and forwarding. MSTP uses the port states of RSTP.

Table 1-7 Port Roles, Port States, and Functions Supported by STP, RSTP, and MSTP

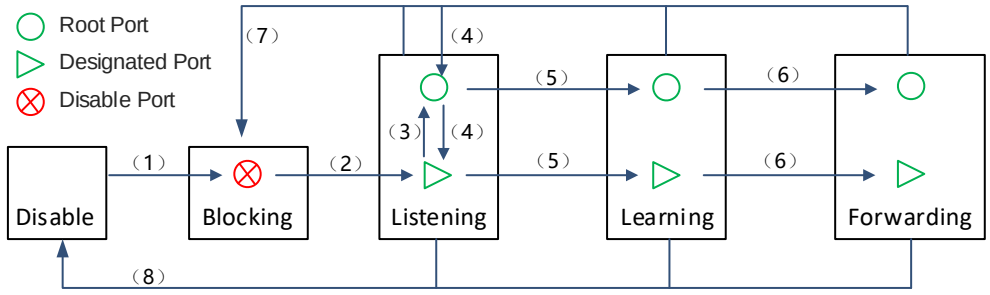
Note: In this table, R = root port, D = designated port, A = alternate port, B = backup port, BLK = blocking port.

STP		RSTP and MSTP		Function				
Port Role	**Port State**	**Port Role**	**Port State**	**Receiving BPDUs**	**Sending BPDUs**	**MAC Address Learning**	**Receiving Data**	**Forwarding Data**
Disable	Disabled	Disable	Discarding	No	No	No	No	No
Blk	Blocking	A, B	Discarding	Blk, A, B	No	No	No	No
R, D	Listening	R, D	Discarding	R, D	D	No	No	No
	Learning		Learning	R, D	D	R, D	No	No
	Forwarding		Forwarding	R, D	D	R, D	R, D	R, D

After a topology becomes stable, only root ports and designated ports are in forwarding state and can send and receive data. Other ports are in blocking state and can only receive BPDUs. Different port roles have different functions in the same port state. For example, a root port cannot send but can receive BPDUs in forwarding state, while a designated port can send and receive BPDUs in forwarding state.

The port states can be transitioned according to the rules shown in Figure 1-3.

Figure 1-3 State Transition Rules of STP Ports



(1) After being initialized or enabled, a port enters blocking state.

(2) When STP is just running on a device, all the ports on the device enter listening state as designated ports, and send and receive BPDUs for election. After STP runs stably, if a blocking port fails to receive a BPDU within **MAX Age**, it becomes a designated port, transitions from blocking state to listening state, and receives and sends BPDUs for re-election.

(3) After election, if a port is elected as a root port, it only receives BPDUs but cannot send BPDUs.

(4) If a root port fails to receive a BPDU within **MAX Age**, the root is considered to be faulty, and the root port becomes a designated port and sends a BPDU.

(5) After a port is elected as a root port or designated port, it waits for one period of **Forward Delay**, transitions from listening state to learning state, and starts MAC address learning.

(6) Then, the port waits for another period of **Forward Delay**, transitions from learning state to forwarding state, and starts sending and receiving BPDUs.

(7) If the port is no longer a root port or designated port, it returns to blocking state.

(8) If a port is disabled or the link is faulty, the port returns to disable state.

7. Generating a Spanning Tree Topology

STP automatically calculates and generates a spanning tree topology for a LAN according to a set of bridge parameters configured by the user. Only appropriate configuration can ensure the fast generation of a stable topology. The steps for STP to generate a spanning tree topology are as follows:

(1) When a network starts to run STP, all switches consider themselves as root bridges and all ports on the switches are considered as designated ports. So, all ports enter listening state, broadcast configuration BPDUs, and receive BPDUs from the network. At this time, the root ID in the BPDU of each switch is the bridge ID of the switch.

(2) Each broadcast domain selects one root bridge. Each switch receives BPDUs from other switches and each BPDU contains the priority vector <**Root ID**, **Root Path Cost**, **Bridge ID**, **Port ID**>. According to the rule that a smaller value indicates a higher priority, each switch compares root IDs (which is the bridge ID of each switch at first) in BPDUs. When a port of a switch receives a smaller root ID, the root ID is updated in the **Root ID** field in the configuration BPDUs stored in all ports of the switch and then sent to other switches. After interaction, the bridge ID of the device with the highest priority is recorded as the root ID, indicating that the device is elected as the root bridge.

(3) All ports on the root bridge are designated ports. After waiting for one period of **Forward Delay**, designated ports transition from listening state to learning state to learn MAC addresses. After waiting for another period of **Forward Delay**, the designated ports transition from learning state to forwarding state and start receiving and sending packets.

(4) Each non-root bridge compares the cumulative path cost from each port to the root and elects the port with the smallest cost as the root port. Root ports only receive configuration BPDUs from designated ports and do not send configuration BPDUs. After waiting for one period of **Forward Delay**, root ports transition from listening state to learning state to learn MAC addresses. After waiting for another period of **Forward Delay**, root ports transition from learning state to forwarding state and start sending and receiving packets.

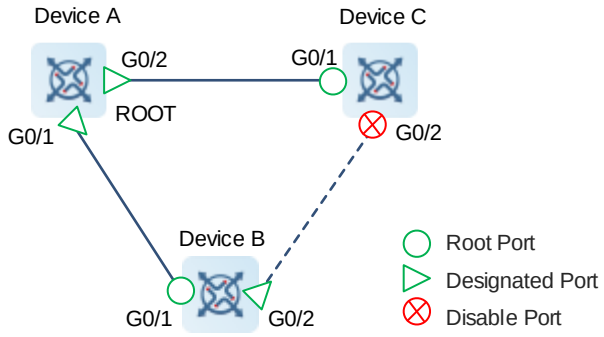
(5) A link between two directly connected switches is called a "segment". The priority vector <**Root ID**, **Root Path Cost**, **Bridge ID**, **Port ID**> is compared in sequence to select a designated port for each segment. After waiting for one period of **Forward Delay**, designated ports transition from listening state to learning state to learn MAC addresses. After waiting for another period of **Forward Delay**, the designated ports transition from learning state to forwarding state and start receiving and sending packets.

(6) The other ports are non-designated ports and are blocked. Such ports return to blocking state and only receive BPDUs. If a blocking port receives a high-priority BPDU within **MAX Age**, the current line from the peer device to the root is better and the blocking port remains in blocking state. If a blocking port fails to receive a high-priority BPDU within **MAX Age**, the current line from the peer device to the root is worse or even faulty. The blocking port transitions to listening state and sends its configuration BPDU to inform the peer device that the communication line passing through this port is better.

Note

- In STP and RSTP, the same device has the same bridge ID and the same port has the same port ID.
- In MSTP, different bridge IDs can be configured for the same device and different port IDs can be configured for the same port by instance.

Figure 1-4 Basic Topology of STP



8. Maintaining a Spanning Tree Topology

After a spanning tree topology becomes stable, only designated ports and root ports are in forwarding state while other ports are in blocking state. The root bridge continues to send configuration BPDUs at an interval of **Hello Time**. A downstream non-root bridge stores the optimal configuration BPDU received from the root port in the device, fills the bridge ID of the root bridge in the root ID, and then sends the configuration BPDU to the downstream device through a designated port. If a downstream device receives a higher-priority BPDU from the upstream device within one period of **Max Age**, the network topology remains unchanged.

9. Updating a Spanning Tree Topology

Taking the topology shown in Figure 1-4 as an example, this section describes the update process of the spanning tree topology. The priority vector in the configuration BPDU packet sent by the interface is <RID, Cost, BID, PID>, and that saved by the interface is (RID, Cost, BID, PID). BIDs of Device A, Device B, and Device C are 1, 2, and 3 respectively. PIDs of ports GigabitEthernet 0/1 and GigabitEthernet 0/2 are 1 and 2 respectively. The lines are 1000M links and the path cost of each segment of line is 4.

- Device B and Device C consider themselves as root devices if the following situation occurs: Root bridge A malfunctions, Device B fails to receive BPDU<1,0,1,1> from the root bridge, and Device C fails to receive BPDU<1,0,1,2> from the root bridge within one period of **Max Age**. Then, port GigabitEthernet 0/2 of Device B sends BPDU<2,0,2,2>. Port GigabitEthernet 0/2 of Device C becomes a designated port and sends BPDU<3,0,3,2>. <2,0,2,2> has a higher priority than <3,0,3,2>. Therefore, Device B is elected as the root bridge and port GigabitEthernet 0/2 of Device C becomes a root port.
- Device C deems that root bridge A is faulty and considers itself as the root bridge if the following situation occurs: Port GigabitEthernet 0/2 of root bridge A or the link from root bridge A to Device C is faulty and downstream Device C fails to receive BPDU<1,0,1,2> from root bridge A through the root port GigabitEthernet 0/1 within one period of **Max Age**. Blocking port GigabitEthernet 0/2 (that can only receive BPDUs but cannot send BPDUs) of Device C becomes a designated port, transitions from blocking state to listening port, and sends BPDU<3,0,3,2>. Port GigabitEthernet 0/2 of Device C receives BPDU<1,4,2,2> from designated port GigabitEthernet 0/2 of Device B and finds that the root ID has a higher priority than the root ID in Device C. Device C considers Device A as the root bridge again, updates the spanning tree topology, and sets port GigabitEthernet 0/2 as the root port. After waiting for one period of **Forward Delay**, GigabitEthernet 0/2 transitions to learning state to learn MAC addresses. After waiting for another period of **Forward Delay**, GigabitEthernet 0/2 transitions to forwarding state to forward data, implementing link auto-recovery. In this case, the link recovery needs **Max Age** + 2 × FWD time. The default link recovery time

is 50s, that is, 20 + 2 × 15 = 50.

- Device B deems that root bridge A is faulty and considers itself as the root bridge if the following situation occurs: Port GigabitEthernet 0/1 of Device A or the link from root bridge A to Device B is faulty and downstream Device B fails to receive BPDU<1,0,1,1> from root bridge A through the root port GigabitEthernet 0/1 within one period of **Max Age**. The BPDU sent by port GigabitEthernet 0/2 of Device B changes from <1,4,2,2> (Device A functions properly) to <2,0,2,2> (Device A is faulty). Blocking port GigabitEthernet 0/2 of Device C finds that the BPDU sent from port GigabitEthernet 0/2 of Device B has a lower priority than the configuration BPDU<1,4,3,2> stored by the blocking port itself, and discards the BPDU sent from Device B. If blocking port GigabitEthernet 0/2 of Device C fails to receive a BPDU with the priority higher than that of the BPDU stored in the blocking port within one period of **Max Age**, the blocking port becomes a designated port and sends its stored BPDU<1,4,3,2> to Device B. Device B find that BPDU<1,4,3,2> has a higher priority than BPDU<2,0,2,2> stored in port GigabitEthernet 0/2 of Device B, deems that Device A is the root bridge, and sets port GigabitEthernet 0/2 as the root port. After two periods of **Forward Delay** elapse, port GigabitEthernet 0/2 of Device B transitions to forwarding state to forward data, implementing link auto-recovery in the STP network.
- If a port on a downstream device changes to down state, the device sends a TCN BPDU to the upstream device to notify the upstream device of the topology change. After receiving the TCN BPDU from the downstream device, the upstream device sets bit 7 (indicating TCA) in the **Flags** field of the configuration BPDU to **1**, and sends the configuration BPDU to the downstream device, indicating that the upstream device has known the topology change and requesting the downstream device to stop sending the TCN BPDU. Only the designated port of the upstream device processes TCN BPDUs. Non-designated ports discard TCN BPDUs directly once they receive such BPDUs. The upstream device copies the TCN BPDU and forwards it to its upstream device through the root port. The TCN BPDU is forwarded segment by segment until it reaches the root bridge. After receiving the TCN BPDU from a downstream device, the root bridge sets bit 7 (indicating TCA) in the **Flags** field of the configuration BPDU to **1** and sends the configuration BPDU to the downstream device, indicating that the upstream device has known the topology change and requesting the downstream device to stop sending the TCN BPDU. In addition, the root bridge sets bit 0 (indicating a TC) in the **Flags** field of the configuration BPDU to **1** to ask all downstream devices to clear dynamic MAC address entries of ports and relearn MAC addresses due to the topology change.
- If a new Device D is connected to Device B and the port status of Device B becomes up, Device B will also notify others of the topology change. Device D will send and receive BPDUs. If Device D has a higher priority than the original root (Device A), great changes may be incurred to the spanning tree. In conclusion, STP is not stable.

10. Drawbacks of STP

- Topology convergence is slow. After port role election is complete, STP waits for twice the period of **Forward Delay** (15s by default) before entering forwarding state, namely, STP needs to wait 30 seconds (that is, 2 × 15). Every time the topology changes, a bridge needs to re-elect the root port and designated ports after the period of **MAX Age** (20s by default), and then wait for twice the period of **Forward Delay** before entering forwarding state. Therefore, it takes about 50s (that is, 20 + 2 × 15 = 50 seconds) to generate a stable topology. STP requires the root bridge to actively send configuration BPDUs and other devices to forward the BPDUs. This mechanism leads to slow network convergence.
- If the network topology changes frequently, the network will be frequently disconnected, which is unacceptable to users.

1.1.4 RSTP

1. Overview

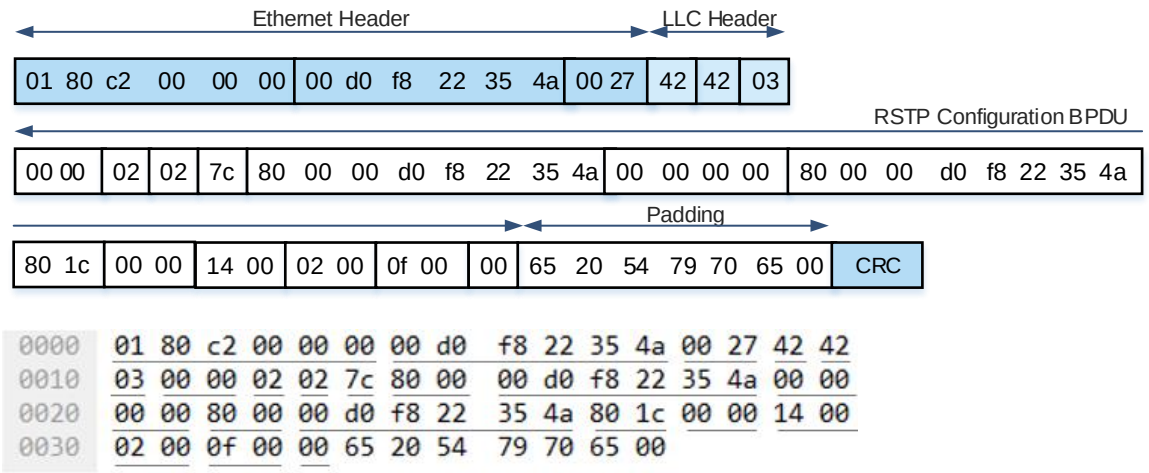
RSTP, defined by IEEE 802.1W, evolves from STP and is compatible with IEEE 802.1D STP downward. RSTP has all functions of the conventional STP protocol and is capable of preventing loops and providing redundant links. Compared with STP, RSTP shortens the time required for network topology convergence. If all the bridges in a LAN support RSTP and are configured properly, once the network topology changes, the time required by RSTP to re-generate the topology tree is less than 1s while STP needs about 50s.

2. RSTP Protocol Packets

RSTP needs only one type of protocol packet: configuration BPDU. The packet format is slightly different from that in STP. The configuration BPDU of RSTP is also called rapid spanning tree BPDU (RST BPDU).

When a topology changes, RSTP no longer uses a TCN BPDU to notify the topology change. Instead, RSTP sets the TC bit in the Flags field of the RST BPDU to 1 and notifies the entire network of the topology change via RST BPDU flooding.

Figure 1-5 Configuration BPDU of RSTP



As shown in Figure 1-5, the Length field is 0x0027, which indicates that the BPDU length is 39 bytes, including a 3-byte LLC header and a 36-byte RST BPDU. The fields of a RST BPDU packet are described in Table 1-8. The 39 bytes, plus the last 7 padding bytes and the 4 CRC bytes, together form a 64 byte Ethernet frame.

Table 1-8 Format of an RST BPDU

Field	Bytes	Description
PID	2	Protocol identifier. The value is **0x0000**.
PVID	1	Protocol version identifier. The value is **0x02**, indicating an RSTP packet. Devices running STP will discard RST BPDUs upon receiving.
BPDU Type	1	BPDU packet type. The value is **0x02**, indicating an RST BPDU.

Field	Bytes	Description
Flags	1	An RST BPDU uses all the eight bits. In the example, the value is 0x7c, that is, 0111 1100. ● Bit 7: TCA. The value **1** indicates that a topology change is known. ● Bit 6: Agreement. The value **1** indicates consent to status switching. ● Bit 5: Forwarding. The value **1** indicates that forwarding is allowed. ● Bit 4: Learning. The value **1** indicates that learning is allowed. ● Bit 3 and bit 2: Port role. ○ **00**: Unknown. ○ **01**: Root port. ○ **10**: Alternate/Backup. ○ **11**: Designated port. ● Bit 1: Proposal. The value **1** indicates sending a proposal to request status switching. ● Bit 0: TC. The value **1** indicates notifying the topology change.
Root ID	8	Identifier of the root bridge recognized by the device. It consists of a 2-byte bridge priority and a 6-byte bridge MAC address.
RPC	4	Root path cost, which is the cumulative path cost from the port to the root bridge.
Bridge ID	8	Bridge identifier of the device. It consists of a 2-byte bridge priority and a 6-byte bridge MAC address.
Port ID	2	Identifier of the port that sends the BPDU. It consists of a 1-byte priority and 1-byte port ID.
Message Age	2	Age of the BPDU, which represents the time that the BPDU has been alive in the network.
Max Age	2	Aging time of the BPDU.
Hello Time	2	Interval for sending two adjacent BPDUs. After an RSTP topology becomes stable, a non-root switch automatically sends configuration BPDUs at an interval of **Hello Timer** regardless of whether it receives the configuration BPDU from the root bridge. If a non-root bridge fails to receive a BPDU from the root within the time of **Time-factor** × **Hello Time**, it considers the root bridge or the link to the root bridge to be faulty. **Time-factor** is a multiple of **Hello Time** and can be configured.
FWD Delay	2	Forwarding delay. This time parameter controls the duration of listening and learning states.
Ver 1 Length	1	The value is 0x00, indicating that the BPDU does not contain content of Version 1. STP does not have this field.

3. Port Roles

RSTP uses alternate ports and backup ports to replace blocking ports, and adds the edge port auto-identification function. Therefore, RSTP supports six types of ports:

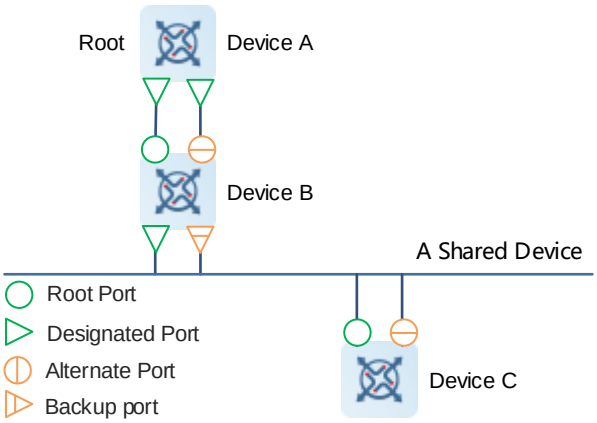
- Root port: A port with the shortest path to the root bridge.
- Alternate port: A backup port of a root port. Once the root port fails, the alternate port becomes the root port immediately.
- Designated port: A port that connects a root bridge or upstream bridge to a downstream device.
- Backup port: A backup port of a designated port. When a root bridge or upstream bridge has two interfaces to connect to the same downstream bridge, the interface with a higher priority is the designated port and the interface with a lower priority is the backup port. Once the designated port fails, the backup port becomes the designated port immediately.
- Edge port: A port connected to a terminal. Edge ports do not participate in spanning tree calculation. They can rapidly enter forwarding state with no need to wait for twice the period of **Forward Delay**. An edge port that receives a BPDU automatically becomes a non-edge port.
- Disable port: A port in down state.

Improvement effects:

- The work mechanism of the alternate port and backup port enables a device to restore the network connectivity within several milliseconds after discovering a topology change, with no need to transmit configurations.
- The status change of an edge port will neither affect the network connectivity nor cause loops. Edge ports can enter forwarding state without delay.

Figure 1-6 shows port roles. The port priority sequence is as follows unless otherwise specified: root port > designated port > alternate port > backup port.

Figure 1-6 RSTP Port Roles



4. BPDU Processing of RSTP

In RSTP, a device processes BPDUs in a different way.

- Independent Hello: After an STP topology becomes stable, the root bridge sends configuration BPDUs at an interval of **Hello Timer**. Non-root switches send configuration BPDUs only after receiving the configuration BPDU from the upstream device, resulting in long convergence time in the case of a topology change. After an RSTP topology is stable, non-root switches automatically send configuration BPDUs at an interval of **Hello Timer** regardless of whether they receive configuration BPDUs from the root bridge.
- Fast fault detection: In STP, if a non-root switch fails to receive a BPDU from the upstream device within one

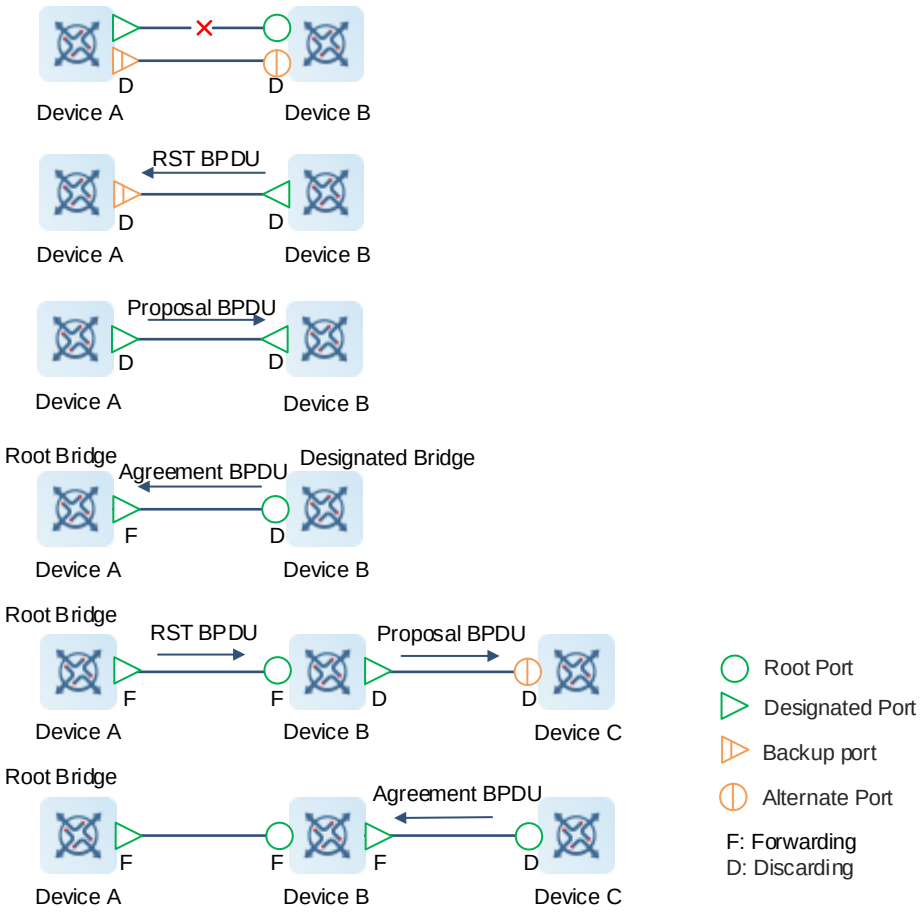
period of **MAX Age**, the switch considers the upstream device to be faulty. In RSTP, if a non-root switch fails to receive a BPDU from the upstream device within three periods of **Hello Time**, it considers the upstream device to be faulty. The value range of **Max Age** is from 6 to 40, in seconds, the default value is **20**, and it takes at least 6s to detect the upstream device fault. The value range of **Hello Time** is from 1 to 10, in seconds, the default value is **2**, and it takes at least 3s to detect the upstream device fault. Therefore, RSTP supports a shorter fault detection time.

- Fast convergence: In STP, all ports discard low-priority BPDUs upon receipt. If a blocking port fails to receive a high-priority BPDU within one period of **MAX Age**, the port changes to the designated port and sends its BPDU to the peer. In RSTP, when a downstream device receives a BPDU from the upstream device and finds that the priority of the BPDU is lower than that of the BPDU stored in the port, it sends its stored BPDU to the upstream device immediately. The upstream device compares the received BPDU with its own BPDU and updates the BPDU stored in the local port. This mechanism can accelerate topology convergence.

5. Fast Convergence of RSTP

In STP, to transition from listening state to learning state and then to forwarding state, a port needs to wait for two periods of Forward Delay. The value range of Forward Delay is from 4 to 30, in seconds, the default value is 15, and therefore, the default waiting time is 30s. In RSTP, a port enters discarding state after role election is complete, and then performs a Proposal/Agreement handshake with the connected bridge so as to quickly enter forwarding state. Figure 1-7 shows a fast convergence process of RSTP for example.

Figure 1-7 Fast Convergence of RSTP



(1) Proposal: If the root port of Device B fails to receive a BPDU from the upstream device within three periods of **Hello Time**, it considers the upstream device to be faulty. Device B returns to the initial state, considers itself as the root, and sets all its ports as designated ports (in discarding state) to send RST BPDUs. The alternate port (if Device A is not the root) or backup port (if Device A is the root) of Device A receives an RST BPDU from Device B. If Device A finds that the RST BPDU from Device B has a lower priority than its RST BPDU, Device A sets its port as a designated port (in discarding state), sets the **Proposal** bit in the configuration BPDU of the port to **1**, and sends a Proposal BPDU to Device B to request Device B to recognize Device A as the root bridge.

(2) Agreement: Device B receives the Proposal BPDU from Device A. If the BPDU of Device A is superior to that of Device B, Device B considers Device A as the root bridge (Device A may be the root or an intermediate device between Device B and the root). The port receiving the Proposal BPDU on Device B becomes the root port, and the root port stops sending RST BPDUs. Before a port transitions to forwarding state, Device B needs to block non-edge designated ports to prevent loops. The time required by ports to transition to -discarding state may be different. Devices use a set of synchronization variables to indicate whether the port state transition is complete. Device B sets the **Sync** variable to **1** for all ports. After all the ports enter discarding state, the **Synced** variable of all the ports is set to **1**. Then, the **Synced** variable of the root port is set to **1**, indicating that all the ports have been transitioned to the correct state. The root port sets the **Proposal** bit to **0** and the **Agreement** bit to **1** in the Proposal BPDU received from Device A and keeps other content of the BPDU unchanged, to create an Agreement BPDU. Then, the root port sends the BPDU to Device A.

(2) Forwarding: The designated port of Device A enters forwarding state after receiving the Agreement BPDU, and sends the RST BPDU with the **Forwarding** bit set to **1** at an interval of **Hello Time**. After receiving the RST BPDU, the root port of Device B enters forwarding state.

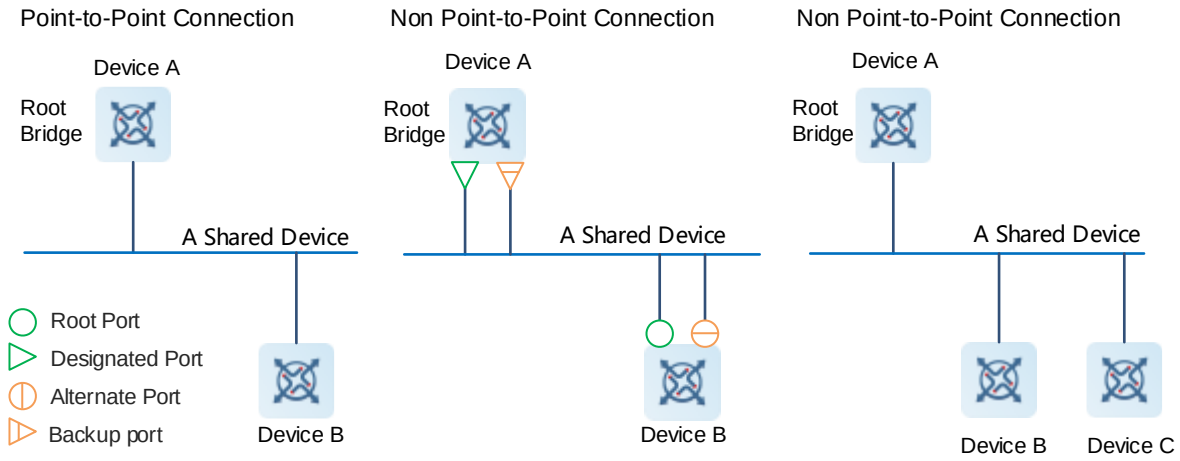
(3) Extension proposal: The designated port of Device B sends a Proposal BPDU with both **Proposal** and **Agreement** bits set to **1** to the downstream device, to request the downstream device to recognize Device B as the upstream device and extend the spanning tree. Theoretically, RSTP can recover the network tree topology immediately and achieve fast convergence once the network topology changes.

Improvement effect: Non-edge designated ports are blocked actively, without waiting for the period of Forward Delay. Hence, the convergence is accelerated. Network connectivity can be restored within the time of exchanging two handshake BPDUs, and the delay incurred by handshakes is less than 1s.

⚠ Caution

The above "handshake" process must meet one condition: Ports must be in a point-to-point connection state. Otherwise, the ports cannot shake hands and fast convergence cannot be achieved. To make RSTP-compliant devices function as they should, you are advised to connect devices in a point-to-point way.

Figure 1-8 Examples of Point-to-Point Connection and Non Point-to-Point Connection



6. Drawbacks of RSTP

- RSTP creates only one spanning tree in the whole switching network, which cannot prevent performance degradation caused by the network expansion. Therefore, when using RSTP, it is not recommended that the number of devices between any two nodes running STP/RSTP in the network exceed 7.
- RSTP can only implement redundant backup but does not support load balancing based on VLAN traffic.
- STP and RSTP calculate the spanning tree topology based on devices (nodes), and the impact of VLANs is not taken into account. In a particular topology, ports in the same VLAN may be blocked and fail to communicate with each other. This problem and the solution will be detailed in the MSTP section.

1.1.5 MSTP

1. Overview

MSTP, defined by the IEEE 802.1s standard, is capable of ironing out drawbacks of STP, RSTP (only one spanning tree is generated for a device), and Per-VLAN Spanning Tree (PVST) (one spanning tree is generated for each VLAN).

- After MSTP is deployed, if a loop arises in a network, MSTP performs topology calculation, and eliminates possible loops in the network by blocking redundant links.
- When an active path fails, MSTP activates the redundant backup link to restore the network connectivity.
- MSTP supports fast convergence.
- MSTP forwards the traffic of different VLANs along their respective spanning trees, which provides a better load balancing mechanism for redundant links.

2. Origin of MSTP

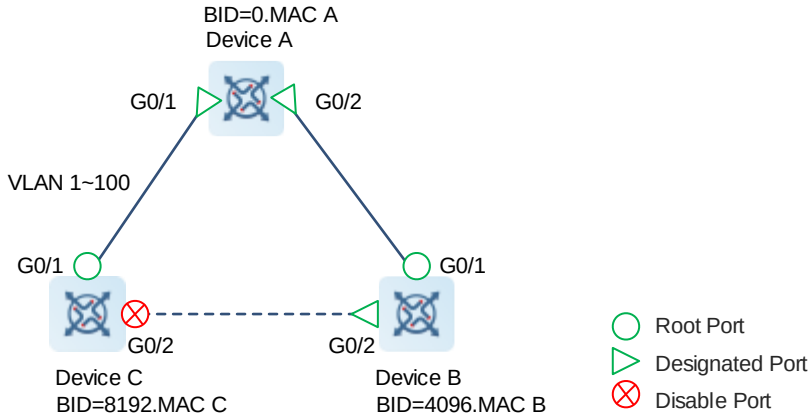
The conventional STP and RSTP protocols have the following two problems and MSTP emerges as a solution to them.

- Both STP and RSTP block redundant lines and concentrate traffic on the active link, which leads to a waste of bandwidth of the redundant link and heavy traffic on the active link, resulting in traffic imbalance.

As shown in Figure 1-9, assume that the rate of each segment of link is 1 Gbps. If Devices A, B, and C run STP, the STP algorithm first elects Device A as the root based on the bridge ID. On non-root bridge B, the STP

algorithm compares <**Root Path Cost**, **Bridge ID**, **Port ID**> in sequence, and selects port GigabitEthernet 0/1 as the designated port based on the root path cost. STP performs the same operation on Device C. Ports on the directly connected segment between Device C and Device B have the same root path cost. Port GigabitEthernet 0/2 of Device B is elected as the designated port based on the bridge ID, and port GigabitEthernet 0/2 of Device C is blocked. Port GigabitEthernet 0/2 of Device C cannot make communication due to STP blocking. VLANs 1–100 can be configured only on port GigabitEthernet 0/1 of Device C. It is impossible to implement traffic sharing by configuring port GigabitEthernet 0/1 of Device C to allow VLANs 1–50 and port GigabitEthernet 0/2 of Device C to allow VLANs 51–100.

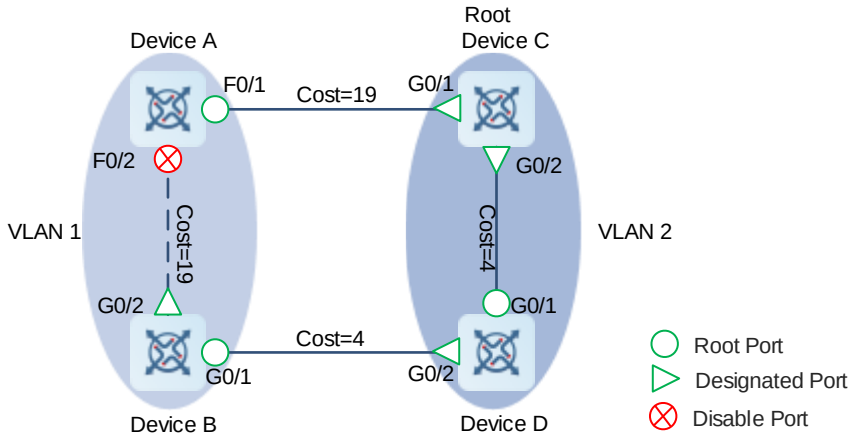
Figure 1-9 Traffic Sharing Failure in STP



- In special cases, STP and RSTP may block links in the same VLAN, causing interfaces within the same VLAN to be unable to communicate with each other.

  As shown in Figure 1-10, Devices A and B are in VLAN 1, and Devices C and D are in VLAN 2. Assume that the topology tree calculated by STP is shown in the figure. Port FastEthernet 0/2 of Device A is blocked, and the link between Devices A and B is blocked. VLAN 1 is not configured on Devices C and D, and Devices C and D cannot forward packets of VLAN 1. As a result, Devices A and B in VLAN 1 cannot communicate with each other.

Figure 1-10 Communication Failure in the Same VLAN Caused by STP



3. Instance and Instance Mapping

The collection of one or more VLANs on one device is called a Multiple Spanning Tree Instance (MSTI) or instance for short. If a device supports a maximum of 65 instances, instance 0 exists by default and cannot be deleted, and instances 1–64 can be created or deleted. Users can allocate VLANs 1–4094 to different instances as required, and unallocated VLANs belong to instance 0 by default. The mapping between MSTIs and VLANs is called instance mapping, as shown in Table 1-9.

Table 1-9 Mappings Between MSTIs and VLANs

Instance	VLAN Mapped
0	1–2, 4, 11–4094
1	3, 5-7
2	8–10

Note

You are advised to disable MSTP before configuring mappings between MSTIs and VLANs, and then enable MSTP after configuration, to ensure stable convergence of the network topology.

4. Multiple Spanning Tree Region

Devices that run the MSTP protocol, have the same configuration name, revision number, and instance mappings constitute a multiple spanning tree (MST) region. Configuration names, revision numbers, and instance mappings are recorded in the MST CFG ID field of MST BPDUs and they can be configured.

- Configuration name: A string of 32 bytes used to identify an MST region.
- Revision number: A 2-byte non-negative integer used to identify an MST region.
- Instance mapping: The mappings between MSTIs and VLANs are encrypted into a 16-byte digest. One MST region can contain multiple MSTIs.

MST regions are independent of each other. If a port on a device receives a BPDU with MST CFG ID same as that of the MST BPDU of the device, the device deems that the peer device and the device belong to the same MST region. Otherwise, the device deems that the peer device belongs to a different MST region. The load sharing advantage of MSTP can be reflected only after multiple devices are configured to the same MST region. Therefore, MST regions need to be properly divided and devices in the same MST region need to have the same MST CFG ID.

5. Instance Spanning Tree

One instance spanning tree can be generated for each instance. Each instance calculates and generates an instance spanning tree based on MSTI configuration messages in MST BPDUs. Information contained in MST configuration messages is described in Table 1-10.

If a device contains only one instance or MST configuration messages of multiple instances are the same, the same instance spanning tree will be generated, which is useless in traffic sharing. If VLANs are allocated to different instances to form multiple VLAN groups and different bridge priorities and port priorities are configured

for different instances, multiple different instance spanning trees can be generated for the instances in the region to share VLAN traffic. The spanning tree of each instance is not affected by the spanning trees of other instances.

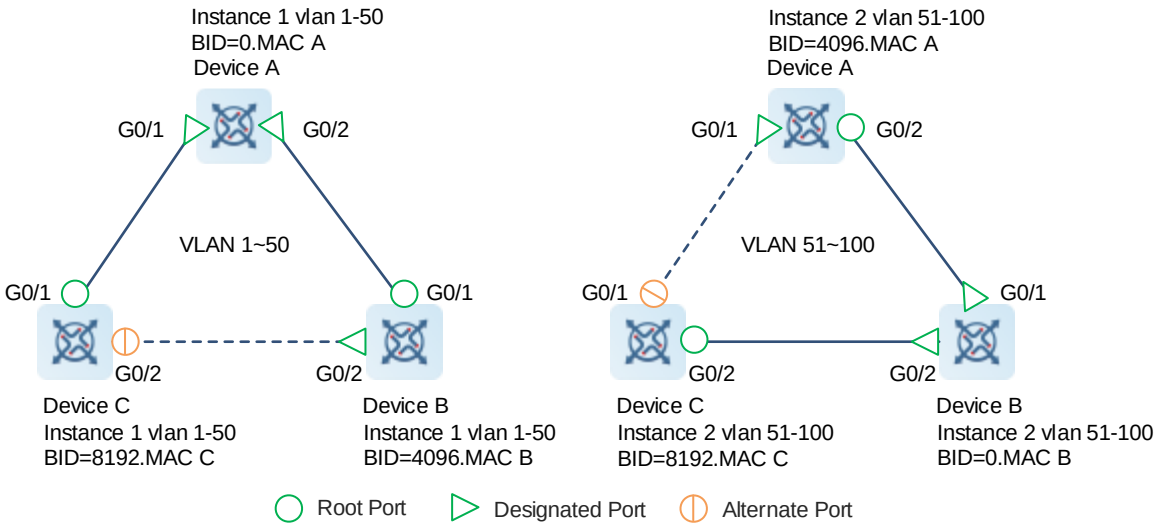
Caution

"MSTI configuration messages" of instances need to be manually modified so that spanning trees of instances are different, that is, the selected root, designated ports, blocking ports, and active links are different for different instances. In this way, VLAN traffic sharing can come into play.

MSTP calculates the topology for each instance in the same way as RSTP. In each instance, MSTP first elects the root bridge by comparing the root ID, and all ports on the root bridge are designated ports. Then, MSTP compares key elements in the priority vector <Root ID, Root Path Cost, Bridge ID, Port ID> on non-root bridges to elect root ports. Afterwards, MSTP compares the two ports of each directly connected network segment, and elects the port with a higher priority as the designated port and the one with a lower priority as the alternate port. Devices produce multiple instance spanning trees by means of MST BPDU interaction and calculation. Each instance provides a single loop-free network topology for VLAN groups in the instance, and the topologies of different VLAN groups are different.

For example, after MSTP runs in the topology as shown in Figure 1-9, the calculated topology is shown in Figure 1-11.

Figure 1-11 Traffic Sharing Implemented by MSTP



In instance 1, devices elect Device A as the root by comparing root IDs (bridge IDs of the devices at this moment). Devices compare the root path cost and elect port GigabitEthernet 0/1 of Device B and port GigabitEthernet 0/1 of Device C as the root ports. In the directly connected network segment between Devices C and B, the two devices have the same root ID (0.MAC A). Bridge ID of Device B is smaller than that of Device C, and therefore, port GigabitEthernet 0/2 of Device B is elected as the designated port and port GigabitEthernet 0/2 of Device C is elected as the alternate port. When port GigabitEthernet 0/1 of Device C malfunctions, port GigabitEthernet 0/2 of Device C becomes the root port immediately.

In instance 2, devices elect Device B as the root by comparing root IDs (bridge IDs of the devices at this moment). Devices compare the root path cost and elect port GigabitEthernet 0/2 of Device A and port

GigabitEthernet 0/2 of Device C as the root ports. In the directly connected network segment between Devices A and C, the two devices have the same root ID (0.MAC B). Bridge ID of Device A is smaller than that of Device C, and therefore, port GigabitEthernet 0/1 of Device A is elected as the designated port and port GigabitEthernet 0/1 of Device C is elected as the alternate port. When port GigabitEthernet 0/2 of Device C malfunctions, port GigabitEthernet 0/1 of Device C becomes the root port immediately.

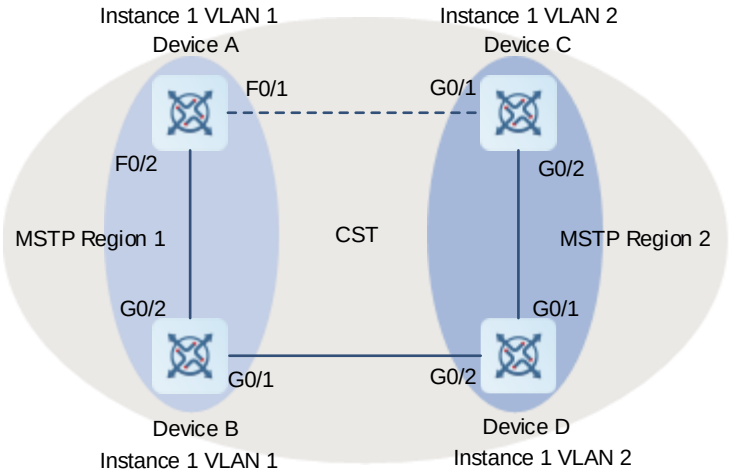
It can be seen that different spanning trees are calculated for instance 1 and instance 2 due to different bridge IDs. Consequently, packets of VLANs 1–50 are forwarded along the spanning tree of instance 1 and packets of VLANs 51–100 are forwarded along the spanning tree of instance 2, implementing VLAN traffic sharing.

Users can, as needed, configure different instances for different VLAN groups, Configuring the Bridge Priority and Port Priority by instance, Configuring the Port Path Cost, and plan VLAN groups to forward traffic through different paths.

6. Common Spanning Tree

Devices with different instances and VLAN mapping configurations do not belong to the same MST region and each MST region is a whole. Therefore, one MST region can be considered as one node. The inter-node spanning tree calculated via STP is called a common spanning tree (CST). Instances with the same ID can exist in different MST regions, for example, instance 1 exists in two MST regions. Though the instance IDs are the same, instance spanning trees are generated independently for the instances and the instances are unrelated to each other and unrelated to the CST. If two MST regions are configured in a topology shown in Figure 1-10, the calculated topology is shown in Figure 1-12 after MSTP runs.

Figure 1-12 Preventing a Communication Failure of Ports in the Same VLAN by MSTP



Devices A and B (both instance 1 VLAN 1) have the same instance mapping. There is no loop in MST region 1, and therefore, no port is blocked in this region.

Devices C and D (both instance 1 VLAN 2) have the same instance mapping. There is no loop in MST region 2, and therefore, no port is blocked in this region.

MST region 1 and MST region 2 can be considered as two nodes and there are loops between the two nodes. Therefore, a link is blocked according to the configuration to prevent loops and ensure that the intra-VLAN communication is not affected.

7. Common and Internal Spanning Trees

In a topology as shown in Figure 1-13, there are three MST regions. In each MST region, one instance spanning tree is generated for each instance. As shown in Figure 1-14, in region 2, one spanning tree is generated for instance 0, instance 1, and instance 2 each. According to the MSTP protocol, the spanning tree of instance 0 represents the region and is called an internal spanning tree (IST).

Each MST region is considered as a node, and one spanning tree is calculated between nodes according to the STP protocol. This spanning tree is called a common spanning tree (CST). As shown in Figure 1-13, a spanning tree connected using blue lines indicates an IST in a region while a spanning tree connected using orange lines indicates an inter-region CST.

ISTs and a CST constitute a common and internal spanning tree (CIST).

CIST root: Root of a CIST. It is also the root of a CST, that is, CST root. In the CST calculation, the device with the smallest bridge ID is elected as the CIST root. As shown in Figure 1-13, assume that Device A has the smallest bridge ID. Device A is elected as the CIST root.

Region root: In an IST, the device with the smallest root path cost to the CIST root is elected as the region root of the IST, which is also called master bridge. CIST root A is also the region root of region 1. As shown in Figure 1-13, assume that Device B in region 2 has the smallest root path cost to Device A. Device B is elected as the region root of region 2. Device B may not be directly connected to Device A but may through other devices in region 1. Likewise, Device C is elected as the region root of region 3.

The region root of region 2 is Device B. Suppose that three spanning trees are calculated for the three instances in region 2, and the roots of the three spanning trees are all Device B. On Device B, the port with the smallest root path cost to the CIST root is the master port. Therefore, compared with RSTP, MSTP adds two port roles: master port and region edge port.

⚠ Caution

The **CIST Regional Root** field in MSTP packets indicates a region root rather than the CIST root.

A region root (represented by **CIST Regional Root**) may not be the device with the smallest bridge ID in a region but is the device with the smallest root path cost from the region to the CIST root.

To make the topology more stable, it is recommended that the "outlets" of different instance spanning trees in each region to the CIST root be configured on the same device of the region. As shown in Figure 1-14, the outlets of MSTI 1, MSTI 2, and MSTI 3 in region 2 to the CIST root are configured on Device B.

Figure 1-13 CIST Composed of ISTs and the CST

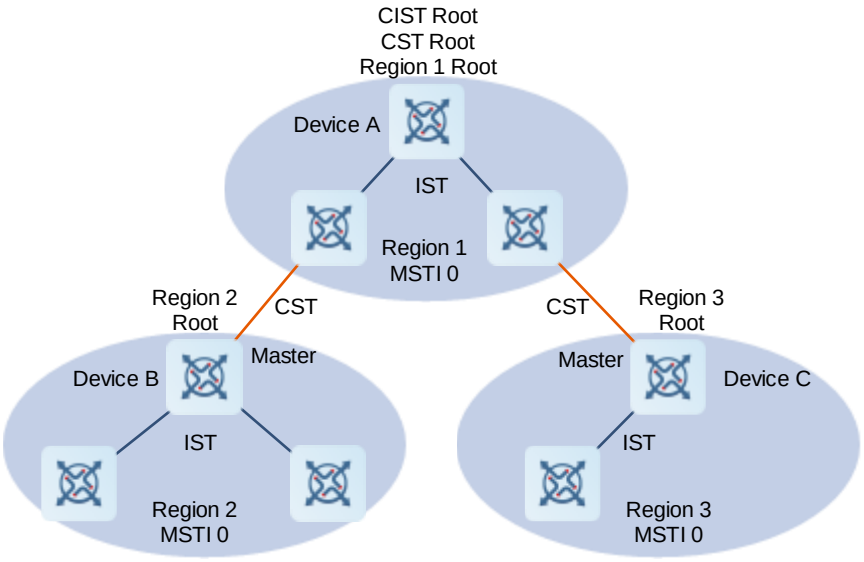
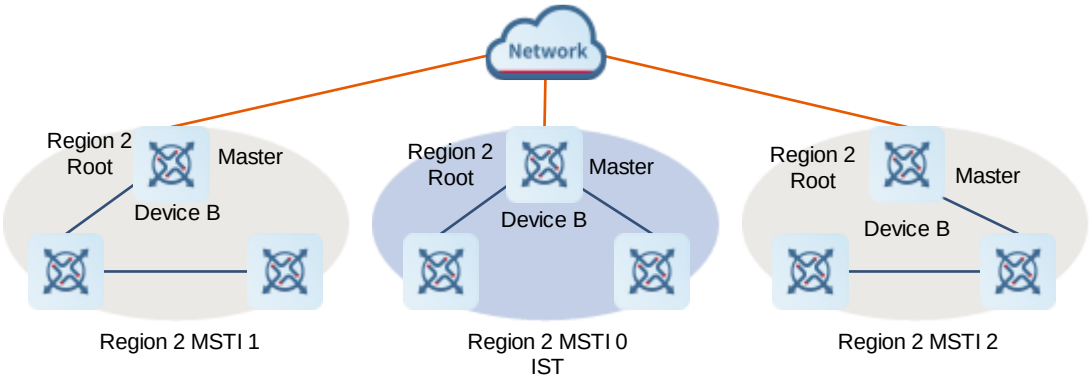


Figure 1-14 Three Spanning Trees Generated for the Three Instances in Region 2



8. Port Roles

MSTP introduces a new port role to the region root: master port, which is the port with the smallest root path cost from a region to the CIST root. The master port is the "outlet" of all instances in a region, and all instances can forward data through the master port.

- Root port: A port with the shortest path to the root bridge.
- Designated port: A port that connects a bridge to the root bridge in each LAN.
- Alternate port: A backup of a root port. Once the root port fails, the alternate port becomes the root port immediately.
- Backup port: A backup of a designated port. When two ports on a bridge connect to the same LAN, the port with a higher priority is the designated port and the one with a lower priority is the backup port. Once the designated port fails, the backup port changes to the designated port immediately.
- Edge port: A port connected to a terminal. Edge ports do not participate in spanning tree calculation.
- Disable port: A port in down state.
- Master port: A port with the shortest path cost to the CIST root. It is the "outlet" of all MSTIs in an MST

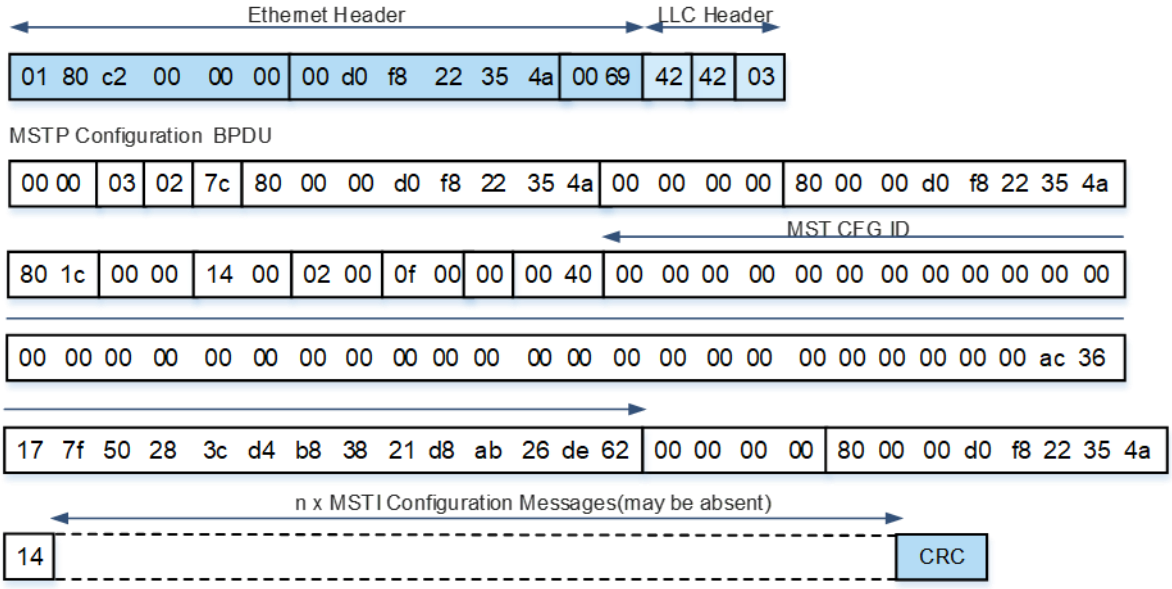
region.

- Region edge port: A port directly connected to other MST regions. A master port is a special region edge port.

9. MSTP Packets

MSTP uses MST BPDUs as the basis for spanning tree calculation. MST BPDUs can be used to calculate spanning tree topologies, maintain network topologies, and communicate topology change records.

Figure 1-15 Configuration BPDU of MSTP



As shown in Figure 1-15, the Length field is 0x0069, which indicates that the packet length is 105 bytes in this example, including a 3-byte LLC header and a 102-byte MST BPDU. This length is the minimum length of an MST BPDU. For both intra-region and inter-region MST BPDUs, the first 36 bytes are the same as those of an RST BPDU. MSTP-specific fields start from the 37th byte, and the MSTI configuration message field at the end is composed of multiple MSTI configuration messages. The fields are described in Table 1-10.

Table 1-10 Format of an MST BPDU

Field	Bytes	Description
PID	2	Protocol identifier. The value is 0x0000.
PVID	1	Protocol version identifier. 0x03 indicates that the packet protocol is MSTP.
BPDU Type	1	BPDU packet type. The value is 0x02 for MSTP. 0x02 indicates that the BPDU type is MST BPDU.

Field	Bytes	Description

CIST Flags	1	Configuration BPDUs of MSTP use all the eight bits. In the example, the value is 0x7c, that is, 0111 1100.
		● Bit 7: TCA. An upstream device sets bit 7 in a configuration BPDU to **1** and sends the BPDU to notify the downstream device of the topology change and request the downstream device to stop sending the TCN BPDU.
		● Bit 6: Agreement. The value **1** indicates consent to status switching.
		● Bit 5: Forwarding. The value **1** indicates forwarding state.
		● Bit 4: Learning. The value **1** indicates learning state.
		● Bit 3 and bit 2: Port role.
		  ○ 00: Unknown.
		  ○ 01: Root port.
		  ○ 10: Alternate/Backup.
		  ○ 11: Designated port.
		● Bit 1: Proposal. The value **1** indicates sending a proposal to request status switching.
		● Bit 0: TC. The root bridge sets bit 0 in a configuration BPDU to **1** and sends the BPDU to instruct the downstream device to delete MAC address entries.

CIST Root ID	8	Bridge ID of the CIST root, which consists of a 2-byte root priority and a 6-byte root MAC address. In the example, the priority value is 0x8000, which indicates 32768, and the bridge MAC address is 00-d0-f8-22-35-4a.
CIST External Path Cost	4	CIST external path cost, which refers to the cumulative path cost from the MST region, to which the device belongs, to the CIST root. The CIST external path cost is calculated based on the link bandwidth, which is 0x00 00 00 00 in the example.
CIST Regional Root ID	8	CIST regional root identifier, that is, bridge ID of the IST master bridge. If the CIST root is in the region, the region root ID is the CIST root ID. In the example, the priority value is 0x8000, which indicates 32768, and the bridge MAC address is 00-d0-f8-22-35-4a.
CIST Port ID	2	ID of a port in an IST. In the example, the priority value is 0x80, which indicates 128, and the port ID is 1c.
Message Age	2	Age of a BPDU, that is, keepalive time of the packet. The value is 0x0000 in the example.
Max Age	2	Maximum timeout time of a BPDU, after which the link to the root switch is considered to be faulty. The value in the example is 0x1400 (low-order bytes on the left), indicating 20 seconds.
Hello Time	2	Hello interval for sending two adjacent BPDUs. The default value is **2**s. The value in the example is 0x0200 (low-order bytes on the left), indicating 2 seconds.
FWD Delay	2	**Forward Delay** timer. The default value is **15**s. This field controls the duration of listening and learning states. The value in the example is 0x0f00 (low-order bytes on the left), indicating 15 seconds.

Field	Bytes	Description
Ver 1 Length	1	Length of a Version1 BPDU. The value is fixed to 0x00.
Ver 3 Length	2	Length of a Version3 BPDU. The value is 0x0040 in the example.
MST CFG ID	51	MST configuration identifier, that is, MST region label information, which contains four fields. Interconnected devices with the same four fields in **MST CFG ID** belong to the same region. ● (1-byte) Configuration Identifier Format Selector: The value is fixed to 0x00. ● (32-byte) Configuration Name: Region name. The default value is all 0's. ● (2-byte) Revision Level: Revision number, which is a non-negative integer. The value is 0x0000. ● (16-byte) Configuration Digest: MSTI-VLAN instance mapping table. The HMAC-MD5 algorithm is used to encrypt the mappings between instances and VLANs in a region into a 16-byte digest.
CIST Internal RPC	4	CIST internal root path cost, which is the cumulative path cost from the local port to the master switch of the IST. The internal root path cost of a CIST is calculated based on the link bandwidth. The value is 0x00-00-00-00 in the example.
CIST Bridge ID	8	ID of the designated switch in the CIST. It consists of a 2-byte bridge priority and a 6-byte bridge MAC address. In the example, the priority value is 0x8000, which indicates 32768, and the bridge MAC address is 00-d0-f8-22-35-4a.
CIST R Hops	1	Number of remaining hops of a BPDU in a CIST. The default value is 0x14, which indicates 20 hops.
MSTI Configuration Messages (may be absent)	16	MSTI configuration message. The configuration message of each MSTI occupies 16 bytes. If there are *n* MSTIs, they occupy *n* × 16 bytes. Fields in the configuration message of a single MSTI are described as follows: ● (1-byte) MSTI Flags: MSTI flag. ● (8-byte) MSTI Regional Root Identifier: MSTI region root ID. ● (4-byte) MSTI Internal Root Path Cost: Cumulative path cost from the local port to the MSTI region root switch. The internal root path cost of an MSTI is calculated based on the link bandwidth. ● (1-byte) MSTI Bridge Priority: Bridge priority of the device in the MSTI. ● (1-byte) MSTI Port Priority: Port priority of the device in the MSTI. ● (1-byte) MSTI Remaining Hops: Number of remaining hops of a BPDU in the MSTI.

1.1.6 Protocols and Standards

- IEEE 802.1D: Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges
- IEEE 802.1w: IEEE Standard for Information Technology -Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Common Specifications - Part 3: Media Access Control (MAC) Bridges: Amendment 2: Rapid Reconfiguration
- IEEE 802.1s: IEEE Standards for Local and Metropolitan Area Networks - Amendment to 802.1Q Virtual Bridged Local Area Networks: Amendment 3: Multiple Spanning Trees

1.2 Restrictions and Guidelines

- The MSTP feature only takes effect on a Layer 2 interface. A Layer 3 interface cannot be directly configured as a dot1q tunnel port, and must be configured as a Layer 2 interface first.
- You can use the **display interface switch-port** command to view the interface type. If the Switchport field corresponding to the interface is "enabled", it means that the interface is a Layer 2 interface. If the Switchport field corresponding to the interface is "disabled", it means that the interface is a Layer 3 interface.
- You can use the **switch-port** command to configure the interface as a Layer 2 interface. First, use the **unset switch-port mode** command to restore the mode of the Layer 2 interface to the default value, and then use the **unset switch-port** command to configure the Layer 2 interface as a Layer 3 interface. When the Layer 2/Layer 3 mode of the interface is modified, all properties of the interface are restored to the default values of the interface type.

1.3 Configuration Task Summary

1.3.1 Configuring STP

The STP basic configuration includes the following tasks:

(1) Configuring the Spanning Tree Mode

(2) (Optional) Configuring the Bridge Priority and Port Priority

(3) (Optional) Configuring the Port Path Cost. Select at least one of the following to configure.

- Configuring the Path Cost Calculation Method
- Configuring the Path Cost Value

(4) (Optional) Configuring Spanning Tree Time Parameters

(5) Enabling the STP Function

1.3.2 Configuring RSTP

The RSTP basic configuration includes the following tasks:

(1) Configuring the Spanning Tree Mode

(2) (Optional) Configuring the Bridge Priority and Port Priority

(3) (Optional) Configuring the Port Path Cost. Select at least one of the following to configure.

- Configuring the Path Cost Calculation Method
- Configuring the Path Cost Value

(4) (Optional) Configuring Spanning Tree Time Parameters

(5) (Optional) Configuring Fast Convergence of Spanning Trees

(6) (Optional) Enabling Protocol Migration

(7) Enabling the STP Function

1.3.3 Configuring MSTP

The MSTP basic configuration includes the following tasks:

(1) Configuring the Spanning Tree Mode

(2) Configuring an MST Region

(3) Configuring the Bridge Priority and Port Priority

(4) (Optional) Configuring the Port Path Cost. Select at least one of the following to configure.

- Configuring the Path Cost Calculation Method
- Configuring the Path Cost Value

(5) (Optional) Configuring Spanning Tree Time Parameters

(6) (Optional) Configuring the Maximum Hop Count for BPDUs

(7) (Optional) Configuring Fast Convergence of Spanning Trees

(1) (Optional) Configuring Spanning Tree Compatibility for Interfaces

(2) Enabling the STP Function

1.3.4 Configuring Advanced Functions for Spanning Trees

Configuring advanced functions for spanning trees includes the following tasks. In networking applications, these functions are used to enhance the stability, robustness, and anti-attack capability of protocols based on the network topology and application characteristics, so as to meet the application requirements of protocols in different user scenarios. The following configuration tasks are all optional. When performing the following configuration, be sure to complete the configuration and then enable spanning tree protocols.

- Enabling the V-STP Function
- Configuring Port Protection
  - Configuring Root Guard
  - Configuring Loop Guard
- Configuring BPDU Source MAC Address Check
- Configuring an Edge Port
  - Configuring Autoedge
  - Configuring the Port Fast Attribute
- Configuring BPDU Guard or BPDU Filter
  - Configuring BPDU Guard
  - Configuring BPDU Filter
- Configuring TC Attack Defense Functions
  - Configuring TC Protection
  - Configuring TC Guard
  - Configuring TC Filter
- Configuring BPDU Tunnel

1.3.5 Configuring BPDU Transparent Transmission

When STP/RSTP/MSTP is not enabled on a device, if the spanning tree between the device and the interconnected device needs to be calculated properly, BPDU transparent transmission needs to be configured on the device to transparently transmit BPDU frames.

- Configuring BPDU Transparent Transmission

1.4 Notes

This section describes precautions for configuring the spanning tree function.

Whether the spanning tree function is supported depends on devices.

Table 1-11 Products Not Supporting the Spanning Tree Function

Function	Product Not Supporting the Function
Configuring BPDU Tunnel	S6980-64QC S6580-48CQ8QC S6580-24DC8QC

1.5 Configuring the Spanning Tree Mode

1.5.1 Overview

According to IEEE 802.1 protocol standards, STP, RSTP, and MSTP are mutually compatible, without any configuration by the user. Some vendors' devices may not implement STP/RSTP/MSTP according to protocol standards, causing incompatibility. Therefore, this spanning tree mode configuration command is provided for the user to switch the spanning tree mode to a lower version if other vendors' devices are incompatible with this device.

1.5.2 Restrictions and Guidelines

- **mstp** indicates the Multiple Spanning Tree Protocol (IEEE 802.1s).
- **rstp** indicates the Rapid Spanning Tree Protocol (IEEE 802.1w).
- **stp** indicates the Spanning Tree Protocol (IEEE 802.1d).
- If the STP mode is switched to RSTP mode, the convergence time is the convergence time of STP at the first topology change and the convergence time of RSTP at the second topology change.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the spanning tree mode.

stp mode { **mstp** | **rstp** | **stp** }

The default spanning tree mode is MSTP.

1.6 Configuring an MST Region

1.6.1 Overview

This section describes how to configure an MST region to change device members in the MST region, so as to affect the spanning tree topology.

1.6.2 Restrictions and Guidelines

- To enable multiple devices to belong to the same MST region, you need to configure the same configuration name, revision number, and instance mappings for the devices.
- One VLAN belongs to only one instance. By default, all VLANs are in Instance 0. In MST configuration mode, you can use the **instance instance**-*id* **vlan vlan**-*range* command to move the VLAN *vlan-range* to the *instance-id*. *instance-id* indicates the spanning tree instance number. The value range is 0-64. VLAN range refers to the list of VLAN IDs. The value range is 1-4094.
- You can use the **unset instance** *instance-id* or **reset instance** *instance-id* command to delete an instance (not Instance 0), and move all VLANs in the instance (not Instance 0) to Instance 0. The value of *instance-id* cannot be 0, and the value range is 1-64.
- You can use the **unset instance** *instance-id* **vlan** *vlan-range* command (with the parameter **vlan** *vlan-range*) to move all VLANs in the instance (not Instance 0) to Instance 0. The value of *instance-id* cannot be 0, and the value range is 1-64. If there are multiple VLANs in a specified instance, you can use the command with parameter to move some VLANs to Instance 0, and some VLANs remain in the specified instance. If all VLANs in the specified instance are moved to Instance 0, the specified instance will be automatically deleted.
- *name* indicates an MST name, which can contain 32 bytes at most. The default value is an empty string.
- *version* indicates the MST revision number. The default value is **0** and the value range is from 0 to 65535.

1.6.3 Prerequisites

Only MSTP supports instance configuration. If the spanning tree mode is not MSTP, run the stp mode mstp command to set the spanning tree mode to MSTP first.

You are advised to configure instance-VLAN mappings when STP is disabled, to ensure the stability and convergence of the network topology. If STP has been enabled, run the unset stp command to disable STP. Enable STP only after completing the instance-VLAN mapping configuration.

1.6.4 Procedure

(1) Create VLANs.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Create a group of VLANs.

vlan { *vlan-id* | **range** *vlan-range* }

Only VLAN 1 exists by default.

d Return to the global configuration mode.

exit

(2) Configure instance-VLAN mappings.

a Enter the MST configuration mode.

stp mst configuration

b Create an instance, and move the VLAN from the original instance to this instance.

instance *instance-id* **vlan** *vlan-range*

Only instance **0** exists and all VLANs belong to instance **0** by default.

(3) (Optional) Configure a name of an MST region.

name *name*

The default name of an MST region is an empty string.

(4) (Optional) Configure a revision number for the MST region.

revision *version*

The default revision number of an MST region is **0**.

1.7 Configuring the Bridge Priority and Port Priority

1.7.1 Overview

The bridge priority concerns the root bridge election as well as the topology of the entire network. Configure the bridge priority when the user needs to change the root or topology of a network.

The port priority concerns the port role election as well as ports entering forwarding state. If a loop arises in a region, the port with a higher priority is preferentially selected to transition to forwarding state. When ports share the same priority, the port with a smaller port ID is selected to transition to forwarding state. Configure the port priority when the user needs to change ports entering forwarding state.

1.7.2 Restrictions and Guidelines

- In **mst** *instance-id*, the value range of an instance ID is from 0 to 64 and the default value is **0**. When the spanning tree works in STP/RSTP mode, the instance ID can only be 0.
- The options of bridge priority indicated by **priority** include 16 values, all of which are multiples of 4096, that is, 0, 4096, 8192, 12288, 16384, 20480, 24576, 28672, 32768, 36864, 40960, 45056, 49152, 53248, 57344, and 61440. The default value is **32768**. It is recommended that the user configure a smaller bridge priority for core devices to ensure the network stability. Different bridge priorities can be configured for different instances, and each instance runs an independent STP based on the configuration. Devices in different regions care about the priority of only instance 0.
- The options of **port-priority** include 16 values, which are multiples of 16, that is, 0, 16, 32, 48, 64, 80, 96, 112, 128, 144, 160, 176, 192, 208, 224, and 240. The default value is **128**. Different port priorities can be configured for different instances on the same port, and an independent spanning tree can be generated for each instance based on the configuration. The changed port priority takes effect only if the port priority is changed on designated ports.

1.7.3 Prerequisites

- Configure the interface as a Layer 2 interface.

1.7.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the bridge priority.

stp [**mst** *instance-id*] **priority** *priority*

The default bridge priority is **32768**.

(4) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(5) Configure the port priority.

stp [**mst** *instance-id*] **port-priority** *port-priority*

The default port priority is **128**.

1.8 Configuring the Port Path Cost

1.8.1 Overview

The port path cost determines the root port election result on a device. A device elects the port with the minimum root path cost as the root port. The root path cost is the sum of path costs of all ports in the path from a device to the root.

The port path cost also affects the BPDU priority as well as the overall network topology. When devices compare the BPDU priority, the root path cost is compared if BPDUs have the same root ID and bridge ID.

The user can configure the port path cost to determine the port or path, through which packets pass preferentially.

1.8.2 Configuration Tasks

The configuration includes the following tasks:

(1) Configuring the Path Cost Calculation Method

(2) Configuring the Path Cost Value

1.8.3 Configuring the Path Cost Calculation Method

1. Overview

By default, the device automatically calculates the port path cost according to the physical interface rate, and the calculation method in the specification (as shown in Table 1-11). The cost of path with high rate is low, and vice versa. This calculation method can calculate the optimal spanning tree topology.

The standard of port path cost must be unified throughout the network. The user should be able to modify the method of automatic calculation of port path cost of a device when necessary.

2. Restrictions and Guidelines

- IEEE 802.1d Short: The range of the port path cost is from 1 to 65535. Aggregate interface cost = Physical port cost × 95%.
- IEEE 802.1t Long: The range of the port path cost is from 1 to 200000000. Aggregate interface cost = Physical port cost × 95%.
- IEEE 802.1t Long Standard: The range of the port path cost is from 1 to 200000000. Aggregate interface cost = Physical port cost/Linkupcnt. At this moment, the cost value of the aggregate interface will change with the number of member ports, which will lead to the network topology change. For configurations of aggregate interfaces and the Link Aggregation Control Protocol (LACP), see *Configuring Aggregated Port*.
  - When an aggregate interface is a static aggregate interface, **Linkupcnt** refers to the number of member ports in Link Up state.
  - When an aggregate interface is an LACP aggregate interface, **Linkupcnt** refers to the number of member ports participating in the aggregate interface data forwarding.
  - When no member port of an aggregate interface is in Link Up state or forwarding data, the value of **Linkupcnt** is **1**.

Table 1-12 Port Path Costs Calculated Based on the Link Rate

Port Rate	Port	IEEE 802.1d Short	IEEE 802.1t Long	IEEE 802.1t Long Standard
10 Mbps	Common port	100	2000000	2000000
	Aggregate interface	95	1900000	2000000/Linkupcnt
100 Mbps	Common port	19	200000	200000
	Aggregate interface	18	190000	200000/Linkupcnt
1000 Mbps	Common port	4	20000	20000
	Aggregate interface	3	19000	20000/Linkupcnt
10000 Mbps	Common port	2	2000	2000
	Aggregate interface	1	1900	20000/Linkupcnt

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure port path cost.

stp pathcost-standard method { **dot1t** | **dot1t standard** | **dot1d** }

The port path cost is calculated using the dot1 method in the IEEE 802.1t standard by default.

1.8.4 Configuring the Path Cost Value

1. Overview

In order to ignore the impact of the physical interface rate and strictly control the spanning tree calculation results, you can configure port path cost as a specified value. After configuration, the device no longer calculates the port path cost based on the physical interface rate.

2. Restrictions and Guidelines

- In MSTP application, you can configure different path costs for different instances on an interface so that different instance spanning trees are generated for the instances to implement load sharing.
- **mst** *instance-id*: Specifies the instance ID. The value range is from 0 to 64 and the default value is **0**.
- **cost** *cost*: Specifies the path cost value. The value range is from 1 to 200000000. A larger value of *cost* indicates a greater path cost.

3. Prerequisites

- Configure the interface as a Layer 2 interface.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(1) Enter the global configuration mode.

config-system terminal

(2) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(3) Configure the port path cost.

stp [**mst** *instance-id*] **cost** *cost*

The port path cost is calculated based on the interface link rate by default.

1.9 Configuring Spanning Tree Time Parameters

1.9.1 Overview

The timer in STP affects the spanning tree election and recovery performance.

You can use the stp command to configure the following time parameters. Each configuration command can only carry one parameter. Only one of the following time parameters can be carried in the stp command:

- Interval for port state change (**Forward Time**)
- BPDU transmission interval (**Hello Time**)
- BPDU expiration time (**Max Age**)
- Number of BPDUs sent per second (**TX Hold Count**)
- Spanning tree timeout factor (**Factor**)

1.9.2 Restrictions and Guidelines

- **forward-time** *forward*: Specifies the port status change interval, in seconds. It refers to the interval for a port to transition from listening state to learning state or from learning state to forwarding when STP runs on a device or RSTP runs in STP compatible mode. It indicates the delay in the port status change. The value range is from 4 to 30, and the default value is **15**. If **Forward Delay** is too small, a temporary loop may occur. If **Forward Delay** is too large, the network recovery time may be very long.
- **hello-time** *hello*: Indicates the interval for the device to send BPDUs, in seconds. The value range is from 1 to 10, and the default value is **2**. If **Hello Time** is too small, the resource consumption of the device and network may increase. If the value is too small, the downstream device may mistakenly consider the uplink to be faulty, especially when packet loss occurs occasionally.
- **max-age** *age*: Indicates the maximum expiration time of BPDUs, in seconds. When the time expires, BPDUs will be discarded. The value range is from 6 to 40, and the default value is **20**. If **Max Age** is too small, link congestion may be wrongly judged as a link fault and the spanning tree is calculated very frequently. If the value is too large, the link fault cannot be found in time, resulting in poor network auto-recovery capability.

  Each device running STP or RSTP is a separate region, and is not allocated to a region with any other devices. Though MSTP uses the hop count mechanism to represent the lifecycle of BPDUs, the **Message Age** and **Max Age** mechanisms are still retained in order to be compatible with STP and RSTP outside regions.
- The value ranges of **forward-time, hello-time, and max-age** are related. Changing one of them will affect the value ranges of the other two parameters. Their restrictive relationship is as follows: 2 × (Hello Time + 1s) ≤ Max Age ≤ 2 × (Forward Delay – 1s). The configured three parameters must meet this condition. Otherwise, the topology may be unstable. In addition, the three parameters need to match the network size. A larger network diameter needs larger values of the parameters. You are advised to adopt automatic calculation according to the protocol to avoid manual configuration.
- **tx-hold-count** *count*: Configures the maximum number of BPDUs that can be sent per second. The value range is from 1 to 10 and the default value is **3**.
- **timer-factor** *factor*: Configures the packet receiving timeout factor. Timeout time = Timeout factor (indicated by *factor*) × Hello Time. If a device fails to receive a BPDU from the upstream device within the timeout time,

the spanning tree is re-calculated. The value range is from 1 to 30, and the default value is **3**.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the interval for the status change of a spanning tree port.

stp forward-time *forward*

The default interval for the status change of a spanning tree port is 15s.

(4) Configure the interval for a spanning tree port to send BPDUs.

stp hello-time *hello*

The default interval for sending BPDUs is 2s.

(5) Configure the maximum number of BPDUs that can be sent by a spanning tree port.

stp tx-hold-count *count*

Three BPDUs are sent per second by default.

(6) Configure the maximum timeout time of spanning tree BPDUs.

stp max-age *age*

The default expiration time of BPDUs is 20s.

(7) Configure the spanning tree timeout factor.

stp timer-factor *factor*

The default spanning tree timeout factor is 3.

1.10 Configuring the Maximum Hop Count for BPDUs

1.10.1 Overview

MST regions do not use the Message Age and Max Age mechanisms to calculate whether BPDUs time out but use the maximum hop count mechanism for calculation. The maximum hop count mechanism is similar to the Time to Live (TTL) of IP packets. The maximum hop count of BPDUs affects the BPDU lifetime, and thus affects the network topology.

In an MST region, the BPDU hop count decreases by 1 each time the BPDU passes through one device from the region root until the remaining hop count is 0, indicating that the BPDU times out. Devices discard received BPDUs with the hop count of 0.

The maximum hop count of BPDUs is 20 by default. The value does not need to be changed for a network with the scale less than 20 hops, but needs to be changed to match the actual network situation when the network scale is greater than 20 hops. Changing the maximum hop count will affect all instances.

1.10.2 Restrictions and Guidelines

Hop-count:-count: Indicates the number of devices that a BPDU can pass through before it is discarded. The value range is from 1 to 40 and the default value is 20.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum hop count for BPDUs.

stp max-hops *hop-count*

The maximum hop count of BPDUs is 20 by default.

1.11 Configuring Fast Convergence of Spanning Trees

1.11.1 Overview

In a full-duplex point-to-point link, an interface supporting RSTP first enters discarding state after port role election is complete, and then performs a Proposal/Agreement handshake with the connected bridge to quickly enter forwarding state.

Only ports in point-to-point connection state support fast convergence. Ports in shared connection state do not support fast convergence. See1.1.4 5 Fast Convergence of RSTP.

1.11.2 Restrictions and Guidelines

- If the connection type is not configured, the device automatically sets the connection type of an interface based on the duplex state of the interface. When an interface works in full-duplex mode, the connection type of the interface is point-to-point. When an interface works in half-duplex mode, the connection type of the interface is shared (non point-to-point).
- You can forcibly configure the interface connection type. The parameter **point-to-point** indicates the point-to-point connection type and **shared** indicates the shared connection type.
- You can use the **display stp** [**mst** *instance-id*] { *port-index* | **interface** *interface-type interface-number* [**link-type**] } command to view the configuration results.

1.11.3 Prerequisites

- Configure the interface as a Layer 2 interface.

1.11.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure fast convergence for spanning trees.

stp link-type { **point-to-point** | **shared** }

By default, if an interface works in full-duplex mode, the connection type is point-to-point. If an interface works in half-duplex mode, the connection type is shared.

1.12 Enabling Protocol Migration

1.12.1 Overview

1. Compatibility of Different Protocols

If a device running STP receives an RST BPDU or MST BPDU, it ignores or discards the BPDU, resulting in incompatibility between the device running STP and other devices running RSTP or MSTP.

However, a device running RSTP or MSTP can recognize and process STP protocol packets from other devices. RSTP and MSTP are capable of automatically determining the protocol type supported by the connected bridge based on the version of received BPDUs. If the peer supports STP, after port role election is complete, RSTP and MSTP wait for twice the period of Forward Delay (30s by default) before entering forwarding state according to the STP status transition mechanism.

RSTP can process the CIST part of MST BPDUs. Therefore, MSTP does not have to send RST BPDUs to ensure RSTP compatibility.

Each device running STP or RSTP is a separate region, and does not belong to the same region with any other devices.

2. Protocol Migration Principle

When both RSTP and STP are used, a problem shown in Figure 1-16 may occur: Device A supports RSTP, Device B supports only STP, and the two devices are connected to each other. Device A judges that the peer device supports only STP, and sends STP BPDUs to ensure compatibility with the peer device. After Device B is replaced with Device C supporting RSTP, Device A still runs STP and sends STP BPDUs to the peer device. In this case, Device C mistakenly judges that Device A supports only STP. Finally, both devices run STP and the spanning tree convergence speed is greatly reduced. Hence, the protocol migration function emerges.

The protocol migration function enabled on Device C clears the original protocol and forces Device C to send RST BPDUs. Device A switches to the RSTP mode when finding that the connected bridge supports RSTP. Then, the two devices run RSTP.

When finding that the peer device supports RSTP, the user can enable the protocol migration function to force the local device to send RST BPDUs to achieve protocol migration of the local and peer devices.

Figure 1-16 Forcing Protocol Migration to RSTP



1.12.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Clear the original protocol and force the device to migrate to the RSTP protocol.

cleanup stp detected-protocols [**interface** *interface-type interface-number*]

1.13 Configuring Spanning Tree Compatibility for Interfaces

1.13.1 Overview

After the spanning tree compatibility function is enabled on an interface, STP calculates whether the interface participates in the calculation of a specified instance based on the VLAN, to which the interface belongs, and the mapping between the VLAN and the instance. When the interface sends a BPDU, only the MSTI configuration message of the instance calculated by the interface, is carried to ensure compatibility with other devices.

For example, instances 0 (exists by default), 1 and 2 exist on a device. Port GigabitEthernet 0/1 belongs only to VLAN 10, and VLAN 10 belongs to instance 1. If the spanning tree compatibility mode is enabled on port GigabitEthernet 0/1, the BPDU sent by port GigabitEthernet 0/1 carries only information of instance 0 (the port participates in calculation of this instance by default) and instance 1, with no information of instance 2.

1.13.2 Prerequisites

- Configure the interface as a Layer 2 interface.

1.13.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure spanning tree compatibility for an interface.

stp compatible enable

The spanning tree compatibility function of interfaces is disabled by default.

1.14 Enabling the STP Function

1.14.1 Overview

Unless otherwise specified, STP should be enabled on each device.

1.14.2 Restrictions and Guidelines

STP and the Transparent Interconnection of Lots of Links (TRILL) protocol of data centers are mutually exclusive.

1.14.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the STP function.

stp

The STP function is enabled by default.

1.15 Enabling the V-STP Function

Note

This function is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.15.1 Overview

The Multichassis Link Aggregation Group (M-LAG) is a virtualization technology that aggregates inter-device links. The M-LAG system consists of two independent devices. Access devices that support link aggregation group connect to this system through aggregate interfaces. The system virtualizes two devices into one logical device to improve the reliability from link level to device level. M-LAG implements only Layer 2 virtualization. The two devices are still independent devices. They can be upgraded independently without affecting the normal operation of the other device.

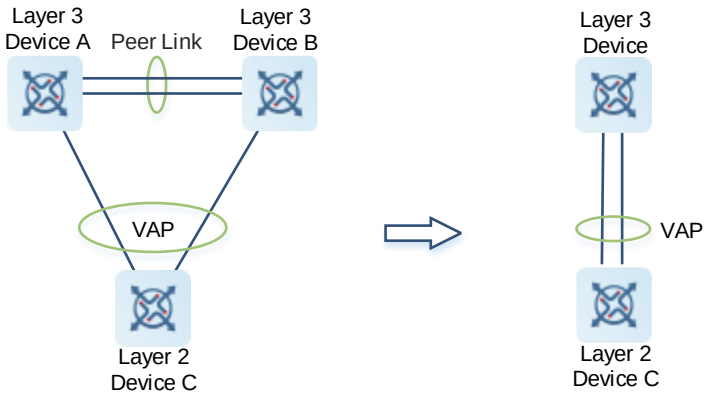
V-STP is not a protocol, but a technology that virtualizes multiple M-LAG devices into a network bridge to participate in spanning tree calculation, without affecting the implementation of the STP, RSTP, and MSTP protocols. The protocol packets and interface status depend on the spanning tree working mode STP, RSTP or MSTP.

V-STP is dedicated for M-LAG scenarios, which can meet the requirements of Layer 2 loop protection in single and multi-level M-LAG scenarios. You can use the stp v-stp enable command to enable the V-STP function.

1. Basic concepts of M-LAG

Figure 1-1 shows the basic topology of M-LAG.

Figure 1-1 Basic Topology of M-LAG



- Domain ID
  Domain ID is the unique identifier of the M-LAG system to distinguish different M-LAGs. Two devices in the M-LAG are paired by synchronizing with each other and comparing their domain ID. If the domain ID of the two devices are the same, an M-LAG system is formed.
- Virtual Aggregation Port (VAP)
  Device A and Device B form a pair of virtual AP groups with a pair of aggregate interface connected to the same access device Device C, also known as VAP member ports, for dual access of access devices. Only a Layer 2 aggregate interface can join the virtual AP group. Different access devices access different virtual groups. For Device C, Device A and Device B are logically the same device.
- Peer Link
  The link between Device A and Device B is called the Peer Link, which is used to synchronize data and transmit some traffic. The interface directly connected to both ends of the Peer Link is called the Peer Link interface. The Peer Link interface must be configured as an aggregate interface. In order to enhance the reliability of data transmission between M-LAG devices, TCP is used to transmit synchronous data between devices. Therefore, it is necessary to configure an SVI on the Peer Link link, and to configure Layer 3 reachability between M-LAG master and slave devices.
- Active and standby devices
  After the devices in the M-LAG system are paired successfully by identifying the domain ID, the active device and the standby device are selected through negotiation by synchronizing and comparing the M-LAG priority and the device MAC address. The device with higher M-LAG priority is selected as the active device. If the M-LAG priorities are the same, the device with a smaller MAC address is selected as the master device.

2. Working principle of V-STP

If the V-STP function is enabled on both the M-LAG active and standby devices, the active and standby devices will be virtualized into a network bridge and participate in the spanning tree calculation of interconnected devices outside the M-LAG domain, so as to achieve the purpose of spanning tree virtualization and block Layer 2 loops in the scenario.

The V-STP function can be enabled in STP, RSTP and MSTP modes. For the V-STP function to take effect, the following preconditions must be met:

- Enable the spanning tree function and V-STP function on the network bridge with M-LAG function enabled.
- M-LAG member devices can be connected successfully. See *Configuring M-LAG* in *Reliability Configuration*

Guide for details.

- The spanning tree configuration of M-LAG member devices must be the same, that is, the MST configuration information, bridge priority, port priority and port path cost of the VAP group must be the same. Otherwise, the V-STP function will not take effect.

The M-LAG active and standby devices with the V-STP function enabled send consistent BPDU packets as a virtual bridge, called V-STP BPDU. Like ordinary BPDU packets, the packets contain the priority vector <Root ID, Root Path Cost, Bridge ID, Port ID>. Interconnected devices outside the M-LAG domain determine the priority of V-STP BPDU by comparing the size of priority elements in the vector in turn.

- M-LAG active and standby devices have the same bridge priority. For example, in Figure 1-1, the bridge priority of Device A and Device B must be the same and higher than that of Device C. After the M-LAG devices successfully establish the connection and complete the active and standby negotiation, the system selects the device with a smaller MAC address from the M-LAG active and standby devices and use the device MAC address as the MAC address of the virtual bridge, and generates the Bridge ID based on the bridge priority to participate in the spanning tree calculation externally.
- The port priority of the VAP group is the same (the default value can be used), and the system selects a smaller port ID to generate the port ID of the VAP group.
- The port path cost of the VAP group is the same (the default value can be used), so the root path cost calculated on the active and standby devices is the same.

The active and standby devices calculate the spanning tree respectively, and then periodically synchronize the spanning tree results calculated by the device to the peer M-LAG member devices. Finally, by comparing the spanning tree calculation results of the device and the peer M-LAG member, the device with a lower path cost is selected as the forwarding path to block redundant links.

1.15.2 Restrictions and Guidelines

- In global configuration mode, run the **stp v-stp enable** command to enable the V-STP function or run the **unset stp v-stp enable** command to disable the V-STP function.

1.15.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the V-STP function.

stp v-stp enable

The V-STP function is disabled by default.

1.16 Configuring Port Protection

1.16.1 Overview

A root bridge may receive a configuration BPDU with a higher priority due to a misconfiguration or a malicious attack using illegitimate packets in the network, and lose the current root bridge role. This can cause an

incorrect network topology change. To prevent this situation, you can configure the root guard function on designated ports of the root bridge.

If the root port, master port, and alternate port on a non-root bridge periodically receive BPDUs with a higher priority from the upstream bridge, the spanning tree topology will keep unchanged. If they fail to receive high-priority BPDUs from the upstream bridge, they judge that the upstream device is faulty. Then, the bridge considers itself as the root and sets all ports as designated ports, which cause loops. To prevent this situation, you can configure the loop guard function on the above ports to enhance the device stability.

1.16.2 Restrictions and Guidelines

A device cannot be a root bridge and a non-root bridge at the same time. Therefore, root guard and loop guard cannot take effect on a port at the same time.

1.16.3 Configuration Tasks

The configuration includes the following tasks. The configuration steps below are mutually exclusive. Select one of them for configuration.

- Configuring Root Guard
- Configuring Loop Guard

1.16.4 Configuring Root Guard

1. Overview

In the network design, the root bridge and backup root bridge are usually classified into the same region. Designated ports on a root bridge may receive configuration BPDUs with a higher priority due to a misconfiguration or a malicious attack, and the root bridge loses the current root bridge role. As a result, incorrect network topology change is incurred. To prevent this situation, you can configure the root guard function on designated ports of the root bridge.

After the root guard function is enabled, the device interfaces are designated ports in all instances. If a port receives a high-priority BPDU, the port enters blocking state due to root-inconsistent. If the port fails to receive a high-priority BPDU within a period of time, it returns to the normal state.

When a port enters blocking state due to root guard, you can manually restore the port to the normal state by using two methods:

- Run the **unset stp guard root** command to disable the root guard function on the port.
- Run the **stp guard none** command to disable the guard function on the port.

⚠ Caution

- When the root guard function is configured on a non-designated port, the function forcibly sets the non-designated port as a designated port. The port will inevitably receive a high-priority BPDU according to the network design, and the port will enter blocking state due to root-inconsistent. Therefore, configuring root guard on a non-designated port can cause a communication failure.
- If a port receives a BPDU with a higher priority in an instance, it enters blocking state in this instance. If a port receives a BPDU with a higher priority in instance **0**, the port enters blocking state in all instances.

2. Restrictions and Guidelines

- In Layer 2 Ethernet interface configuration mode or Layer 2 aggregate port configuration mode, run the **stp**

guard root command to enable the root guard function on an interface, or run the **unset stp guard root** command to disable the root guard function of the interface.

- When a port is blocked due to root guard, you can run the **stp guard none** command in Layer 2 Ethernet interface configuration mode or Layer 2 aggregate interface configuration mode to disable the root guard function and restore the port to the normal state.

3. Prerequisites

- Configure the interface as a Layer 2 interface.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Enable the root guard function.

stp guard root

The root guard function is disabled on an interface by default.

1.16.5 Configuring Loop Guard

1. Overview

The root port or alternate port of a non-root bridge may fail to receive BPDUs due to a unidirectional link failure, and the port becomes a designated port and enters forwarding state. As a result, loops occur in the network. To prevent this situation, you can configure loop guard on a non-root bridge.

After loop guard is enabled, when the root port or alternate port fails to receive BPDUs and changes to a designated port, the port will remain in discarding state until it receives a BPDU for spanning tree calculation.

⚠ Caution

- Before the MSTP process is restarted, a port enters the loop guard blocking state. After the MSTP process is restarted, if the port still fails to receive BPDUs, the port changes to a designated port and enters forwarding state, causing loops in the network. Therefore, you are advised to figure out the cause that a port enters the loop guard blocking state and rectify the fault before the MSTP process is restarted. Otherwise, the spanning tree topology is still abnormal after the MSTP process is restarted.

2. Restrictions and Guidelines

- You can configure the loop guard function in the global configuration mode or interface configuration mode. Please select one configuration mode to configure this function.

- In global configuration mode, run the **stp loopguard default** command to enable the loop guard function on all Layer 2 Ethernet interface and Layer 2 aggregate interfaces, or run the **unset stp loopguard default** command to disable the loop guard function of all Layer 2 Ethernet interface and Layer 2 aggregate interfaces.
- In Layer 2 Ethernet interface or Layer 2 aggregate interface configuration mode, run the **stp guard loop** command to enable the loop guard function on an interface or run the **unset stp guard loop** command to disable the loop guard function of the interface.

3. Prerequisites

- Configure the interface as a Layer 2 interface.

4. Procedure (global configuration mode)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the loop guard function on all Layer 2 Ethernet interface and Layer 2 aggregate interfaces.

stp loopguard default

The loop guard function is enabled on all Layer 2 Ethernet interface and Layer 2 aggregate interfaces by default.

5. Procedure (interface configuration mode)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.
  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }
- Enter the Layer 2 aggregate interface configuration mode.
  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Enable the loop guard function.

stp guard loop

The loop guard function is disabled on an interface by default.

1.17 Configuring BPDU Source MAC Address Check

1.17.1 Overview

This section describes how to enable the BPDU source MAC address check function to prevent BPDU attacks. After the function is enabled on a device, the device receives only BPDU frames from specified source MAC addresses and discards other BPDU frames.

1. BPDU Source MAC Address Check

If the peer device connects to the local device in a point-to-point manner and the MAC address of the peer device is certain, the BPDU source MAC address check function can be configured on the local device. After this function is enabled, the device receives only BPDU frames matching the designated source MAC address and discards all the other BPDU frames. Therefore, when the device encounters BPDU packet attacks, illegitimate BPDU packets can be identified and discarded to prevent the MSTP function failure due to the attacks.

2. Filtering of BPDUs with Invalid Length

If the Ethernet length field in a BPDU frame exceeds 1,500, this illegitimate BPDU frame will be discarded.

1.17.2 Restrictions and Guidelines

- In Layer 2 Ethernet interface configuration mode or Layer 2 aggregate interface configuration mode, run the **bpdu src-mac-check** *mac-address* command to enable the BPDU source MAC address check function on an interface, or run the **unset bpdu src-mac-check** command to disable the BPDU source MAC address check function of the interface.
- *mac-address* indicates that only BPDU frames with the source MAC address matching this address are received. Only one filter MAC address can be configured for one interface.
- If the configured MAC address is incorrect, required BPDU frames may be discarded.

1.17.3 Prerequisites

Configure an interface as a Layer 2 interface.

1.17.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Enable BPDU source MAC address check so that only BPDUs with the source MAC address matching *mac-address* are received.

bpdu src-mac-check *mac-address*

The BPDU source MAC address check function is disabled on an interface by default.

1.18 Configuring an Edge Port

1.18.1 Overview

After the spanning tree role election is complete, a designated port needs to wait for twice the period of Forward Delay (2 × 15 = 30s) before entering forwarding state. Designated ports can be converted into edge ports via automatic recognition or manual configuration. Edge ports can rapidly enter forwarding state without waiting for twice the period of Forward Delay. If an edge port receives a BPDU, it transitions from forwarding state to disabled state, changes to a non-edge port, and participates in the spanning tree calculation.

- Autoedge

  If the autoedge function is enabled on a designated port and the port fails to receive a BPDU from the downstream device within a certain period of time (3s), the port judges that the connected peer device is a network terminal. Then, the designated port changes to an edge port and rapidly enters forwarding state.

- Port fast

  If you are sure that a designated port is directly connected to a network terminal, you can manually configure port fast rather than rely on the autoedge function. A manually configured edge port rapidly enters forwarding state without waiting for twice the period of **Forward Delay**.

1.18.2 Restrictions and Guidelines

When autoedge conflicts with the edge port manually configured via port fast, the manual configuration prevails.

Unless otherwise specified, do not disable the autoedge function.

- You can use the **display stp** [**mst** *instance-id*] { *port-index* | **interface** *interface-type interface-number* [**edged-port**] } command to check the configuration result.

1.18.3 Configuration Tasks

The configuration includes the following tasks:

- Configuring Autoedge
- Configuring the Port Fast Attribute

1.18.4 Configuring Autoedge

1. Overview

After the spanning tree role election is complete, a designated port needs to wait for twice the period of Forward Delay (2 × 15 = 30s) before entering forwarding state. If the autoedge function is enabled on a designated port and the port fails to receive a BPDU from the downstream device within a certain period of time (3s), the port judges that the connected peer device is a network terminal. Then, the designated port changes to an edge port and directly enters forwarding state. If an edge port receives a BPDU, it transitions from forwarding state to disabled state, changes to a non-edge port, and participates in the spanning tree calculation.

⚠ Caution

- STP does not support this function.
- This function applies only to designated ports, and takes effect only in the process of fast negotiation between a designated port and a downstream port and the designated port has not entered forwarding state in this process. If a designated port has been in forwarding state, the autoedge function does not take effect. For example, if the BPDU filter function is enabled on a port and the port has transitioned to

forwarding state, the autoedge function does not take effect. It takes effect only when the designated port restarts fast negotiation (for example, the network cable is removed and then inserted).

2. Restrictions and Guidelines

You can use the stp autoedge disabled command to disable the autoedge function.

3. Prerequisites

- Configure the interface as a Layer 2 interface.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure the autoedge function.

stp autoedge [**disabled**]

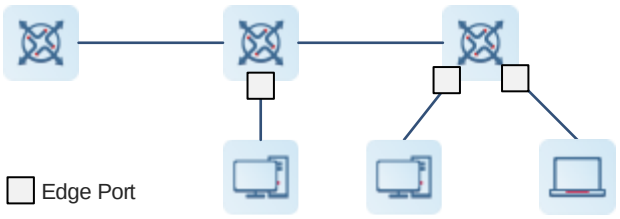
The autoedge function is enabled by default.

1.18.5 Configuring the Port Fast Attribute

1. Overview

After the spanning tree role election is complete, a designated port needs to wait for twice the period of Forward Delay (2 × 15 = 30s) before entering forwarding state. The autoedge function is enabled on a designated port by default. If the port fails to receive a BPDU from the downstream device within a period of time (3s), it is automatically recognized as an edge port. If a designated port fails to receive a BPDU within the timeout time due to poor network condition, packet loss, or packet transmission/receiving delay, the device mistakenly judges that the designated port connects to a network terminal and automatically identifies the designated port as an edge port. In this network condition, you are advised to disable the autoedge function, and manually configure the interface connected to a terminal as an edge port according to the network plan. When you are sure that a designated port is directly connected to a network terminal, configure the port fast attribute manually rather than rely on the autoedge function. A manually configured edge port enters forwarding state without waiting for twice the period of Forward Delay. If an edge port receives a BPDU, it transitions from forwarding state to disabled state, changes to a non-edge port, and participates in the spanning tree calculation.

Figure 1-17 Positions of Edge Ports



2. Restrictions and Guidelines

- Only a Layer 2 Ethernet interface or Layer 2 aggregate interface can be configured as an edge port.
- In global configuration mode, you can use the **stp edged-port default** command to configure all Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces as edge ports, use the **unset stp edged-port default** command to restore all Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces as non-edge ports.
- In Layer 2 Ethernet interface configuration mode or Layer 2 aggregate interface configuration mode, you can use the **stp edged-port** command to configure this interface as an edge port, and use the **stp edged-port disabled** command to restore this interface as a non-edge port.

3. Prerequisites

- Configure the interface as a Layer 2 interface.

4. Procedure (global configuration mode)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

- Configure all Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces as edge ports.

  stp edged-port default

  by default, an interface is a non-edge port.

5. Procedure (interface configuration mode)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Configure this interface as an edge port.

stp edged-port

1.19 Configuring BPDU Guard or BPDU Filter

1.19.1 Overview

If a device interface is directly connected to a network terminal, BPDU guard or BPDU filter can be configured on the interface to prevent illegitimate access, BPDU attacks, or loops in downstream devices. BPDU guard and BPDU filter can be enabled globally and take effect only on edge ports, or enabled and take effect only on specific interfaces.

1.19.2 Restrictions and Guidelines

BPDU filter has a higher priority than BPDU guard. If both BPDU filter and BPDU guard are enabled on an interface, only BPDU filter takes effect.

- You can use the **display stp** [**mst** *instance-id*] { *port-index* | **interface** *interface-type interface-number* [**bpdufilter** | **bpduguard**] } command to check the configuration result.

1.19.3 Configuration Tasks

The configuration includes the following tasks:

- Configuring BPDU Guard
- Configuring BPDU Filter

1.19.4 Configuring BPDU Guard

1. Overview

Edge ports do not receive BPDUs in normal cases. If an edge port is attacked by forged BPDUs or an illegitimate device is added to the network, the edge port will receive BPDUs, change to a non-edge port, and participate in spanning tree calculation, resulting in network flapping.

The BPDU guard function aims to defend against BPDU packet attacks and illegitimate access. If this function is enabled, a port enters the error-disabled state when receiving a BPDU, indicating that a port exception occurs. Then, the device disables the port, which can be restored to the normal state only through the errdisable recovery function.

2. Restrictions and Guidelines

BPDU guard can be enabled globally or on specific interfaces. The global BPDU guard function takes effect only on edge ports. The BPDU guard function takes effect on interfaces regardless of whether the interfaces are edge ports.

- In the global BPDU guard configuration command, the **unset** parameter is used to disable the global BPDU guard function.
- In the interface BPDU guard configuration command, the **disabled** parameter is used to disable the BPDU guard function on a specific interface.
- A port in error-disabled state can be restored automatically or manually. The **errdisable recovery** [**interval** *seconds*] command is used to configure the auto-recovery interval, in seconds. The value range is from 30 to 86400. The errdisable recovery function is disabled on a port by default and no default recovery interval is available. To restore a port in error-disabled state manually, run the **errdisable recovery** command. See *Configuring Ethernet Interface* in the *Interface Configuration Guide*.

3. Prerequisites

- Configure the interface as a Layer 2 interface.

4. Procedure (global configuration mode)

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

- Enable the global BPDU guard function. The function takes effect only on edge ports.
stp edged-port bpduguard default

  By default, the global BPDU guard function is disabled.

5. Procedure (interface configuration mode)

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the interface configuration mode.
- Enter the Layer 2 Ethernet interface configuration mode.
interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }
- Enter the Layer 2 aggregate interface configuration mode.
interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Enable the BPDU guard function.
stp bpduguard enabled

By default, the interface BPDU guard function is disabled.

1.19.5 Configuring BPDU Filter

1. Overview

BPDU filter is a method of preventing BPDU attacks. When BPDU filter is enabled, a port neither sends nor receives BPDUs, but directly enters forwarding state. If a port receives a BPDU, it transitions from forwarding state to disabled state, the BPDU filter function automatically fails, and the port participates in spanning tree calculation.

2. Restrictions and Guidelines

BPDU filter can be enabled globally or on specific interfaces. The global BPDU filter function takes effect only on edge ports. The BPDU guard function takes effect on interfaces regardless of whether the interfaces are edge ports.

- In the global BPDU filter configuration command, the **unset** parameter is used to disable the global BPDU filter function.
- In the interface BPDU filter configuration command, the **enabled** parameter is used to enable the BPDU filter function and the **disabled** parameter is used to disable the BPDU filter function.

3. Prerequisites

- Configure the interface as a Layer 2 interface.

4. Procedure (global configuration mode)

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

- Enable the global BPDU filter function. The function takes effect only on edge ports.
  stp edged-port bpdufilter default

  By default, the global BPDU filter function is disabled.

5. Procedure (interface configuration mode)

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.
  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }
- Enter the Layer 2 aggregate interface configuration mode.
  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Enable the BPDU filter function.
stp bpdufilter enabled

By default, the interface BPDU filter function is disabled.

1.20 Configuring TC Attack Defense Functions

1.20.1 Overview

When the downstream network topology in port forwarding state changes (from Blocking to Forwarding, for example), a port generates a TC packet (that is, TCN BPDU; see Table 1-4 for details) to notify the upstream device of the spanning tree change. After receiving the TCN BPDU, the port copies the BPDU and forwards it to upstream devices until the root bridge receives the BPDU. This is called TC diffusion. After receiving a TC packet, a device deletes dynamic MAC addresses and ARP entries that have been learned, and a Layer 3 device also enables the fast forwarding module to change the port status of ARP entries.

If a device is attacked by forged TC packets, it frequently performs the deletion operation, which occupies excessive device resources and makes the device overburdened. After TC packet attacks diffuse to the whole network, the performance of devices throughout the network will deteriorate and the network stability will be affected. Therefore, TC protection, TC guard, and TC filter arise to solve this problem.

- TC protection

After TC protection is enabled on a device, the device performs only one deletion operation within a period of time (generally 4s) after receiving TC packets, and monitors whether any TC packet is received in this period. If

the device receives TC packets in this period, it performs another deletion operation after the period expires, to prevent frequent deletion of MAC address entries and ARP entries.

- TC guard

TC protection can reduce the frequency of deleting dynamic MAC address entries and ARP entries, but many deletion operations still need to be performed in the case of TC packet attacks, which affects the normal operation of devices. In addition, TC packets still are diffused in the network and the entire network is affected. For this, the TC guard function emerges.

After TC guard is enabled on a port, TC packets generated or received by the port will not be diffused to other ports of the device that participate in spanning tree calculation, so as to effectively control possible TC attacks in the network and retain the network stability. Especially on Layer 3 devices, this function can effectively prevent interruption of core routes caused by access device flapping.

- TC filter

TC guard blocks the diffusion of all TC packets, including normal TC packets generated in the case of a topology change. As a result, devices in the network cannot perceive the topology change and fail to clear MAC address entries and ARP entries of relevant ports in time, and errors occur in data forwarding. Hence, the TC filter function emerges.

TC filter refers that a port does not process TC packets received from other ports but processes TC packets generated due to the normal topology change of the port.

This function solves the problems of MAC address entry clearing and core route interruption caused by frequent up/down state switching of ports with port fast not configured. It also ensures that core routing entries are updated in time when a topology change occurs.

1.20.2 Restrictions and Guidelines

It is recommended that the TC attack defense functions be enabled only when the network is attacked by TC packets.

1.20.3 Configuration Tasks

The configuration includes the following tasks:

- Configuring TC Protection
- Configuring TC Guard
- Configuring TC Filter

1.20.4 Configuring TC Protection

1. Overview

After TC protection is enabled, the device performs only one deletion operation within a period of time (generally 4s) after receiving TC packets. This can prevent frequent deletion of MAC address entries and ARP entries.

2. Restrictions and Guidelines

TC protection can be enabled or disabled only globally and takes effect on all Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces.

- Run the **stp tc-protection** command to enable TC protection, or run the **unset stp tc-protection** command

to disable TC protection.

3. Prerequisites

- Configure the interface as a Layer 2 interface.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable TC guard on all Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces.

stp tc-protection

TC guard is disabled globally by default.

1.20.5 Configuring TC Guard

1. Overview

After TC guard is enabled on a port, the port shields received TC packets and the TC packets generated by the port so that the TC packets are not diffused to other ports. In this way, possible TC attacks in the network are effectively contained, ensuring network stability.

The TC guard function is divided into global TC guard function and interface TC guard function.

- In the global configuration mode, TC guard function takes effect on all Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces if enabled. In this case, all ports will not diffuse TC packets to other ports. TC guard is suitable to be enabled on desktop access devices.
- In the Layer 2 Ethernet interface configuration mode or the Layer 2 aggregate interface configuration mode, the TC guard function only takes effect on the current interface. In this case, this interface will not diffuse TC packets to other ports. TC guard is suitable to be enabled on the uplink port, especially the port connecting the aggregation switch to the core switch.

2. Restrictions and Guidelines

- In global configuration mode, run the **stp tc-protection tc-guard** command to enable the TC guard function globally, or run the **unset stp tc-protection tc-guard** command to disable the TC guard function globally.
- In interface configuration mode, run the **stp tc-guard** command to enable the TC guard function on an interface, or run the **unset stp tc-guard** command to disable the TC guard function on an interface.
- If no TC attack exists in a network but TC guard is configured, the spread of normal TC packets is contained. When the topology changes, the device fails to clear learned MAC addresses and a packet forwarding error may occur. Enabling the TC guard function will also limit the transmission of normal TC packets. When the spanning tree network topology changes, the device cannot obtain the topology change information, and cannot clear MAC addresses in time, which may lead to packet forwarding errors. If the topology cannot be restored automatically, communication may be interrupted. Therefore, it is recommended to enable this function only when there is an illegal TC packet attack in the network.

3. Prerequisites

- Configure the interface as a Layer 2 interface.

4. Procedure (global configuration mode)

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

- Enable TC guard on all Layer 2 Ethernet interfaces and Layer 2 aggregate interfaces.
stp tc-protection tc-guard

  By default, TC guard is disabled globally.

5. Procedure (interface configuration mode)

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the interface configuration mode.
- Enter the Layer 2 Ethernet interface configuration mode.
interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }
- Enter the Layer 2 aggregate interface configuration mode.
interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Enable TC guard on the interface.
stp tc-guard

By default, TC guard is disabled on an interface.

1.20.6 Configuring TC Filter

1. Overview

TC filter enables a port not to process received TC packets but to process TC packets generated by the port due to topology changes.

2. Restrictions and Guidelines

TC filter can be enabled or disabled only on interfaces and takes effect on specific interfaces.

- In Layer 2 Ethernet interface configuration mode or Layer 2 aggregate interface configuration mode, run the **stp ignore tc** command to enable the TC filter function or run the **unset stp ignore tc** command to disable the TC filter function.

3. Prerequisites

Configure the interface as a Layer 2 interface.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(4) Enable TC filter on an interface.

stp ignore tc

TC filter is disabled on an interface by default.

1.21 Configuring BPDU Tunnel

Note

S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC do not support the BPDU tunnel function.

1.21.1 Overview

BPDU tunnel is a Layer 2 tunnel function based on the Layer 2 Tunnel Protocol (L2TP).

A 802.1Q (QinQ) network is usually divided into the customer network and service provider network. The basic principle of QinQ is described as follows: When a user packet is sent to a provider edge (PE), the PE considers the VLAN tag in the user packet as data, and encapsulates the VLAN tag of the service provider network into the packet so that the user packet carries two VLAN tags to traverse the service provider network. In the service provider network, the packet is forwarded based on the outer VLAN tag. When the packet leaves the service provider network, the outer VLAN tag is removed.

STP BPDUs use the BPDU dedicated address 0180.c200.0000 as the destination MAC address. Therefore, the BPDUs of the customer network and service provider network have the same destination MAC address. When a Dot1q-tunnel interface of a PE receives a BPDU from the customer network, the PE recognizes the BPDU as an STP packet and does not forward it because the destination MAC address of the BPDU is the BPDU dedicated address (0180.c200.0000). In addition, the PE performs spanning tree calculation based on the BPDU content, which affects the topology of the service provider network. In order to prevent the customer network BPDUs from interfering with the spanning tree calculation of the service provider network, you need to differentiate customer network BPDUs from service provider network BPDUs, that is, modify the destination MAC address in customer network BPDUs.

After the BPDU tunnel function is enabled on a Dot1q-tunnel interface of a PE, if the interface receives a customer network BPDU, it changes the destination MAC address in the BPDU from the BPDU dedicated address (0180.c200.0000) to the tunnel address (01d0.f800.0005 by default) and forwards the BPDU in the service provider network. When the BPDU reaches a PE at the other end, the PE restores the destination MAC address of the BPDU from the tunnel address (01d0.f800.0005 by default) to the BPDU dedicated address (0180.c200.0000) and forwards the BPDU to the peer customer network. Customer network BPDUs are transmitted through BPDU tunnels in the service provider network. In this way, the STP calculations of the

customer network and service provider network are performed separately without interfering with each other. For details, see *Configuring QinQ* and *Configuring VPDN*.

1.21.2 Restrictions and Guidelines

- This function takes effect only when it is enabled in both global configuration mode and interface configuration mode.
  - Run the **l2protocol-tunnel stp** command to enable the global BPDU tunnel function or run the **unset l2protocol-tunnel stp** command to disable the global BPDU tunnel function.
  - Run the **l2protocol-tunnel stp enable** command to enable the BPDU tunnel function or run the **unset l2protocol-tunnel stp enable** command to disable the BPDU tunnel function.
- Run the **l2protocol-tunnel stp tunnel-dmac** *mac-address* command to configure a BPDU tunnel address. The optional tunnel addresses of STP packets include 01d0.f800.0005 (default), 011a.a900.0005, 010f.e200.0003, 0100.0ccd.cdd0, 0100.0ccd.cdd1, and 0100.0ccd.cdd2.

1.21.3 Prerequisites

Configure the interface as a Layer 2 interface.

1.21.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the global BPDU tunnel function.

l2protocol-tunnel stp

The global BPDU tunnel function is disabled by default.

(4) (Optional) Configure a BPDU tunnel address.

l2protocol-tunnel stp tunnel-dmac *mac-address*

The default BPDU tunnel address is 01d0.f800.0005.

(5) Enter the interface configuration mode of the port that connects the customer network.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 2 aggregate interface configuration mode.

  interface { **aggregateport** *interface-number* | **range aggregateport** *interface-range* }

(6) Enable the STP BPDU tunnel function on the interface.

l2protocol-tunnel stp enable

The STP BPDU tunnel function is disabled on the interface by default.

1.22 Configuring BPDU Transparent Transmission

1.22.1 Overview

According to the IEEE 802.1Q standard, BPDUs use 0180.c200.0000 as the destination MAC address. When a device supporting IEEE 802.1Q receives a frame with the destination address of 0180.c200.0000, it recognizes the frame as a BPDU and will not forward it.

However, in the actual network deployment, some BPDU frames need to be transparently transmitted by devices. For example, STP is disabled on Device A but enabled on Device B and Device C connected through Device A. In this case, Device A needs to transparently transmit BPDU frames so that Device B and Device C can normally calculate and generate spanning trees.

1.22.2 Restrictions and Guidelines

- BPDU transparent transmission takes effect only when STP is disabled. If STP is enabled on a device, the device does not transparently transmit BPDU frames. If STP is enabled on a service provider network device and the device is required to transmit customer network BPDUs without affecting the spanning tree of the service provider network, use the BPDU tunnel function.
- In global configuration mode, run the **bridge-frame forwarding protocol bpdu** command to enable the BPDU transparent transmission function, or run the **unset bridge-frame forwarding protocol bpdu** command to disable the BPDU transparent transmission function.

1.22.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BPDU transparent transmission function.

bridge-frame forwarding protocol bpdu

BPDU transparent transmission is disabled by default.

1.23 Monitoring

Run the display commands to check the running status of a configured function to verify the configuration effect.

Run the debugging commands to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging switch immediately after use.

Run the cleanup commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-13 Monitoring

Command	Purpose
display stp [**forward-time** | **hello-time** | **max-age** | **max-hops** | **mst** *instance-id* | **pathcost-standard method** | **tx-hold-count**]	Displays the global spanning tree configuration.
display stp counters	Displays statistics on sent and received STP packets.
display stp inconsistentports	Displays ports blocked due to root guard or loop guard.
display stp summary	Displays the spanning tree topology and port forwarding status.
display stp v-stp information	Displays information about the V-STP function.
display stp [**mst** *instance-id*] **interface** *interface-type interface-number* [**bpdufilter** | **bpduguard** | **link-type** | **edged-port**]	Displays the spanning tree configuration and status information of an interface by interface type and number.
display stp [**mst** *instance-id*] *port-index*	Displays the spanning tree configuration and status information of an interface by interface number.
display stp mst configuration	Displays the MST region configuration.
display stp mst *instance-id* **topochange record**	Displays the topology change information of an interface in a specified instance.
display l2protocol-tunnel stp	Displays BPDU tunnel information.
debugging mstp all	Debugs STP, RSTP, and MSTP.
debugging mstp gr	Debugs the spanning tree GR function.
debugging mstp rx	Debugs the BPDU receiving.
debugging mstp tx	Debugs BPDU transmission.
debugging mstp event	Debugs spanning tree events.
debugging mstp loopguard	Debugs spanning tree loop guard.
debugging mstp rootguard	Debugs spanning tree root guard.
debugging mstp bridgedetect	Debugs the bridge detect state machine.
debugging mstp portinfo	Debugs the port information state machine.
debugging mstp protomigrat	Debugs the port protocol migration state machine.
debugging mstp topochange	Debugs the spanning tree topology changes.
debugging mstp receive	Debugs the spanning tree receiving state machine.
debugging mstp roletran	Debugs the port role transition state machine.

Command	Purpose
debugging mstp statetran	Debugs the port state transition state machine.
debugging mstp transmit	Debugs the MSTP transmission state machine.
cleanup stp counters [**interface** *interface-type interface-number*]	Clears statistics on packets sent and received by a port.
cleanup stp mst *instance-id* **topochange record**	Clears the STP topology change information.

1.24 Configuration Examples

1.24.1 Configuring MSTP

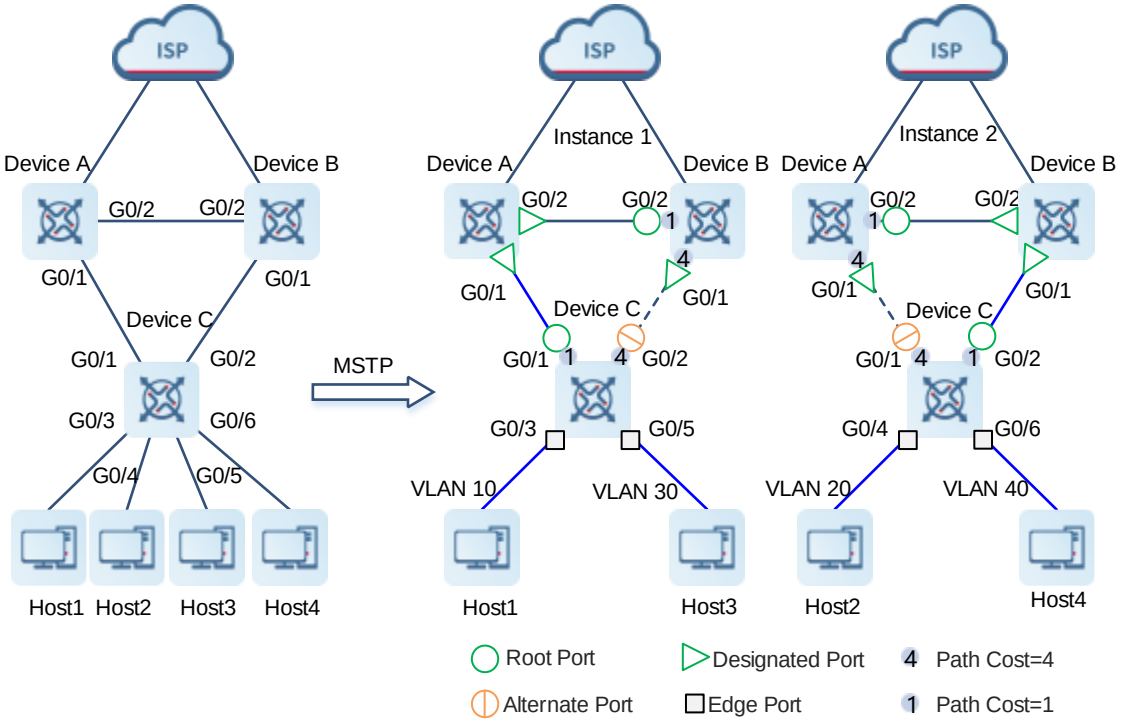
1. Requirements

The topology shown in Figure 1-18 includes two layers: Device A and Device B are in the core layer, Device C is in the access layer, and Device C connects to intranet terminals. There are four VLANs in the intranet. The configured MSTP needs to meet the following requirements:

- The spanning tree root bridge of VLAN 10 and VLAN 30 is Device A and data is forwarded through port GigabitEthernet 0/1 of Device C.
- The spanning tree root bridge of VLAN 20 and VLAN 40 is Device B and data is forwarded through port GigabitEthernet 0/2 of Device C.
- Enable loop guard on the root ports of Device C. Enable BPDU guard on the edge ports of Device C that connect to terminals.

2. Topology

Figure 1-18 Configuring Basic MSTP Function



3. Notes

- Create the same VLANs on Devices A, B, and C and configure the same instance mappings: Map instance 1 to VLANs 10 and 30 and map instance 2 to VLANs 20 and 40. Instance 0 contains other VLANs that have not been created.
- The devices elect the device roles and port roles in the spanning tree by comparing the priority vector <**Root Identifier**, **Root Path Cost**, **Bridge ID**, **Port ID**>. Configure the bridge priorities and port path costs for instances so that the required topology is calculated. For the ease of management, configure the same spanning tree for instance 0 and instance 1.

  Instance 0 and instance 1:

  - Set the bridge priority to 4096 for Device A and 8192 for Device B, and use the default bridge priority 32768 (no configuration is required) for Device C so that Device A becomes the root bridge.
  - On Device B, set the path cost to 1 for port GigabitEthernet 0/2 and 4 for port GigabitEthernet 0/1 so that port GigabitEthernet 0/2 becomes the root port of Device B.
  - On Device C, set the path cost to 1 for port GigabitEthernet 0/1 and 4 for port GigabitEthernet 0/2 so that port GigabitEthernet 0/1 becomes the root port of Device C.
  - The bridge priority of Device B is 8192, which is higher than that of Device C (32768). Therefore, port GigabitEthernet 0/1 of Device B is elected as the designated port and port GigabitEthernet 0/2 of Device C is elected as the alternate port.

  Instance 2:

  - Set the bridge priority to 4096 for Device B and 8192 for Device A, and use the default bridge priority

32768 (no configuration is required) for Device C so that Device B becomes the root bridge.

- On Device A, set the path cost to 1 for port GigabitEthernet 0/2 and 4 for port GigabitEthernet 0/1 so that port GigabitEthernet 0/2 becomes the root port of Device A.
- On Device C, set the path cost to 1 for port GigabitEthernet 0/2 and 4 for port GigabitEthernet 0/1 so that port GigabitEthernet 0/2 becomes the root port of Device C.
- The bridge priority of Device A is 8192, which is higher than that of Device C (32768). Therefore, port GigabitEthernet 0/1 of Device A is elected as the designated port and port GigabitEthernet 0/1 of Device C is elected as the alternate port.

- Configure MSTP loop guard on ports GigabitEthernet 0/1–0/2 of Device C. When the root port or backup port fails to receive BPDUs and changes to a designated port, the port will remain in discarding state until it receives BPDUs for spanning tree calculation.
- On Device C, configure ports GigabitEthernet 0/3–0/6 that connect to terminals as edge ports. The autoedge function is enabled by default. If ports GigabitEthernet 0/3–0/6 fail to receive BPDUs within 3s after being elected as designated ports, they are automatically recognized as edge ports and immediately enters forwarding state. Packet loss or packet transmission/receiving delay in the network may affect the autoedge function. Therefore, disable the autoedge function, manually configure the ports as edge ports, and enable the BPDU guard function.
- Enable the STP function globally on Devices A, B, and C and use the default MSTP mode.
- Configure interconnected ports among Devices A, and B, and C as trunk ports and configure the ports to allow all VLAN traffic to pass. Add the interfaces of Device C that connect to terminals to their respective VLANs.

4. Procedure

(1) Configure Device A.

Create VLANs and configure instance mappings.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan range 10,20,30,40
DeviceA(cfg-vlan-range)# exit
DeviceA(cfg)# stp mst configuration
DeviceA(cfg-mst)# instance 1 vlan 10,30
DeviceA(cfg-mst)# instance 2 vlan 20,40
```

Set the bridge priority to 4094 for instances 0 and 1 and 8192 for instance 2.

```
DeviceA(cfg-mst)# stp mst 0 priority 4096
DeviceA(cfg)# stp mst configuration
DeviceA(cfg-mst)# stp mst 1 priority 4096
DeviceA(cfg)# stp mst configuration
DeviceA(cfg-mst)# stp mst 2 priority 8192
```

Configure port GigabitEthernet 0/2 as a trunk port. In instance 2, set the path cost of the port to 1.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port mode trunk
```

```
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-GigabitEthernet 0/2)# stp mst 2 cost 1
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

Configure port GigabitEthernet 0/1 as a trunk port. In instance 2, set the path cost of the port to 4.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-GigabitEthernet 0/1)# stp mst 2 cost 4
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Enable the STP function globally.

```
DeviceA(cfg)# stp
DeviceA(cfg)# end
DeviceA# write
```

(2) Configure Device B.

Create VLANs and configure instance mappings.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan range 10,20,30,40
DeviceB(cfg-vlan-range)# exit
DeviceB(cfg)# stp mst configuration
DeviceB(cfg-mst)# instance 1 vlan 10,30
DeviceB(cfg-mst)# instance 2 vlan 20,40
```

Set the bridge priority to 4094 for instance 2 and 8192 for instance 0 and instance 1.

```
DeviceB(cfg-mst)# stp mst 0 priority 8192
DeviceB(cfg)# stp mst configuration
DeviceB(cfg-mst)# stp mst 1 priority 8192
DeviceB(cfg)# stp mst configuration
DeviceB(cfg-mst)# stp mst 2 priority 4096
```

Configure port GigabitEthernet 0/2 as a trunk port. In instances 0 and 1, set the path cost of the port to 1.

```
DeviceB(cfg)# interface gigabitethernet 0/2
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-GigabitEthernet 0/2)# stp mst 0 cost 1
DeviceB(cfg-if-GigabitEthernet 0/2)# stp mst 1 cost 1
DeviceB(cfg-if-GigabitEthernet 0/2)# exit
```

Configure port GigabitEthernet 0/1 as a trunk port. In instances 0 and 1, set the path cost of the port to 4.

```
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port
```

```
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-GigabitEthernet 0/1)# stp mst 0 cost 4
DeviceB(cfg-if-GigabitEthernet 0/1)# stp mst 1 cost 4
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Enable the STP function globally.

```
DeviceB(cfg)# stp
DeviceB(cfg)# end
DeviceB# write
```

(3) Configure Device C.

Create VLANs and configure instance mappings.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# vlan range 10,20,30,40
DeviceC(cfg-vlan-range)# exit
DeviceC(cfg)# stp mst configuration
DeviceC(cfg-mst)# instance 1 vlan 10,30
DeviceC(cfg-mst)# instance 2 vlan 20,40
DeviceC(cfg-mst)# exit
```

Configure uplink port GigabitEthernet 0/1 as a trunk port. In instances 0 and 1, set the path cost of the port to 1. In instance 2, set the path cost of the port to 4. Configure loop guard on the uplink port.

```
DeviceC(cfg)# interface gigabitethernet 0/1
DeviceC(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceC(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-GigabitEthernet 0/1)# stp mst 0 cost 1
DeviceC(cfg-if-GigabitEthernet 0/1)# stp mst 1 cost 1
DeviceC(cfg-if-GigabitEthernet 0/1)# stp mst 2 cost 4
DeviceC(cfg-if-GigabitEthernet 0/1)# stp guard loop
DeviceC(cfg-if-GigabitEthernet 0/1)# exit
```

Configure uplink port GigabitEthernet 0/2 as a trunk port. In instance 2, set the path cost of the port to 1. In instances 0 and 1, set the port path cost to 4. Configure loop guard on the uplink port.

```
DeviceC(cfg)# interface gigabitethernet 0/2
DeviceC(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/2)# switch-port mode trunk
DeviceC(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
DeviceC(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan all
DeviceC(cfg-if-GigabitEthernet 0/2)# stp mst 0 cost 4
DeviceC(cfg-if-GigabitEthernet 0/2)# stp mst 1 cost 4
DeviceC(cfg-if-GigabitEthernet 0/2)# stp mst 2 cost 1
DeviceC(cfg-if-GigabitEthernet 0/2)# stp guard loop
DeviceC(cfg-if-GigabitEthernet 0/2)# exit
```

Add downlink ports GigabitEthernet 0/3–0/6 to VLANs. Configure the ports as edge ports and configure BPDU guard.

```
DeviceC(cfg)# interface gigabitethernet 0/3
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port mode access
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port access vlan 10
DeviceC(cfg-if-GigabitEthernet 0/3)# stp autoedge disabled
DeviceC(cfg-if-GigabitEthernet 0/3)# stp edged-port
DeviceC(cfg-if-GigabitEthernet 0/3)# stp bpduguard enable
DeviceC(cfg-if-GigabitEthernet 0/3)# exit
DeviceC(cfg)# interface gigabitethernet 0/4
DeviceC(cfg-if-GigabitEthernet 0/4)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/4)# switch-port mode access
DeviceC(cfg-if-GigabitEthernet 0/4)# switch-port access vlan 20
DeviceC(cfg-if-GigabitEthernet 0/4)# stp autoedge disabled
DeviceC(cfg-if-GigabitEthernet 0/4)# stp edged-port
DeviceC(cfg-if-GigabitEthernet 0/4)# stp bpduguard enable
DeviceC(cfg-if-GigabitEthernet 0/4)# exit
DeviceC(cfg)# interface gigabitethernet 0/5
DeviceC(cfg-if-GigabitEthernet 0/5)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/5)# switch-port mode access
DeviceC(cfg-if-GigabitEthernet 0/5)# switch-port access vlan 30
DeviceC(cfg-if-GigabitEthernet 0/5)# stp autoedge disabled
DeviceC(cfg-if-GigabitEthernet 0/5)# stp edged-port
DeviceC(cfg-if-GigabitEthernet 0/5)# stp bpduguard enable
DeviceC(cfg-if-GigabitEthernet 0/5)# exit
DeviceC(cfg)# interface gigabitethernet 0/6
DeviceC(cfg-if-GigabitEthernet 0/6)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/6)# switch-port mode access
DeviceC(cfg-if-GigabitEthernet 0/6)# switch-port access vlan 40
DeviceC(cfg-if-GigabitEthernet 0/6)# stp autoedge disabled
DeviceC(cfg-if-GigabitEthernet 0/6)# stp edged-port
DeviceC(cfg-if-GigabitEthernet 0/6)# stp bpduguard enable
DeviceC(cfg-if-GigabitEthernet 0/6)# exit
```

Enable the STP function globally.

```
DeviceC(cfg)# stp
DeviceC(cfg)# end
DeviceC# write
```

5. Verification

(1) Verify that instance mappings on the devices are the same.

Check instance mappings on Device A.

```
DeviceA# display stp mst configuration
Multi spanning tree protocol : Enable
Name      :
```

```
Revision : 0
Instance    Vlans Mapped
-------- ---------------------------------------------
0            : 1-9, 11-19, 21-29, 31-39, 41-4094
1            : 10, 30
2            : 20, 40
---------------------------------------------------------
```

Check instance mappings on Device B.

```
DeviceB# display stp mst configuration
Multi spanning tree protocol : Enable
Name         :
Revision : 0
Instance    Vlans Mapped
-------- ---------------------------------------------
0            : 1-9, 11-19, 21-29, 31-39, 41-4094
1            : 10, 30
2            : 20, 40
---------------------------------------------------------
```

Check instance mappings on Device C.

```
DeviceC# display stp mst configuration
Multi spanning tree protocol : Enable
Name         :
Revision : 0
Instance    Vlans Mapped
-------- ---------------------------------------------
0            : 1-9, 11-19, 21-29, 31-39, 41-4094
1            : 10, 30
2            : 20, 40
---------------------------------------------------------
```

(2) Check the instance spanning tree topology and port forwarding status on Device A.

In instances 0 and 1, Device A (0074.9cee.f49e) is the root bridge. Ports GigabitEthernet 0/1–0/2 of Device A are designated ports. They are all in forwarding state.

```
DeviceA# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-29, 31-39, 41-4094
  Root ID       Priority       4096
                MAC Address         0074.9cee.f49e
                this bridge is root
                Hello Time     2 sec    Forward Delay 15 sec    Max Age 20 sec

  Bridge ID    Priority       4096
                MAC Address         0074.9cee.f49e
                Hello Time     2 sec    Forward Delay 15 sec    Max Age 20 sec
```

```
Interface            Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1                Desg FWD 200000        128         False       P2p
Gi0/2                Desg FWD 20000         128         False       P2p

MST 1 vlans map : 10, 30
  Region Root Priority     4096
                MAC Address        0074.9cee.f49e
                this bridge is region root

  Bridge ID   Priority       4096
                MAC Address        0074.9cee.f49e

Interface            Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1                Desg FWD 200000        128         False       P2p
Gi0/2                Desg FWD 20000         128         False       P2p
```

In instance 2, Device B (00d0.f8ee.8c1e) is the root bridge. Port GigabitEthernet 0/2 of Device A is the root port and the port path cost is 1. Port GigabitEthernet 0/1 of Device A is the designated port and the port path cost is 4. The two ports are in forwarding state.

```
MST 2 vlans map : 20, 40
  Region Root Priority     4096
                MAC Address        00d0.f8ee.8c1e
                this bridge is region root

  Bridge ID   Priority       8192
                MAC Address        0074.9cee.f49e

Interface            Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1                Desg FWD 4             128         False       P2p
Gi0/2                Root FWD 1             128         False       P2p
```

(3) Check the instance spanning tree topology and port forwarding status on Device B.

In instances 0 and 1, Device A (0074.9cee.f49e) is the root bridge. Port GigabitEthernet 0/1 of Device B is the root port and the port path cost is 1. Port GigabitEthernet 0/2 of Device B is the designated port and the port path cost is 4. The two ports are in forwarding state.

```
DeviceB# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-29, 31-39, 41-4094
  Root ID      Priority       4096
                MAC Address        0074.9cee.f49e
                this bridge is root
                Hello Time     2 sec    Forward Delay 15 sec    Max Age 20 sec
```

```
  Bridge ID  Priority    8192
             MAC Address     00d0.f8ee.8c1e
             Hello Time   2 sec  Forward Delay 15 sec  Max Age 20 sec

Interface          Role Sts Cost      Prio    OperEdge Type
---------------- ---- --- --------- -------- -------- ----------------
Gi0/1              Root FWD 1           128      False    P2p
Gi0/2              Desg FWD 4           128      False    P2p

MST 1 vlans map : 10, 30
  Region Root Priority   4096
             MAC Address     0074.9cee.f49e
             this bridge is region root

  Bridge ID  Priority    8192
             MAC Address     00d0.f8ee.8c1e

Interface          Role Sts Cost      Prio    OperEdge Type
---------------- ---- --- --------- -------- -------- ----------------
Gi0/1              Root FWD 1           128      False    P2p
Gi0/2              Desg FWD 4           128      False    P2p
```

In instance 2, Device B (00d0.f8ee.8c1e) is the root bridge. Ports GigabitEthernet 0/1–0/2 of Device B are designated ports. They are all in forwarding state.

```
MST 2 vlans map : 20, 40
  Region Root Priority   4096
             MAC Address     00d0.f8ee.8c1e
             this bridge is region root

  Bridge ID  Priority    4096
             MAC Address     00d0.f8ee.8c1e

Interface          Role Sts Cost      Prio    OperEdge Type
---------------- ---- --- --------- -------- -------- ----------------
Gi0/1              Desg FWD 20000       128      False    P2p
Gi0/2              Desg FWD 20000       128      False    P2p
```

(4) Check the instance spanning tree topology and port forwarding status on Device C.

In instances 0 and 1, Device A (0074.9cee.f49e) is the root bridge. Port GigabitEthernet 0/1 of Device C is the root port and the port path cost is 1. The port is in forwarding state. Port GigabitEthernet 0/2 of Device C is the alternate port and the port path cost is 4. The port is in blocking state. Ports GigabitEthernet 0/3–0/6 of Device C are edge ports. They are all in forwarding state.

```
DeviceC# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-29, 31-39, 41-4094
  Root ID     Priority    4096
              MAC Address     0074.9cee.f49e
```

```
                this bridge is root
                Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID   Priority      32768
                MAC Address       0074.9cee.53ca
                Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1               Root FWD 1             128        False      P2p
Gi0/2               Altn BLK 4             128        False      P2p
Gi0/3               Desg FWD 20000         128        True       P2p
Gi0/4               Desg FWD 20000         128        True       P2p
Gi0/5               Desg FWD 20000         128        True       P2p
Gi0/6               Desg FWD 20000         128        True       P2p

MST 1 vlans map : 10, 30
  Region Root Priority    4096
                MAC Address       0074.9cee.f49e
                this bridge is region root

  Bridge ID   Priority      32768
                MAC Address       0074.9cee.53ca

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1               Root FWD 1             128        False      P2p
Gi0/2               Altn BLK 4             128        False      P2p
Gi0/3               Desg FWD 20000         128        True       P2p
Gi0/4               Desg FWD 20000         128        True       P2p
Gi0/5               Desg FWD 20000         128        True       P2p
Gi0/6               Desg FWD 20000         128        True       P2p
```

In instance 2, Device B (00d0.f8ee.8c1e) is the root bridge. Port GigabitEthernet 0/2 of Device C is the root port and the port path cost is 1. The port is in forwarding state. Port GigabitEthernet 0/1 of Device C is the alternate port and the port path cost is 4. The port is in blocking state. Ports GigabitEthernet 0/3–0/6 of Device C are edge ports. They are all in forwarding state.

```
MST 2 vlans map : 20, 40
  Region Root Priority    4096
                MAC Address       00d0.f8ee.8c1e
                this bridge is region root

  Bridge ID   Priority      32768
                MAC Address       0074.9cee.53ca

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
```

```
Gi0/1           Altn BLK 4          128       False     P2p
Gi0/2           Root FWD 1          128       False     P2p
Gi0/3           Desg FWD 20000      128       True      P2p
Gi0/4           Desg FWD 20000      128       True      P2p
Gi0/5           Desg FWD 20000      128       True      P2p
Gi0/6           Desg FWD 20000      128       True      P2p
```

(5) Check "DesignatedRoot", "RootPort", and "RootCost" of each device in instance 1. Check the "DesignatedRoot", "RootPort", and "RootCost" of each device in instances 0 and 2.

Device A is the root and the root path cost is 0.

```
DeviceA# display stp mst 1
###### MST 1 vlans mapped : 10, 30
BridgeAddr : 0074.9cee.f49e
Priority: 4096
TimeSinceTopologyChange : 0d:0h:16m:29s
TopologyChanges : 7
DesignatedRoot : 4097.0074.9cee.f49e
RootCost : 0
RootPort : 0
```

The root path cost from Device B to the root (4097.0074.9cee.f49e) through root port GigabitEthernet 0/2 is 1.

```
DeviceB# display stp mst 1
###### MST 1 vlans mapped : 10, 30
BridgeAddr : 00d0.f8ee.8c1e
Priority: 8192
TimeSinceTopologyChange : 0d:0h:16m:27s
TopologyChanges : 7
DesignatedRoot : 4097.0074.9cee.f49e
RootCost : 1
RootPort : GigabitEthernet 0/2
```

The root path cost from Device C to the root (4097.0074.9cee.f49e) through root port GigabitEthernet 0/1 is 1.

```
DeviceC# display stp mst 1
###### MST 1 vlans mapped : 10, 30
BridgeAddr : 0074.9cee.53ca
Priority: 32768
TimeSinceTopologyChange : 0d:0h:19m:48s
TopologyChanges : 5
DesignatedRoot : 4097.0074.9cee.f49e
RootCost : 1
RootPort : GigabitEthernet 0/1
```

(6) Verify that the loop guard function ("PortGuardmode:Guard loop") on uplink ports GigabitEthernet 0/1–0/2 of Device C is enabled. The following uses GigabitEthernet 0/1 as an example.

```
DeviceC# display stp interface gigabitethernet 0/1
PortAdminPortFast : Disabled
```

```
PortOperPortFast : Disabled
PortAdminAutoEdge : Enabled
PortOperAutoEdge : Disabled
PortAdminLinkType : auto
PortOperLinkType : point-to-point
PortBPDUGuard : Disabled
PortBPDUFilter : Disabled
PortGuardmode    : Guard loop

###### MST 0 vlans mapped :1-9, 11-19, 21-29, 31-39, 41-4094
PortState : forwarding
PortPriority : 128
PortDesignatedRoot : 4096.0074.9cee.f49e
PortDesignatedCost : 0
PortDesignatedBridge :4096.0074.9cee.f49e
PortDesignatedPortPriority : 128
PortDesignatedPort : 1
PortForwardTransitions : 1
PortAdminPathCost : 1
PortOperPathCost : 1
Inconsistent states : normal
PortRole : rootPort

###### MST 1 vlans mapped :10, 30
PortState : forwarding
PortPriority : 128
PortDesignatedRoot : 4097.0074.9cee.f49e
PortDesignatedCost : 0
PortDesignatedBridge :4097.0074.9cee.f49e
PortDesignatedPortPriority : 128
PortDesignatedPort : 1
PortForwardTransitions : 1
PortAdminPathCost : 1
PortOperPathCost : 1
Inconsistent states : normal
PortRole : rootPort

###### MST 2 vlans mapped :20, 40
PortState : discarding
PortPriority : 128
PortDesignatedRoot : 4098.00d0.f8ee.8c1e
PortDesignatedCost : 0
PortDesignatedBridge :8194.0074.9cee.f49e
PortDesignatedPortPriority : 128
PortDesignatedPort : 1
PortForwardTransitions : 3
```

```
PortAdminPathCost : 4
PortOperPathCost : 4
Inconsistent states : normal
PortRole : alternatePort
```

(7) Verify that port fast is configured on ports GigabitEthernet 0/3–0/6 of Device C ("PortAdminPortFast : Enabled"), the ports work in port fast state ("PortOperPortFast : Enabled"), and BPDU guard is enabled on the ports (“PortBPDUGuard: Enabled”). The following uses port GigabitEthernet 0/3 as an example.

```
DeviceC# display stp interface gigabitethernet 0/3

PortAdminPortFast : Enabled
PortOperPortFast : Enabled
PortAdminAutoEdge : Disabled
PortOperAutoEdge : Enabled
PortAdminLinkType : auto
PortOperLinkType : point-to-point
PortBPDUGuard : Enabled
PortBPDUFilter : Disabled
PortGuardmode   : None

###### MST 0 vlans mapped :1-9, 11-19, 21-29, 31-39, 41-4094
PortState : forwarding
PortPriority : 128
PortDesignatedRoot : 4096.0074.9cee.f49e
PortDesignatedCost : 0
PortDesignatedBridge :32768.0074.9cee.53ca
PortDesignatedPortPriority : 128
PortDesignatedPort : 1
PortForwardTransitions : 1
PortAdminPathCost : 20000
PortOperPathCost : 20000
Inconsistent states : normal
PortRole : designatedPort

###### MST 1 vlans mapped :10, 30
PortState : forwarding
PortPriority : 128
PortDesignatedRoot : 4097.0074.9cee.f49e
PortDesignatedCost : 0
PortDesignatedBridge :32769.0074.9cee.53ca
PortDesignatedPortPriority : 128
PortDesignatedPort : 1
PortForwardTransitions : 1
PortAdminPathCost : 20000
PortOperPathCost : 20000
Inconsistent states : normal
PortRole : designatedPort
```

```
###### MST 2 vlans mapped :20, 40
PortState : forwarding
PortPriority : 128
PortDesignatedRoot : 4098.00d0.f8ee.8c1e
PortDesignatedCost : 0
PortDesignatedBridge :32770.0074.9cee.53ca
PortDesignatedPortPriority : 128
PortDesignatedPort : 1
PortForwardTransitions : 1
PortAdminPathCost : 20000
PortOperPathCost : 20000
Inconsistent states : normal
PortRole : designatedPort
```

6. Configuration Files

- Device A configuration file

```
stp mst configuration
  instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
  instance 1 vlan 10, 30
  instance 2 vlan 20, 40
!
stp mst 0 priority 4096
stp mst 1 priority 4096
stp mst 2 priority 8192
stp
!
sysmac 0074.9cee.f49e
!
vlan range 1,10,20,30,40
!
interface GigabitEthernet 0/1
  switch-port
  switch-port mode trunk
  switch-port trunk pvid vlan 1
  switch-port trunk allow-pass vlan only 1-4094
  stp mst 2 cost 4
!
interface GigabitEthernet 0/2
  switch-port
  switch-port mode trunk
  switch-port trunk pvid vlan 1
  switch-port trunk allow-pass vlan only 1-4094
  stp mst 2 cost 1
```

- Device B configuration file

```
stp mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 30
 instance 2 vlan 20, 40
!
stp mst 0 priority 8192
stp mst 1 priority 8192
stp mst 2 priority 4096
stp
!
sysmac 00d0.f8ee.8c1e
!
vlan range 1,10,20,30,40
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 stp mst 1 cost 4
 stp mst 0 cost 4
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 stp mst 1 cost 1
 stp mst 0 cost 1
```

- Device C configuration file

```
stp mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 30
 instance 2 vlan 20, 40
!
stp
!
sysmac 0074.9cee.53ca
!
vlan range 1,10,20,30,40
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
```

```
 stp guard loop
 stp mst 2 cost 4
 stp mst 1 cost 1
 stp mst 0 cost 1
!
interface GigabitEthernet 0/2
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 stp guard loop
 stp mst 2 cost 1
 stp mst 1 cost 4
 stp mst 0 cost 4
!
interface GigabitEthernet 0/3
 switch-port
 switch-port mode access
 switch-port access vlan 10
 stp bpduguard enable
 stp edged-port
 stp autoedge disabled
!
interface GigabitEthernet 0/4
 switch-port
 switch-port mode access
 switch-port access vlan 20
 stp bpduguard enable
 stp edged-port
 stp autoedge disabled
!
interface GigabitEthernet 0/5
 switch-port
 switch-port mode access
 switch-port access vlan 30
 stp bpduguard enable
 stp edged-port
 stp autoedge disabled
!
interface GigabitEthernet 0/6
 switch-port
 switch-port mode access
 switch-port access vlan 40
 stp bpduguard enable
 stp edged-port
 stp autoedge disabled
```

7. Common Errors

- When no bridge priority lower than that of the root bridge but higher than that of access devices is configured for non-root devices, the non-root core devices and access devices will determine upstream devices by comparing MAC addresses in the spanning tree calculation. As a result, the calculation result may not meet networking requirements.
- When no port path cost is configured for non-root devices, the spanning tree calculation result may not meet networking requirements due to different link conditions.
- The root guard function is configured on a root port, master port, or alternate port, and ports may be blocked incorrectly.

1.24.2 Configuring MSTP+VRRP

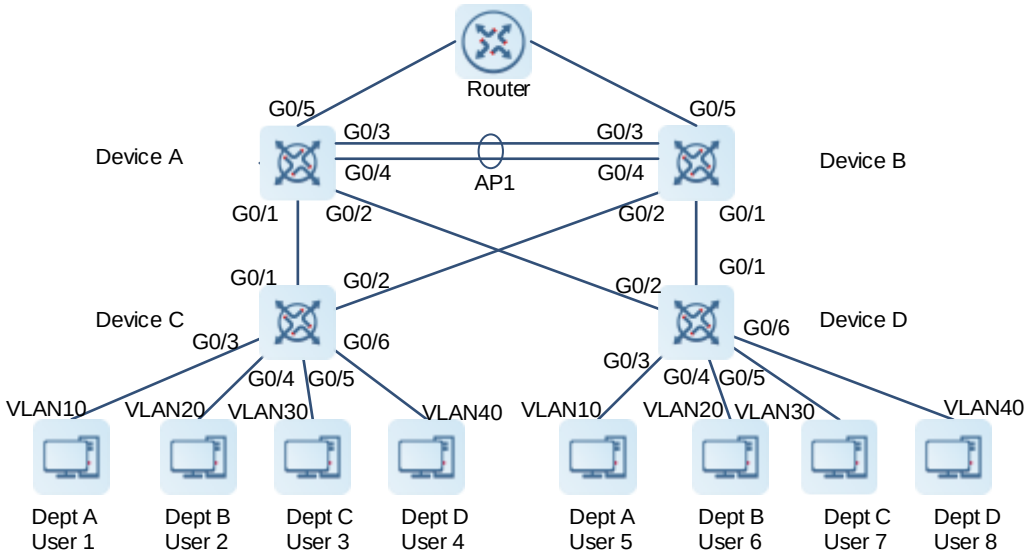
1. Requirements

The MSTP+VRRP dual-core solution is a typical application of MSTP. This solution uses a hierarchical network architecture and MSTP and Virtual Router Redundancy Protocol (VRRP) to implement redundant backup and VLAN load balancing, to enhance the network system availability. The highlight of this architecture is the hierarchical network structure. The capacity indicators, characteristics, and functions of network devices in each layer can be optimized based on their network locations and functions to boost the system stability and availability. This solution usually uses a three-layer (core layer, aggregation layer, and access layer) or two-layer (core layer and access layer) architecture. The topology shown in Figure 1-19 adopts a two-layer architecture. The networking requirements are as follows:

- Core layer: Configure multiple MSTP instances to achieve load balancing. Create instance 1 and instance 2. Map instance 1 to VLAN 10 and VLAN 30 and instance 2 to VLAN 20 and VLAN 40. Device A is the root bridge of instances 0 and instance 1 (instance 0 exists by default) as well as the VRRP master device of VLAN 10 and VLAN 30. Device B is the root bridge of instance 2 as well as the VRRP master device of VLANs 20 and 40.
- Access layer: Configure ports that are directly connected to terminals (such as PCs or servers) as edge ports and enable BPDU guard to prevent the access of unauthorized user devices.

2. Topology

Figure 1-19 Configuring MSTP and VRRP



3. Notes

- Configure MSTP.
  - Set the spanning tree mode to MSTP. The default spanning tree mode is MSTP and no configuration is required.
  - Configure an MST region. On Devices A, B, C, and D, map instance 1 to VLAN 10 and VLAN 30 and instance 2 to VLAN 20 and VLAN 40.
  - Set the bridge priority of instances 0 and 1 to 4096 on Device A and 8192 on Device B to make Device A become the root bridge of instances 0 and 1.
  - Set the bridge priority of instance 2 to 8192 on Device A and 4096 on Device B to make Device B become the root bridge of instance 2.
  - On access Devices C and D, instances 0, 1, and 2 use the default bridge priority (32768). Configure ports that connect to user terminals as edge ports so that they do not participate in spanning tree calculation. Enable BPDU guard on the ports.
- Configure monitoring ports for a VRRP group. Configure the uplink port of the master device as the VLAN monitoring port. Configure port GigabitEthernet 0/5 of Device A to monitor VLAN 10 and VLAN 30, and port GigabitEthernet 0/5 of Device B to monitor VLAN 20 and VLAN 40.
  - In interface configuration mode of the uplink port, run the **unset switch-port** command to configure the monitoring port as a Layer 3 interface and configure an IP address for the port. Layer 3 interfaces do not participate in spanning tree calculation.
  - In switch virtual interface (SVI) configuration mode of a VLAN, run the **vrrp** *group-id* **track** *interface-type interface-number* [*priority decrement*] command to configure a monitoring interface *interface-type interface-number* for the VRRP group *group-id* and VRRP priority change value *priority decrement*. When the link state of the monitored interface or IP route reachability status changes, VRRP priority will change, and the change value is *decrement*. When the link is interrupted, the priority is reduced. When

the link recovers, the priority is restored. The value range is from 1 to 255 and the default value is **10**.

- Configure priority parameters: The VRRP priority change value needs to be taken into account together with the VRRP priority. In SVI configuration mode of a VLAN, run the **vrrp** *group-id* **priority** *priority* command to configure the VRRP priority. The value range of *priority* is from 1 to 254 and the default value is **100**. The default value of *priority decrement* is **10**. When the monitoring port is down, the VRRP priority is reduced to 90 (that is, 100 – 10).
  - In this example, raise the VRRP priority of VLANs 10 and 30 to 120 on the master device (Device A), use the default priority **100** on Device B, and set *priority decrement* to 30. When monitoring port GigabitEthernet 0/5 of Device A is down due to a fault, the VRRP priority of VLAN 10 and VLAN 30 is reduced by 30 to 90, which is lower than the default priority (100) on Device B. Therefore, data of VLAN 10 and VLAN 30 will be transmitted through Device B. When monitoring port GigabitEthernet 0/5 of Device A recovers, the VRRP priority of VLAN 10 and VLAN 30 is restored to 120, which is greater than the default priority on Device B. Therefore, data of VLAN 10 and VLAN 30 is transmitted through Device A. Then, VLAN 10 and VLAN 30 use Device A as the VRRP master device and Device B as the VRRP backup device.
  - Likewise, raise the VRRP priority of VLAN 20 and VLAN 40 to 120 on Device B and use the default priority (**100**) on Device A.
- Configure VRRP: Add SVIs of VLANs to a VRRP group and configure a virtual IP address for the VRRP group.
- Pay attention to the following when configuring SVI addresses: If the monitoring port is down and the VRRP priorities are the same on two devices after priority reduction, the SVI addresses of a VLAN on the two devices will be compared. A larger SVI address indicates a higher priority. Therefore, it is recommended that the SVI IP addresses of VLAN 10 and VLAN 30 on the master device (Device A) be greater than those of VLAN 10 and VLAN 30 on Device B. The same configuration applies to other VLANs.
- Configure an aggregate link between core devices. Configure the downlink ports of core devices and uplink ports of access devices as trunk ports. On access devices, configure ports that connect to user terminals as access ports and add them to VLANs.

4. Procedure

(1) Configure the MSTP function.

Configure the MSTP function for core Device A. Create VLANs, and map instance 1 to VLAN 10 and VLAN 30 and instance 2 to VLAN 20 and VLAN 40. Set the bridge priority to 4096 for instances 0 and 1 and 8192 for instance 2 to make Device A become the root bridge of instances 0 and 1. Enable MSTP.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan range 10,20,30,40
DeviceA(cfg-vlan-range)# exit
DeviceA(cfg)# stp mode mstp
DeviceA(cfg)# stp mst configuration
DeviceA(cfg-mst)# instance 1 vlan 10,30
DeviceA(cfg-mst)# instance 2 vlan 20,40
DeviceA(cfg-mst)# exit
DeviceA(cfg)# stp mst 0 priority 4096
```

```
DeviceA(cfg)# stp mst 1 priority 4096
DeviceA(cfg)# stp mst 2 priority 8192
DeviceA(cfg)# stp
```

Configure the MSTP function for core Device B. Create VLANs, and map instance 1 to VLAN 10 and VLAN 30 and instance 2 to VLAN 20 and VLAN 40. Set the bridge priority to 8192 for instances 0 and 1 and 4096 for instance 2 to make Device B become the root bridge of instance 2. Enable MSTP.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan range 10,20,30,40
DeviceB(cfg-vlan-range)# exit
DeviceB(cfg)# stp mode mstp
DeviceB(cfg)# stp mst configuration
DeviceB(cfg-mst)# instance 1 vlan 10,30
DeviceB(cfg-mst)# instance 2 vlan 20,40
DeviceB(cfg-mst)# exit
DeviceB(cfg)# stp mst 0 priority 8192
DeviceB(cfg)# stp mst 1 priority 8192
DeviceB(cfg)# stp mst 2 priority 4096
DeviceB(cfg)# stp
```

Configure the MSTP function for access Devices C and D. The bridge priority does not need to be configured on access devices. Configure ports that are directly connected to user terminals as edge ports and enable BPDU guard on the ports. The configurations on Device D are similar to those on Device C. The following uses Device C as an example.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# vlan range 10,20,30,40
DeviceC(cfg-vlan-range)# exit
DeviceC(cfg)# stp mode mstp
DeviceC(cfg)# stp mst configuration
DeviceC(cfg-mst)# instance 1 vlan 10,30
DeviceC(cfg-mst)# instance 2 vlan 20,40
DeviceC(cfg-mst)# exit
DeviceC(cfg)# stp
DeviceC(cfg)# interface list gigabitethernet 0/3-6
DeviceC(cfg-if-list)# stp edged-port
DeviceC(cfg-if-list)# stp bpduguard enable
DeviceC(cfg-if-list)# exit
```

(2) Configure monitoring ports for the VRRP group.

Configure port GigabitEthernet 0/5 of Device A as a Layer 3 interface and set the IP address to 10.10.1.1/24. Then, this port serves as the monitoring port of VLAN 10 and VLAN 30.

```
DeviceA(cfg)# interface gigabitethernet 0/5
DeviceA(cfg-if-GigabitEthernet 0/5)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/5)# ip address 10.10.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/5)# exit
```

Configure port GigabitEthernet 0/5 of Device B as a Layer 3 interface and set the IP address to 10.10.2.1/24. Then, this port serves as the monitoring port of VLAN 20 and VLAN 40.

```
DeviceB(cfg)# interface gigabitethernet 0/5
DeviceB(cfg-if-GigabitEthernet 0/5)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/5)# ip address 10.10.2.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/5)# exit
```

(3) Configure VRRP for VLAN 10 and VLAN 30 on core devices.

Configure the master device (Device A). Enter an SVI and configure an SVI address. Add the SVI to the VRRP group and configure a virtual gateway IP address for the VRRP group. Raise the VRRP priority to 120, configure port GigabitEthernet 0/5 as a monitoring port, and set priority decrement to 30.

```
DeviceA(cfg)# interface vlan 10
DeviceA(cfg-if-VLAN 10)# ip address 192.168.10.3 255.255.255.0
DeviceA(cfg-if-VLAN 10)# vrrp 10 ip 192.168.10.1
DeviceA(cfg-if-VLAN 10)# vrrp 10 priority 120
DeviceA(cfg-if-VLAN 10)# vrrp 10 track gigabitethernet 0/5 30
DeviceA(cfg-if-VLAN 10)# exit
DeviceA(cfg)# interface vlan 30
DeviceA(cfg-if-VLAN 30)# ip address 192.168.30.3 255.255.255.0
DeviceA(cfg-if-VLAN 30)# vrrp 30 ip 192.168.30.1
DeviceA(cfg-if-VLAN 30)# vrrp 30 priority 120
DeviceA(cfg-if-VLAN 30)# vrrp 30 track gigabitethernet 0/5 30
DeviceA(cfg-if-VLAN 30)# exit
```

Configure the backup device (Device B). Enter an SVI and configure an SVI address. Add the SVI to the VRRP group and configure a virtual gateway IP address for the VRRP group. On the backup device, use the default VRRP priority (100). No monitoring port needs to be configured.

```
DeviceB(cfg)# interface vlan 10
DeviceB(cfg-if-VLAN 10)# ip address 192.168.10.2 255.255.255.0
DeviceB(cfg-if-VLAN 10)# vrrp 10 ip 192.168.10.1
DeviceB(cfg-if-VLAN 10)# exit
DeviceB(cfg)# interface vlan 30
DeviceB(cfg-if-VLAN 30)# ip address 192.168.30.2 255.255.255.0
DeviceB(cfg-if-VLAN 30)# vrrp 30 ip 192.168.30.1
DeviceB(cfg-if-VLAN 30)# exit
```

(4) Configure VRRP for VLAN 20 and VLAN 40 on core devices.

Configure the master device (Device B). Enter an SVI and configure an SVI address. Add the SVI to the VRRP group and configure a virtual gateway IP address for the VRRP group. Raise the VRRP priority to 120, configure port GigabitEthernet 0/5 as a monitoring port, and set priority decrement to 30.

```
DeviceB(cfg)# interface vlan 20
DeviceB(cfg-if-VLAN 20)# ip address 192.168.20.3 255.255.255.0
DeviceB(cfg-if-VLAN 20)# vrrp 20 ip 192.168.20.1
DeviceB(cfg-if-VLAN 20)# vrrp 20 priority 120
DeviceB(cfg-if-VLAN 20)# vrrp 20 track gigabitethernet 0/5 30
DeviceB(cfg-if-VLAN 20)# exit
DeviceB(cfg)# interface vlan 40
```

```
DeviceB(cfg-if-VLAN 40)# ip address 192.168.40.3 255.255.255.0
DeviceB(cfg-if-VLAN 40)# vrrp 40 ip 192.168.40.1
DeviceB(cfg-if-VLAN 40)# vrrp 40 priority 120
DeviceB(cfg-if-VLAN 40)# vrrp 40 track gigabitethernet 0/5 30
DeviceB(cfg-if-VLAN 40)# exit
```

Configure the backup device (Device A). Enter an SVI and configure an SVI address. Add the SVI to the VRRP group and configure a virtual gateway IP address for the VRRP group. On the backup device, use the default VRRP priority (100). No monitoring port needs to be configured.

```
DeviceA(cfg)# interface vlan 20
DeviceA(cfg-if-VLAN 20)# ip address 192.168.20.2 255.255.255.0
DeviceA(cfg-if-VLAN 20)# vrrp 20 ip 192.168.20.1
DeviceA(cfg-if-VLAN 20)# exit
DeviceA(cfg)# interface vlan 40
DeviceA(cfg-if-VLAN 40)# ip address 192.168.40.2 255.255.255.0
DeviceA(cfg-if-VLAN 40)# vrrp 40 ip 192.168.40.1
DeviceA(cfg-if-VLAN 40)# exit
```

(5) Configure an aggregate link between VRRP core devices.

On Device A, configure ports GigabitEthernet 0/3 and 0/4 as AggregatePort 1 and set the aggregate interface to work in trunk mode.

```
DeviceA(cfg)# interface list gigabitethernet 0/3-4
DeviceA(cfg-if-list)# port-group 1
DeviceA(cfg-if-list)# exit
DeviceA(cfg)# interface aggregateport 1
DeviceA(cfg-if-AggregatePort 1)# switch-port
DeviceA(cfg-if-AggregatePort 1)# switch-port mode trunk
DeviceA(cfg-if-AggregatePort 1)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-AggregatePort 1)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-AggregatePort 1)# exit
```

On Device B, configure ports GigabitEthernet 0/3 and 0/4 as AggregatePort 1 and set the aggregate interface to work in trunk mode.

```
DeviceB(cfg)# interface list gigabitethernet 0/3-4
DeviceB(cfg-if-list)# port-group 1
DeviceB(cfg-if-list)# exit
DeviceB(cfg)# interface aggregateport 1
DeviceB(cfg-if-AggregatePort 1)# switch-port
DeviceB(cfg-if-AggregatePort 1)# switch-port mode trunk
DeviceB(cfg-if-AggregatePort 1)# switch-port trunk pvid vlan 1
DeviceB(cfg-if-AggregatePort 1)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-AggregatePort 1)# exit
```

(6) Configure the downlink ports of core devices and the uplink ports of access devices as trunk ports.

Configure downlink ports GigabitEthernet 0/1–0/2 of Device A as trunk ports.

```
DeviceA(cfg)# interface list gigabitethernet 0/1-2
DeviceA(cfg-if-list)# switch-port
```

```
DeviceA(cfg-if-list)# switch-port mode trunk
DeviceA(cfg-if-list)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-list)# switch-port trunk allow-pass vlan all DeviceA(cfg-if-list)# end
DeviceA# write
```

Configure downlink ports GigabitEthernet 0/1–0/2 of Device B as trunk ports.

```
DeviceB(cfg)# interface list gigabitethernet 0/1-2
DeviceB(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode trunk
DeviceA(cfg-if-list)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-list)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-list)# end
DeviceB# write
```

Configure uplink ports GigabitEthernet 0/1–0/2 of Device C as trunk ports.

```
DeviceC(cfg)# interface list gigabitethernet 0/1-2
DeviceC(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode trunk
DeviceA(cfg-if-list)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-list)# switch-port trunk allow-pass vlan all
DeviceC(cfg-if-list)# exit
```

Configure uplink ports GigabitEthernet 0/1–0/2 of Device D as trunk ports.

```
DeviceD(cfg)# interface list gigabitethernet 0/1-2
DeviceD(cfg-if-list)# switch-port
DeviceA(cfg-if-list)# switch-port mode trunk
DeviceA(cfg-if-list)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-list)# switch-port trunk allow-pass vlan all
DeviceD(cfg-if-list)# exit
```

(7) Set user ports on access devices to work in access mode and add them to VLANs.

Configure Device C.

```
DeviceC(cfg)# interface gigabitethernet 0/3
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port mode access
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port access vlan 10
DeviceC(cfg-if-GigabitEthernet 0/3)# exit
DeviceC(cfg)# interface gigabitethernet 0/4
DeviceC(cfg-if-GigabitEthernet 0/4)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port mode access
DeviceC(cfg-if-GigabitEthernet 0/4)# switch-port access vlan 20
DeviceC(cfg-if-GigabitEthernet 0/4)# exit
DeviceC(cfg)# interface gigabitethernet 0/5
DeviceC(cfg-if-GigabitEthernet 0/5)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port mode access
DeviceC(cfg-if-GigabitEthernet 0/5)# switch-port access vlan 30
DeviceC(cfg-if-GigabitEthernet 0/5)# exit
DeviceC(cfg)# interface gigabitethernet 0/6
```

```
DeviceC(cfg-if-GigabitEthernet 0/6)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port mode access
DeviceC(cfg-if-GigabitEthernet 0/6)# switch-port access vlan 40
DeviceC(cfg-if-GigabitEthernet 0/6)# end
DeviceC# write
```

Configure Device D.

```
DeviceD(cfg)# interface gigabitethernet 0/3
DeviceD(cfg-if-GigabitEthernet 0/3)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port mode access
DeviceD(cfg-if-GigabitEthernet 0/3)# switch-port access vlan 10
DeviceD(cfg-if-GigabitEthernet 0/3)# exit
DeviceD(cfg)# interface gigabitethernet 0/4
DeviceD(cfg-if-GigabitEthernet 0/4)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port mode access
DeviceD(cfg-if-GigabitEthernet 0/4)# switch-port access vlan 20
DeviceD(cfg-if-GigabitEthernet 0/4)# exit
DeviceD(cfg)# interface gigabitethernet 0/5
DeviceD(cfg-if-GigabitEthernet 0/5)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port mode access
DeviceD(cfg-if-GigabitEthernet 0/5)# switch-port access vlan 30
DeviceD(cfg-if-GigabitEthernet 0/5)# exit
DeviceD(cfg)# interface gigabitethernet 0/6
DeviceD(cfg-if-GigabitEthernet 0/6)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/3)# switch-port mode access
DeviceD(cfg-if-GigabitEthernet 0/6)# switch-port access vlan 40
DeviceD(cfg-if-GigabitEthernet 0/6)# end
DeviceD# write
```

5. Verification

Run the display stp summary command to display the spanning tree topology.

View spanning tree topology on Device A. Device A (00d0.f822.3344) is the root.

```
DeviceA# display stp summary
Spanning tree enabled protocol mstp

MST 0 vlans map : 1-9, 11-19, 21-29, 31-39, 41-4094
  Root ID       Priority       4096
                MAC Address        00d0.f822.3344
                this bridge is root
                Hello Time     4 sec    Forward Delay 18 sec    Max Age 25 sec

  Bridge ID    Priority       4096
                MAC Address        00d0.f822.3344
                Hello Time     4 sec    Forward Delay 18 sec    Max Age 25 sec

Interface            Role Sts Cost           Prio        OperEdge Type
```

```
---------------- ---- --- ---------- -------- -------- -----------------
Ag1                 Desg FWD 19000        128       False      P2p
Gi0/1               Desg FWD 200000       128       False      P2p
Gi0/2               Desg FWD 200000       128       False      P2p

MST 1 vlans map : 10,30
  Region Root Priority    4096
                MAC Address       00d0.f822.3344
                this bridge is region root

  Bridge ID   Priority      4096
                MAC Address       00d0.f822.3344

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- -----------------
Ag1                 Desg FWD 19000        128       False      P2p
Gi0/1               Desg FWD 200000       128       False      P2p
Gi0/2               Desg FWD 200000       128       False      P2p

MST 2 vlans map : 20,40
  Region Root Priority    4096
                MAC Address       001a.a917.78cc
                this bridge is region root

  Bridge ID   Priority      8192
                MAC Address       00d0.f822.3344

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- -----------------
Ag1                 Root FWD 19000        128       False      P2p
Gi0/1               Desg FWD 200000       128       False      P2p
Gi0/2               Desg FWD 200000       128       False      P2p
```

View spanning tree topology on Device B. Device A (00d0.f822.3344) is the root.

```
DeviceB# display stp summary

Spanning tree enabled protocol mstp

MST 0 vlans map : 1-9, 11-19, 21-29, 31-39, 41-4094
  Root ID      Priority      4096
                MAC Address       00d0.f822.3344
                this bridge is root
                Hello Time    4 sec   Forward Delay 18 sec   Max Age 25 sec

  Bridge ID   Priority      8192
                MAC Address       001a.a917.78cc
                Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec
```

```
Interface            Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Ag1                  Root FWD 19000         128        False      P2p
Gi0/1                Desg FWD 200000        128        False      P2p
Gi0/2                Desg FWD 200000        128        False      P2p

MST 1 vlans map : 10,30
  Region Root Priority     4096
                MAC Address        00d0.f822.3344
                this bridge is region root

  Bridge ID   Priority       8192
                MAC Address        001a.a917.78cc

Interface            Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Ag1                  Root FWD 19000         128        False      P2p
Gi0/1                Desg FWD 200000        128        False      P2p
Gi0/2                Desg FWD 200000        128        False      P2p

MST 2 vlans map : 20,40
  Region Root Priority     4096
                MAC Address        001a.a917.78cc
                this bridge is region root

  Bridge ID   Priority       4096
                MAC Address        001a.a917.78cc

Interface            Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Ag1                  Desg FWD 19000         128        False      P2p
Gi0/1                Desg FWD 200000        128        False      P2p
Gi0/2                Desg FWD 200000        128        False      P2p
```

View spanning tree topology on Device C and Device D. Device C is taken as an example. Device A (00d0.f822.3344) is the root.

```
DeviceC# display stp summary

Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-29, 31-39, 41-4094
  Root ID      Priority       4096
                MAC Address        00d0.f822.3344
                this bridge is root
                Hello Time     4 sec    Forward Delay 18 sec    Max Age 25 sec

  Bridge ID   Priority       32768
```

```
                MAC Address       001a.a979.00ea
                Hello Time    2 sec  Forward Delay 15 sec  Max Age 20 sec

Interface          Role Sts Cost      Prio      Type  OperEdge
---------------- ---- --- ---------- -------- ----- ----------
Gi0/2              Altn BLK 200000      128       P2p   False
Gi0/1              Root FWD 200000      128       P2p   False

MST 1 vlans map : 10,30
  Region Root Priority   4096
                MAC Address       00d0.f822.3344
                this bridge is region root

  Bridge ID  Priority     32768
                MAC Address       001a.a979.00ea

Interface          Role Sts Cost      Prio      Type  OperEdge
---------------- ---- --- ---------- -------- ----- ----------
Gi0/2              Altn BLK 200000      128       P2p   False
Gi0/1              Root FWD 200000      128       P2p   False

MST 2 vlans map : 20,40
  Region Root Priority   4096
                MAC Address       001a.a917.78cc
                this bridge is region root

  Bridge ID  Priority     32768
                MAC Address       001a.a979.00ea

Interface          Role Sts Cost      Prio      Type  OperEdge
---------------- ---- --- ---------- -------- ----- ----------
Gi0/2              Root FWD 200000      128       P2p   False
Gi0/1              Altn BLK 200000      128       P2p   False
```

Run the display vrrp brief command to check whether VRRP active/standy relationship is successfully established. VLAN 10 and VLAN 30 use Device A as the master device and Device B as the standby device. VLAN 20 and VLAN 40 use Device B as the master device and Device A as the standby device.

```
DeviceA# display vrrp brief
Interface   Grp   Pri   timer   Own   Pre   State     Master addr     Group addr
VLAN 10     10    120   3       -     P     Master    192.168.10.3    192.168.10.1
VLAN 20     20    100   3       -     P     Backup    192.168.20.2    192.168.20.1
VLAN 30     30    120   3       -     P     Master    192.168.30.3    192.168.30.1
VLAN 40     40    100   3       -     P     Backup    192.168.40.2    192.168.40.1
```

Run the display vrrp brief command on Device B to check whether VRRP active/standy relationship is successfully established.

```
DeviceB# display vrrp brief
```

```
Interface   Grp  Pri  timer  Own  Pre  State    Master addr     Group addr
VLAN 10     10   100  3      -    P    Backup   192.168.10.2    192.168.10.1
VLAN 20     20   120  3      -    P    Master   192.168.20.3    192.168.20.1
VLAN 30     30   100  3      -    P    Backup   192.168.30.2    192.168.30.1
VLAN 40     40   120  3      -    P    Master   192.168.40.3    192.168.40.1
```

Disconnect uplink GigabitEthernet 0/5 of Device A. VLAN 10 and VLAN 30 use Device A as the master device and Device B as the standby device.

Run the display vrrp brief command on Device A to check VRRP status change.

```
DeviceA#display vrrp brief
Interface   Grp  Pri  timer  Own  Pre  State    Master addr     Group addr
VLAN 10     10   90   3      -    P    Backup   192.168.10.3    192.168.10.1
VLAN 20     20   100  3      -    P    Backup   192.168.20.2    192.168.20.1
VLAN 30     30   90   3      -    P    Backup   192.168.30.3    192.168.30.1
VLAN 40     40   100  3      -    P    Backup   192.168.40.2    192.168.40.1
```

Run the display vrrp brief command on Device B to check VRRP status change.

```
DeviceB# display vrrp brief
Interface   Grp  Pri  timer  Own  Pre  State    Master addr     Group addr
VLAN 10     10   100  3      -    P    Master   192.168.10.2    192.168.10.1
VLAN 20     20   120  3      -    P    Master   192.168.20.3    192.168.20.1
VLAN 30     30   100  3      -    P    Master   192.168.30.2    192.168.30.1
VLAN 40     40   120  3      -    P    Master   192.168.40.3    192.168.40.1
```

Disconnect uplink GigabitEthernet 0/5 of Device B. VLAN 20 and VLAN 40 use Device B as the master device and Device A as the standby device.

Run the display vrrp brief command on Device A to check VRRP status change.

```
DeviceA# display vrrp brief
Interface   Grp  Pri  timer  Own  Pre  State    Master addr     Group addr
VLAN 10     10   120  3      -    P    Master   192.168.10.3    192.168.10.1
VLAN 20     20   100  3      -    P    Master   192.168.20.2    192.168.20.1
VLAN 30     30   120  3      -    P    Master   192.168.30.3    192.168.30.1
VLAN 40     40   100  3      -    P    Master   192.168.40.2    192.168.40.1
```

Run the display vrrp brief command on Device B to check VRRP status change.

```
DeviceB# display vrrp brief
Interface   Grp  Pri  timer  Own  Pre  State    Master addr     Group addr
VLAN 10     10   100  3      -    P    Backup   192.168.10.2    192.168.10.1
VLAN 20     20   90   3      -    P    Backup   192.168.20.3    192.168.20.1
VLAN 30     30   100  3      -    P    Backup   192.168.30.2    192.168.30.1
VLAN 40     40   90   3      -    P    Backup   192.168.40.3    192.168.40.1
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
stp mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
```

```
 instance 1 vlan 10, 30
 instance 2 vlan 20, 40
!
stp mst 0 priority 4096
stp mst 1 priority 4096
stp mst 2 priority 8192
stp
!
sysmac 00d0.f822.3344
!
vlan range 1,10,20,30,40
!
interface GigabitEthernet 0/1
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/3
 port-group 1
!
interface GigabitEthernet 0/4
 port-group 1
!
interface AggregatePort 1
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface gigabitethernet 0/5
 unset switch-port
 ip address 10.10.1.1 255.255.255.0
!
interface vlan 10
 ip address 192.168.10.3 255.255.255.0
 vrrp 10 ip 192.168.10.1
 vrrp 10 priority 120
 vrrp 10 track gigabitethernet 0/5 30
!
```

```
interface vlan 20
 ip address 192.168.20.2 255.255.255.0
 vrrp 20 ip 192.168.20.1
!
interface vlan 30
 ip address 192.168.30.3 255.255.255.0
 vrrp 30 ip 192.168.30.1
 vrrp 30 priority 120
 vrrp 30 track gigabitethernet 0/5 30
!
interface vlan 40
 ip address 192.168.40.2 255.255.255.0
 vrrp 40 ip 192.168.40.1
```

- Device B configuration file

```
hostname DeviceB
stp mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 30
 instance 2 vlan 20, 40
!
stp mst 0 priority 8192
stp mst 1 priority 8192
stp mst 2 priority 4096
stp
!
sysmac 001a.a917.78cc
!
vlan range 1,10,20,30,40
!
interface GigabitEthernet 0/1
switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/3
 port-group 1
!
interface GigabitEthernet 0/4
 port-group 1
```

```
!
interface AggregatePort 1
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface gigabitethernet 0/5
 unset switch-port
 ip address 10.10.2.1 255.255.255.0
!
interface vlan 10
 ip address 192.168.10.2 255.255.255.0
 vrrp 10 ip 192.168.10.1
!
interface vlan 20
 ip address 192.168.20.3 255.255.255.0
 vrrp 20 ip 192.168.20.1
 vrrp 20 priority 120
 vrrp 20 track gigabitethernet 0/5 30
!
interface vlan 30
 ip address 192.168.30.2 255.255.255.0
 vrrp 30 ip 192.168.30.1
!
interface vlan 40
 ip address 192.168.40.3 255.255.255.0
 vrrp 40 ip 192.168.40.1
 vrrp 40 priority 120
 vrrp 40 track gigabitethernet 0/5 30
```

- Configuration files of Devices C and D

```
stp mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 30
 instance 2 vlan 20, 40
!
stp
!
vlan range 1,10,20,30,40
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
```

```
interface GigabitEthernet 0/2
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/3
 switch-port
switch-port mode access
 switch-port access vlan 10
 stp bpduguard enable
 stp edged-port
!
interface GigabitEthernet 0/4
 switch-port
 switch-port mode access
 switch-port access vlan 20
 stp bpduguard enable
 stp edged-port
!
interface GigabitEthernet 0/5
 switch-port
 switch-port mode access
 switch-port access vlan 30
 stp bpduguard enable
 stp edged-port
!
interface GigabitEthernet 0/6
 switch-port
 switch-port mode access
 switch-port access vlan 40
 stp bpduguard enable
 stp edged-port
```

7. Common Errors

- In the MSTP+VRRP topology, the MST region configurations on different devices are different.
- VLANs are not created before the mappings between instances and VLANs are configured.
- In the MSTP+VRRP topology, if some devices run STP or RSTP, the devices will be considered as different MST regions for spanning tree calculation. In this case, the calculated spanning tree will be different from expectations.

1.24.3 Configuring Spanning Tree Compatibility for MSTP Interfaces

1. Requirements

In the topology as shown in Figure 1-20, enable MSTP on Device A and Device B and configure the same instance mappings on the devices: Associate instance 1 with VLAN 10 and add port GigabitEthernet 0/1 to

VLAN 10; associate instance 2 with VLAN 20 and add port GigabitEthernet 0/2 to VLAN 20. Configure spanning tree compatibility for interfaces so that BPDUs sent by interfaces carry only information about instances corresponding to the VLANs, to which the interfaces belong.

2. Topology

Figure 1-20 Configuring Spanning Tree Compatibility for MSTP Interfaces



3. Notes

- Adopt the same configurations on Device A and Device B. The following uses Device A as an example.
- Create VLAN 10, VLAN 20, instance 1, and instance 2. Associate instance 1 with VLAN 10 and instance 2 with VLAN 20. Add port GigabitEthernet 0/1 to VLAN 10 and port GigabitEthernet 0/2 to VLAN 20, and configure spanning tree compatibility for the ports. Enable the STP function.

4. Procedure

(1) Create VLAN 10, VLAN 20, instance 1, and instance 2. Associate instance 1 with VLAN 10 and instance 2 with VLAN 20.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan range 10,20
DeviceA(cfg-vlan-range)# exit
DeviceA(cfg)# stp mst configuration
DeviceA(cfg-mst)# instance 1 vlan 10
DeviceA(cfg-mst)# instance 2 vlan 20
DeviceA(cfg-mst)# exit
```

(2) Add port GigabitEthernet 0/1 to VLAN 10 and port GigabitEthernet 0/2 to VLAN 20, and configure spanning tree compatibility for the ports.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port mode access
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port access vlan 10
DeviceA(cfg-if-GigabitEthernet 0/1)# stp compatible enable
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port mode access
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 20
DeviceA(cfg-if-GigabitEthernet 0/2)# stp compatible enable
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

(3) Enable the STP function.

```
DeviceA(cfg)# stp
DeviceA(cfg)# end
DeviceA# write
```

5. Verification

(1) If no spanning tree compatibility is configured for interfaces, check the spanning tree configuration as follows.

Device A becomes the root of Device B because its bridge ID is smaller. Ports GigabitEthernet 0/1 and GigabitEthernet 0/2 of Device A are designated ports in their respective instances.

```
DeviceA# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-4094
  Root ID      Priority    32768
               MAC Address      0074.9cee.53ca
               this bridge is root
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID   Priority    32768
               MAC Address      0074.9cee.53ca
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface          Role Sts Cost      Prio      OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1              Desg FWD 20000      128       False     P2p
Gi0/2              Desg FWD 20000      128       False     P2p

MST 1 vlans map : 10
  Region Root Priority    32768
               MAC Address      0074.9cee.53ca
               this bridge is region root

  Bridge ID   Priority    32768
              MAC Address      0074.9cee.53ca

Interface          Role Sts Cost      Prio      OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1              Desg FWD 20000      128       False     P2p
Gi0/2              Desg FWD 20000      128       False     P2p

MST 2 vlans map : 20
  Region Root Priority    32768
               MAC Address      0074.9cee.53ca
               this bridge is region root

  Bridge ID   Priority    32768
               MAC Address      0074.9cee.53ca
```

```
Interface            Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1                Desg FWD 20000        128          False       P2p
Gi0/2                Desg FWD 20000         128          False       P2p
```

Check Device B. In the spanning tree of instance 1, port GigabitEthernet 0/1 belonging to VLAN 10 is the root port and port GigabitEthernet 0/2 belonging to VLAN 20 is the alternate port. Port GigabitEthernet 0/2 rejects traffic of VLAN 10 to pass. Therefore, the alternate port cannot serve as an alternative to the root port. In the spanning tree of instance 2, port GigabitEthernet 0/2 belonging to VLAN 20 is the root port and port GigabitEthernet 0/1 belonging to VLAN 10 is the alternate port. Port GigabitEthernet 0/1 rejects traffic of VLAN 20 to pass. Therefore, the alternate port cannot serve as an alternative to the root port.

```
DeviceB# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-4094
  Root ID       Priority       32768
                MAC Address        0074.9cee.53ca
                this bridge is root
                Hello Time     2 sec    Forward Delay 15 sec    Max Age 20 sec

  Bridge ID    Priority       32768
                MAC Address        0074.9cee.f49e
                Hello Time     2 sec    Forward Delay 15 sec    Max Age 20 sec

Interface            Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1                Root FWD 20000         128          False       P2p
Gi0/2                Altn BLK 20000         128          False       P2p

MST 1 vlans map : 10
  Region Root Priority     32768
                MAC Address        0074.9cee.53ca
                this bridge is region root

  Bridge ID    Priority       32768
                MAC Address        0074.9cee.f49e

Interface            Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1                Root FWD 20000         128          False       P2p
Gi0/2                Altn BLK 20000         128          False       P2p

MST 2 vlans map : 20
  Region Root Priority     32768
                MAC Address        0074.9cee.53ca
                this bridge is region root
```

```
  Bridge ID   Priority      32768
              MAC Address       0074.9cee.f49e

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
Gi0/1               Altn BLK 20000         128        False      P2p
Gi0/2               Root FWD 20000         128        False      P2p
```

(2) After spanning tree compatibility is configured for interfaces, check the spanning tree topology on Device A. Device A becomes the root of Device B because its bridge ID is smaller.

The spanning tree is not pruned in instance 0. Ports GigabitEthernet 0/1 and GigabitEthernet 0/2 are designated ports.

```
DeviceA# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-4094
  Root ID      Priority      32768
               MAC Address       0074.9cee.53ca
               this bridge is root
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID   Priority      32768
               MAC Address       0074.9cee.53ca
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
Gi0/1               Desg FWD 20000         128        False      P2p
Gi0/2               Desg FWD 20000         128        False      P2p
```

The spanning tree is pruned in instance 1. Only port GigabitEthernet 0/1 belonging to VLAN 10 is the designated port.

```
MST 1 vlans map : 10
  Region Root Priority    32768
               MAC Address       0074.9cee.53ca
               this bridge is region root

  Bridge ID   Priority      32768
               MAC Address       0074.9cee.53ca

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
Gi0/1               Desg FWD 20000         128        False      P2p
```

The spanning tree is pruned in instance 2. Only port GigabitEthernet 0/2 belonging to VLAN 20 is the designated port.

```
MST 2 vlans map : 20
```

```
  Region Root Priority     32768
                MAC Address        0074.9cee.53ca
                this bridge is region root

  Bridge ID    Priority      32768
                MAC Address        0074.9cee.53ca

Interface            Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/2                Desg FWD 20000          128         False       P2p
```

(3) Check the spanning tree topology on Device B. Device A becomes the root of Device B because its bridge ID is smaller.

The spanning tree is not pruned in instance 0. Port GigabitEthernet 0/1 is the root port and port GigabitEthernet 0/2 is the alternate port.

```
DeviceB# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-4094
  Root ID       Priority      32768
                MAC Address        0074.9cee.53ca
                this bridge is root
                Hello Time     2 sec    Forward Delay 15 sec    Max Age 20 sec

  Bridge ID    Priority      32768
               MAC Address        0074.9cee.f49e
                Hello Time     2 sec    Forward Delay 15 sec    Max Age 20 sec

Interface            Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1                Root FWD 20000          128         False       P2p
Gi0/2                Altn BLK 20000          128         False       P2p
```

The spanning tree is pruned in instance 1. Only port GigabitEthernet 0/1 belonging to VLAN 10 is the root port and there is no alternate port.

```
MST 1 vlans map : 10
  Region Root Priority     32768
                MAC Address        0074.9cee.53ca
                this bridge is region root

  Bridge ID    Priority      32768
                MAC Address        0074.9cee.f49e

Interface            Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1                Root FWD 20000          128         False       P2p
```

The spanning tree is pruned in instance 2. Only port GigabitEthernet 0/2 belonging to VLAN 20 is the root port and there is no alternate port.

```
MST 2 vlans map : 20
  Region Root Priority     32768
                 MAC Address         0074.9cee.53ca
                 this bridge is region root

  Bridge ID   Priority       32768
                MAC Address         0074.9cee.f49e

Interface             Role Sts Cost           Prio        OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/2                  Root FWD 20000          128          False       P2p
```

6. Configuration Files

Configuration files of Devices A and B

```
stp mst configuration
 instance 0 vlan 1-9, 11-19, 21-4094
 instance 1 vlan 10
 instance 2 vlan 20
!
stp
!
vlan range 1,10,20
!
interface GigabitEthernet 0/1
switch-port
switch-port mode access
 switch-port access vlan 10
 stp compatible enable
!
interface GigabitEthernet 0/2
switch-port
switch-port mode access
 switch-port access vlan 20
 stp compatible enable
```

7. Common Errors

- If the VLAN lists configured on interfaces at both ends of a link are inconsistent, communication may fail after VLANs are pruned by the spanning tree compatibility function of the interfaces.
- If different instances are not configured or interfaces permit the traffic of all VLANs to pass, the spanning tree compatibility function cannot perform spanning tree pruning based on instances. In this case, there is no necessary to configure the spanning tree compatibility function.

1.24.4 Configuring BPDU Tunnel

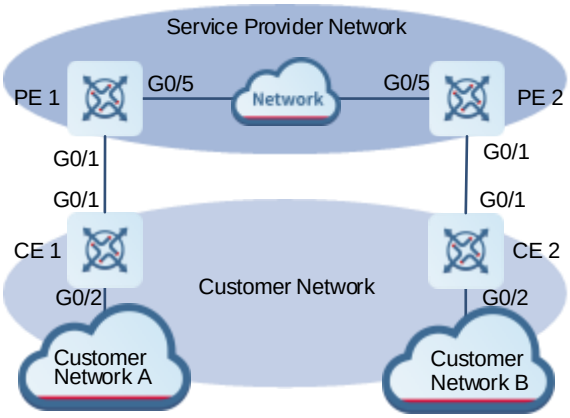
1. Requirements

In the typical QinQ network topology shown in Figure 1-21, the upper part is the service provider network while the lower part is the customer network. The service provider network consists of two PEs: PE 1 and PE 2. Customer Network A and Customer Network B are two sites of the same customer in different regions. CE 1 and CE 2 are access devices for connecting the customer networks to the service provider network, and they connect to the service provider network respectively through PE 1 and PE 2.

Configure a Layer 2 protocol tunnel to transparently transmit Layer 2 protocol packets from the customer network at one end to the customer network at the other end so that the spanning trees of the customer network and service provider network are separately calculated without interference. According to the needs of the customer, configure the STP tunnel function to transmit BPDUs from the customer network through a tunnel in the service provider network so that the spanning tree is calculated in a unified manner for the cross-region customer network, independent of the service provider network.

2. Topology

Figure 1-21 Configuring BPDU Tunnel



3. Notes

- Configure MSTP on PE 1 and PE 2. Configure the bridge priority to make PE 1 become the root of instances 0 and 1, and PE 2 the root of instance 2. Enable STP.
- Configure the QinQ encapsulation mode (for example, basic QinQ) on PE 1 and PE 2. Configure VLAN 10 as the native VLAN of Dot1q-tunnel ports and add VLAN 10 to the permitted untagged VLAN list of the tunnel ports. Configure VLAN 10 as the public network transmission channel. VLAN 10 is encapsulated into the user data packets before they are forwarded in the public network. Configure port GigabitEthernet 0/1 as a Dot1q-tunnel port and enable the STP tunnel function. Enable the global STP tunnel function.
- Configure the uplink ports of PE 1 and PE 2 as the uplink ports.
- Configure STP on CE 1 and CE 2. Set the bridge priority to 0, configure timer parameters, and enable STP on CE 1.
- Configure the uplink and downlink ports of CE 1 and CE 2 as trunk ports.

4. Procedure

(1) Configure MSTP on PE 1 and PE 2.

Set the bridge priority to 4094 for instances 0 and 1 and 8192 for instance 2 on PE 1.

```
PE1> privilege
PE1# config-system terminal
PE1(cfg)# vlan range 10,20,30,40
PE1(cfg-vlan-range)# exit
PE1(cfg)# stp mst configuration
PE1(cfg-mst)# instance 1 vlan 10,30
PE1(cfg-mst)# instance 2 vlan 20,40
PE1(cfg-mst)# stp mst 0 priority 4096
PE1(cfg)# stp mst configuration
PE1(cfg-mst)# stp mst 1 priority 4096
PE1(cfg)# stp mst configuration
PE1(cfg-mst)# stp mst 2 priority 8192
PE1(cfg)# stp
```

Set the bridge priority to 4094 for instance 2 and 8192 for instances 0 and 1 on PE 2.

```
PE2> privilege
PE2# config-system terminal
PE2(cfg)# vlan range 10,20,30,40
PE2(cfg-vlan-range)# exit
PE2(cfg)# stp mst configuration
PE2(cfg-mst)# instance 1 vlan 10,30
PE2(cfg-mst)# instance 2 vlan 20,40
PE2(cfg-mst)# stp mst 0 priority 8192
PE2(cfg)# stp mst configuration
PE2(cfg-mst)# stp mst 1 priority 8192
PE2(cfg)# stp mst configuration
PE2(cfg-mst)# stp mst 2 priority 4096
PE2(cfg)# stp
```

(2) Configure the basic QinQ transmission channel VLAN 10 on PE 1 and PE 2. Enable the interface STP tunnel function and global STP tunnel function. The configurations on PE 1 are the same as those on PE 2. The following uses PE 1 as an example.

```
PE1(cfg)# interface gigabitethernet 0/1
PE1(cfg-if-GigabitEthernet 0/1)# switch-port
PE1(cfg-if-GigabitEthernet 0/1)# switch-port mode dot1q-tunnel
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel pvid vlan 10
PE1(cfg-if-GigabitEthernet 0/1)# switch-port dot1q-tunnel allow-pass vlan add untagged 10
PE1(cfg-if-GigabitEthernet 0/1)# l2protocol-tunnel stp enable
PE1(cfg-if-GigabitEthernet 0/1)# exit
PE1(cfg)# l2protocol-tunnel stp
```

(3) Configure the uplink ports of PE 1 and PE 2 as the uplink ports. The configurations on PE 1 are the same as those on PE 2. The following uses PE 1 as an example.

```
PE1(cfg)# interface gigabitethernet 0/5
PE1(cfg-if-GigabitEthernet 0/5)# switch-port
PE1(cfg-if-GigabitEthernet 0/5)# switch-port mode uplink
PE1(cfg-if-GigabitEthernet 0/5)# switch-port trunk pvid vlan 1
PE1(cfg-if-GigabitEthernet 0/5)# switch-port trunk allow-pass vlan all
PE1(cfg-if-GigabitEthernet 0/5)# end
PE1# write
```

(4) Configure STP on CE 1 and CE 2.

Set the spanning tree mode to STP and bridge priority to 0, configure timer parameters, and enable STP on CE 1.

```
CE1> privilege
CE1# config-system terminal
CE1(cfg)# stp mode stp
CE1(cfg)# stp priority 0
CE1(cfg)# stp hello-time 4
CE1(cfg)# stp max-age 25
CE1(cfg)# stp forward-time 18
CE1(cfg)# stp
```

Set the spanning tree mode to STP and enable STP on CE 2.

```
CE2> privilege
CE2# config-system terminal
CE2(cfg)# stp mode stp
CE2(cfg)# stp
```

(5) Configure the uplink and downlink ports of CE 1 and CE 2 as trunk ports. The configurations on CE 1 are the same as those on CE 2. The following uses CE 1 as an example.

```
CE1(cfg)# interface list gigabitethernet 0/1-2
CE1(cfg-if-list)# switch-port
CE1(cfg-if-list)# switch-port mode trunk
CE1(cfg-if-list)# switch-port trunk pvid vlan 1
CE1(cfg-if-list)# switch-port trunk allow-pass vlan all
CE1(cfg-if-list)# end
CE1# write
```

5. Verification

(1) Check Dot1q-tunnel port configurations on PE 1 and PE 2. The configurations on PE 1 are the same as those on PE 2. The following uses PE 1 as an example.

```
PE1# display interfaces dot1q-tunnel
========Interface Gi0/1========
Native vlan: 10
Allowed vlan list:1,10,
Tagged vlan list:
```

(2) Verify that the STP tunnel function is enabled in both global configuration mode and interface configuration mode on PE 1 and PE 2. The configurations on PE 1 are the same as those on PE 2. The following uses PE 1 as an example.

```
PE1# display l2protocol-tunnel stp
L2protocol-tunnel: Stp Enable
L2protocol-tunnel destination mac address: 01d0.f800.0005
GigabitEthernet 0/1 l2protocol-tunnel stp enable
```

(3) Check that PE 1 and PE 2 have independent spanning tree topologies.

Check the spanning tree topology on PE 1. For instances 0 and 1, PE 1 is the root of PE 2 and port GigabitEthernet 0/5 is the designated port. For instance 2, PE 2 is the root of PE 1 and port GigabitEthernet 0/5 is the root port. Port GigabitEthernet 0/1 that connects to CE 1 of the customer network is not included in the spanning tree topologies of PE 1 and PE 2. The spanning tree time parameters (**Hello Time**: **2 sec**; **Forward Delay**: **15 sec**; **Max Age**: **20 sec**) of PE 1 use default values and are not affected by the configurations of CE 1.

```
PE1# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-29, 31-39, 41-4094
  Root ID      Priority    4096
               MAC Address      0074.9cee.f49e
               this bridge is root
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID   Priority    4096
               MAC Address      0074.9cee.f49e
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost      Prio      OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/5               Desg FWD 20000     128       False    P2p

MST 1 vlans map : 10, 30
  Region Root Priority    4096
               MAC Address      0074.9cee.f49e
               this bridge is region root

  Bridge ID   Priority    4096
               MAC Address      0074.9cee.f49e

Interface           Role Sts Cost      Prio      OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/5               Desg FWD 20000     128       False    P2p

MST 2 vlans map : 20, 40
  Region Root Priority    4096
               MAC Address      00d0.f8ee.8c1e
               this bridge is region root

  Bridge ID   Priority    8192
```

```
                MAC Address       0074.9cee.f49e

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/5               Root FWD 20000             128        False      P2p
```

Check the spanning tree topology on PE 2. For instances 0 and 1, PE 1 is the root of PE 2 and port GigabitEthernet 0/5 is the root port. For instance 2, PE 2 is the root of PE 1 and port GigabitEthernet 0/5 is the designated port. Port GigabitEthernet 0/1 that connects to CE 2 of the customer network is not included in the spanning tree topologies of PE 1 and PE 2. The spanning tree time parameters (**Hello Time**: **2 sec**; **Forward Delay**: **15 sec**; **Max Age**: **20 sec**) of PE 2 use default values and are not affected by the configurations of CE 1.

```
PE2# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : 1-9, 11-19, 21-29, 31-39, 41-4094
  Root ID      Priority      4096
               MAC Address       0074.9cee.f49e
               this bridge is root
               Hello Time     2 sec    Forward Delay 15 sec    Max Age 20 sec

  Bridge ID    Priority      8192
               MAC Address       00d0.f8ee.8c1e
               Hello Time     2 sec    Forward Delay 15 sec    Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/5               Root FWD 20000             128        False      P2p

MST 1 vlans map : 10, 30
  Region Root Priority     4096
               MAC Address       0074.9cee.f49e
               this bridge is region root

  Bridge ID    Priority      8192
               MAC Address       00d0.f8ee.8c1e

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/5               Root FWD 20000             128        False      P2p

MST 2 vlans map : 20, 40
  Region Root Priority     4096
               MAC Address       00d0.f8ee.8c1e
               this bridge is region root

  Bridge ID    Priority      4096
```

```
                MAC Address       00d0.f8ee.8c1e

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/5               Desg FWD 20000         128        False      P2p
```

(4) Check that CE 1 and CE 2 have independent spanning tree topologies.

Check the spanning tree topology on CE 1. CE 1 is the root of CE 2 and port GigabitEthernet 0/1 of CE 1 is the designated port. The spanning tree time parameters (**Hello Time**: **4 sec**; **Forward Delay**: **18 sec**; **Max Age**: **25 sec**) of CE 1 use configured values.

```
CE1# display stp summary
Spanning tree enabled protocol stp
  Root ID      Priority    0
               MAC Address       0074.9cee.53ca
               this bridge is root
               Hello Time    4 sec   Forward Delay 18 sec   Max Age 25 sec

  Bridge ID   Priority    0
               MAC Address       0074.9cee.53ca
               Hello Time    4 sec   Forward Delay 18 sec   Max Age 25 sec

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1               Desg FWD 20000         128        False      P2p
Gi0/2               Desg FWD 20000         128        False      P2p
```

Check the spanning tree topology on CE 2. CE 1 is the root of CE 2 and port GigabitEthernet 0/1 of CE 2 is the root port. The spanning tree time parameters (**Hello Time**: **4 sec**; **Forward Delay**: **18 sec**; **Max Age**: **25 sec**) of CE 2 use configured values of the root CE 1.

```
CE2# display stp summary
Spanning tree enabled protocol stp
  Root ID      Priority    0
               MAC Address       0074.9cee.53ca
               this bridge is root
               Hello Time    4 sec   Forward Delay 18 sec   Max Age 25 sec

  Bridge ID   Priority    32768
               MAC Address       0074.9c11.c9e6
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1               Root FWD 20000         128        False      P2p Bound(STP)
Gi0/2               Desg FWD 20000         128        False      P2p
```

6. Configuration Files

- PE 1 configuration file

```
l2protocol-tunnel stp
stp mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 30
 instance 2 vlan 20, 40
!
stp mst 0 priority 4096
stp mst 1 priority 4096
stp mst 2 priority 8192
stp
!
sysmac 0074.9cee.f49e
!
vlan range 1,10,20,30,40
!
interface GigabitEthernet 0/1
   switch-port
 switch-port mode dot1q-tunnel
 switch-port dot1q-tunnel allow-pass vlan add untagged 10
 switch-port dot1q-tunnel pvid vlan 10
 l2protocol-tunnel stp enable
!
interface GigabitEthernet 0/5
 switch-port
switch-port mode uplink
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

- PE 2 configuration file

```
l2protocol-tunnel stp
stp mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 30
 instance 2 vlan 20, 40
!
stp mst 0 priority 8192
stp mst 1 priority 8192
stp mst 2 priority 4096
stp
!
sysmac 00d0.f8ee.8c1e
!
vlan range 1,10,20,30,40
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode dot1q-tunnel
```

```
 switch-port dot1q-tunnel allow-pass vlan add untagged 10
 switch-port dot1q-tunnel pvid vlan 10
 l2protocol-tunnel stp enable
!
interface GigabitEthernet 0/5
 switch-port
switch-port mode uplink
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

- CE 1 configuration file

```
stp max-age 25
stp forward-time 18
stp hello-time 4
stp mode stp
stp mst 0 priority 0
stp
!
sysmac 0074.9cee.53ca
!
interface GigabitEthernet 0/1
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

- CE 2 configuration file

```
stp mode stp
stp
!
sysmac 0074.9c11.c9e6
!
interface GigabitEthernet 0/1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
   switch-port
 switch-port mode trunk
```

```
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

7. Common Errors

- The Layer 2 protocol tunnel function is not enabled in both global configuration mode and interface configuration mode. As a result, the BPDU tunnel function does not take effect.
- The BPDU tunnel addresses configured in the service provider network are inconsistent. As a result, BPDU frames from the customer network cannot be correctly transmitted.

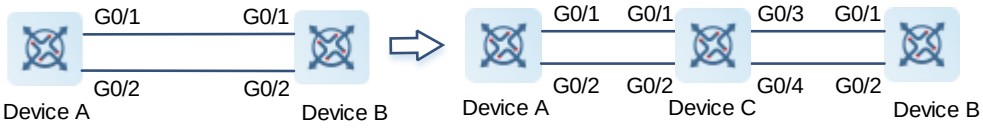
1.24.5 Configuring BPDU Transparent Transmission

1. Requirements

In the topology as shown in Figure 1-22, Device A is connected to Device B and MSTP is enabled on them. Device A is the root of Device B. One Device C with STP disabled is added between Device A and Device B. Configure the BPDU transparent transmission function on Device C to transmit STP packets of Device A and Device B so that Device A and Device B can still conduct calculation together and generate spanning trees.

2. Topology

Figure 1-22 Configuring BPDU Transparent Transmission



3. Notes

- On Device A, configure STP, set the bridge priority to 0 (smaller than the default value **32768**), and set the priority of port GigabitEthernet 0/2 to 16 (smaller than the default value **128**). In this way, Device A becomes the root bridge of Device B and port GigabitEthernet 0/2 of Device B becomes the root port.
- Enable STP on Device B.
- Configure BPDU transparent transmission on Device C.
- Configure interconnected ports as trunk ports.

4. Procedure

(1) On Device A, configure the interconnected ports as trunk ports. Configure the device priority and port priority and enable the STP function.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# stp mst 0 priority 0
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
```

```
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-GigabitEthernet 0/2)# stp mst 0 port-priority 16
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
DeviceA(cfg)# stp
```

(2) On Device B, enable the STP function and configure the interconnected ports as trunk ports.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# stp
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port trunk pvid vlan 1
DeviceB(cfg-if-GigabitEthernet 0/1)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
DeviceB(cfg)# interface gigabitethernet 0/2
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port trunk pvid vlan 1
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port trunk allow-pass vlan all
```

(3) On Device C, configure BPDU transparent transmission and configure the interconnected ports as trunk ports.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# unset stp
DeviceC(cfg)# bridge-frame forwarding protocol bpdu
DeviceC(cfg)# interface list gigabitethernet 0/1-4
DeviceC(cfg-if-rang)# switch-port
DeviceC(cfg-if-rang)# switch-port mode trunk
DeviceC(cfg-if-rang)# switch-port trunk pvid vlan 1
DeviceC(cfg-if-rang)# switch-port trunk allow-pass vlan all
```

5. Verification

(1) When Device C is not connected, check the spanning tree topology.

Check the spanning tree topology of Device A. Device A is the root.

```
DeviceA# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : ALL
  Root ID      Priority      0
               MAC Address       0074.9cee.53ca
               this bridge is root
```

```
                Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID   Priority      0
                MAC Address       0074.9cee.53ca
                Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
Gi0/1               Desg FWD 20000         128        False      P2p
Gi0/2               Desg FWD 20000         16         False      P2p
```

Check the spanning tree topology of Device B. Device A is the root of Device B. The port priority of Device A is changed. Therefore, port GigabitEthernet 0/2 of Device B becomes the root port and port GigabitEthernet 0/1 becomes an alternate port.

```
DeviceB# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : ALL
  Root ID      Priority      0
                MAC Address       0074.9cee.53ca
                this bridge is root
                Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID   Priority      8192
                MAC Address       00d0.f8ee.8c1e
                Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
Gi0/1               Altn BLK 20000         128        False      P2p
Gi0/2               Root FWD 20000         128        False      P2p
```

(2) After Device C is connected to the topology, check the spanning tree topology.

Check the spanning tree topology on Device A. The spanning tree topology of Device A keeps unchanged.

```
DeviceA# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : ALL
  Root ID      Priority      0
                MAC Address       0074.9cee.53ca
                this bridge is root
                Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID   Priority      0
                MAC Address       0074.9cee.53ca
                Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
```

```
Gi0/1              Desg FWD 20000        128        False      P2p
Gi0/2              Desg FWD 20000        16         False      P2p
```

Check the spanning tree topology on Device B. Device A is still the root of Device B, but port GigabitEthernet 0/1 of Device B becomes the root port, and port GigabitEthernet 0/2 becomes an alternate port. After Device C is configured to transparently transmit STP packets, ports GigabitEthernet 0/1 and 0/2 of Device B can receive BPDUs from port GigabitEthernet 0/2 of Device A. Therefore, the root port cannot be elected based on the port priority and port GigabitEthernet 0/1 of Device B is elected as the root port because of its smaller port ID.

```
DeviceB# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : ALL
  Root ID      Priority      0
               MAC Address       0074.9cee.53ca
               this bridge is root
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID    Priority      8192
               MAC Address       00d0.f8ee.8c1e
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface            Role Sts Cost          Prio       OperEdge Type
---------------- ---- --- ---------- -------- -------- ----------------
Gi0/1                Root FWD 20000         128        False      P2p
Gi0/2                Altn BLK 20000         128        False      P2p
```

6. Configuration Files

- Device A configuration file

```
stp mst 0 priority 0
stp
!
interface GigabitEthernet 0/1
  switch-port
  switch-port mode trunk
  switch-port trunk pvid vlan 1
  switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
  switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 stp mst 0 port-priority 16
```

- Device B configuration file

```
stp
```

```
interface GigabitEthernet 0/1
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

- Device C configuration file

```
bridge-frame forwarding protocol bpdu
!
interface GigabitEthernet 0/1
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/2
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/3
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
!
interface GigabitEthernet 0/4
   switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
switch-port trunk allow-pass vlan only 1-4094
```

1.24.6 Configuring V-STP

1. Requirements

In the topology as shown in Figure 1-23, Device A and Device B form an M-LAG Domain 1. aggregate interface

- GigabitEthernet 0/3 and 0/4 join AggregatePort 3, which is used as the peer link of M-LAG to synchronize data and transmit partial traffic.

As the dual-active detection link of Device A and Device B, Gigabit Ethernet 0/5 is used to detect the dual-active status of M-LAG devices when the Peer Link fails.

- The two GigabitEthernet 0/1 interfaces of the downstream Device C are added to AggregatePort 1 as VAP 1 of the M-LAG.
- The two GigabitEthernet 0/2 interfaces of the downstream Device D are added to AggregatePort 2 as VAP 2 of the M-LAG.

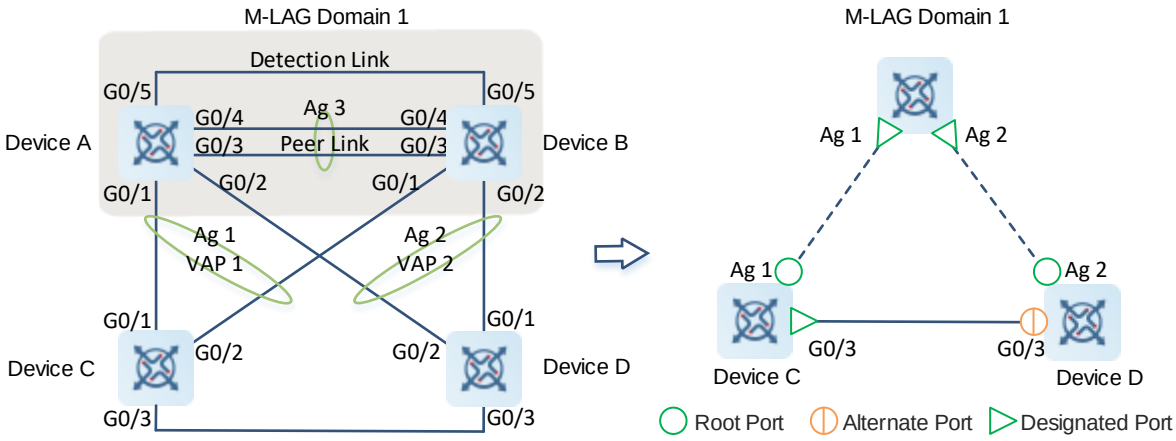
Uplink ports GigabitEthernet 0/1 and 0/2 of Device C are added to AggregatePort 1.

Uplink ports GigabitEthernet 0/1 and 0/2 of Device D are added to AggregatePort 2.

To prevent Layer 2 loops, you can configure a bridge priority of 4094 for Device A and Device B, enable the V-STP virtual spanning tree function on Device A and Device B to virtualize Device A and Device B into M-LAG Domain 1 for spanning tree calculation, and enable the normal spanning tree function on Device C and Device D.

2. Topology

Figure 1-23 Configuring V-STP and M-LAG



3. Procedure

(1) Configure the M-LAG function on Device A.

(2) Create AggregatePort 3 and configure it as a Peer Link port. Configure GigabitEthernet 0/3 and 0/4 as member ports of AggregatePort 3.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface aggregateport 3
DeviceA(cfg-if-AggregatePort 3)# switch-port
DeviceA(cfg-if-AggregatePort 3)# switch-port mode trunk
DeviceA(cfg-if-AggregatePort 3)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-AggregatePort 3)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-AggregatePort 3)# peer-link
DeviceA(cfg-if-AggregatePort 3)# exit
DeviceA(cfg)# interface gigabitethernet 0/3
DeviceA(cfg-if-GigabitEthernet 0/3)# descriptive-name link-to-DeviceB-G0/3-peerlink
```

```
DeviceA(cfg-if-GigabitEthernet 0/3)# port-group 3 mode active
DeviceA(cfg-if-GigabitEthernet 0/3)# exit
DeviceA(cfg)# interface gigabitethernet 0/4
DeviceA(cfg-if-GigabitEthernet 0/4)# descriptive-name link-to-DeviceB-G0/4-peerlink
DeviceA(cfg-if-GigabitEthernet 0/4)# port-group 3 mode active
DeviceA(cfg-if-GigabitEthernet 0/4)# exit
```

Configure the IP address of the Layer 3 SVI of the Peer Link as 20.20.20.1/24.

```
DeviceA(cfg)# interface vlan 1
DeviceA(cfg-if-VLAN 1)# ip address 20.20.20.1/24
DeviceA(cfg-if-VLAN 1)# exit
```

Configure the IP address of the local port Gigabit Ethernet 0/5 of the dual-active detection channel as 192.168.3.1/24.

```
DeviceA(cfg)# interface gigabitethernet 0/5
DeviceA(cfg-if-GigabitEthernet 0/5)# descriptive-name link-to-DeviceB-G0/5
DeviceA(cfg-if-GigabitEthernet 0/5)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/5)# ip address 192.168.3.1/24
DeviceA(cfg-if-GigabitEthernet 0/5)# exit
```

Configure the virtual aggregate interface Domain ID as 1 and enter the M-LAG configuration mode. Configure the local IP address of the data synchronization channel as 20.20.20.1, and the peer IP address as 20.20.20.2. Configure the local IP address of the dual-active detection channel as 192.168.3.1, and the peer IP address as 192.168.3.2.

```
DeviceA(cfg)# vap domain 1
DeviceA(cfg-vap-domain)# data-sync local 20.20.20.1 peer 20.20.20.2
DeviceA(cfg-vap-domain)# peer-keepalive local 192.168.3.1 peer 192.168.3.2
DeviceA(cfg-vap-domain)# exit
```

Configure VAP 1: Create AggregatePort 1 and join VAP 1. Configure GigabitEthernet 0/1 of the downstream Device C as a member port of AggregatePort 1.

```
DeviceA(cfg)# interface aggregateport 1
DeviceA(cfg-if-AggregatePort 1)# switch-port
DeviceA(cfg-if-AggregatePort 1)# switch-port mode trunk
DeviceA(cfg-if-AggregatePort 1)# switch-port trunk pvid vlan 1
DeviceA(cfg-if-AggregatePort 1)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-AggregatePort 1)# vap 1
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# descriptive-name  link-to-DeviceC-G0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# port-group 1 mode active
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Configure VAP 2: Create AggregatePort 2 and join VAP 2. Configure GigabitEthernet 0/2 of the downstream Device D as a member port of AggregatePort 2.

```
DeviceA(cfg)# interface aggregateport 2
DeviceA(cfg-if-AggregatePort 2)# switch-port
DeviceA(cfg-if-AggregatePort 2)# switch-port mode trunk
DeviceA(cfg-if-AggregatePort 2)# switch-port trunk pvid vlan 1
```

```
DeviceA(cfg-if-AggregatePort 2)# switch-port trunk allow-pass vlan all
DeviceA(cfg-if-AggregatePort 2)# vap 2
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# descriptive-namelink-to-DeviceD-G0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# port-group 2 mode active
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

(3) Configure the M-LAG function on Device B.

Create AggregatePort 3 and configure it as the Peer Link port. Configure GigabitEthernet 0/3 and 0/4 as member ports of AggregatePort 3.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface aggregateport 3
DeviceB(cfg-if-AggregatePort 3)# switch-port
DeviceB(cfg-if-AggregatePort 3)# switch-port mode trunk
DeviceB(cfg-if-AggregatePort 3)# switch-port trunk pvid vlan 1
DeviceB(cfg-if-AggregatePort 3)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-AggregatePort 3)# peer-link
DeviceB(cfg-if-AggregatePort 3)# exit
DeviceB(cfg)# interface gigabitethernet 0/3
DeviceB(cfg-if-GigabitEthernet 0/3)# descriptive-name link-to-DeviceA-G0/3-peerlink
DeviceB(cfg-if-GigabitEthernet 0/3)# port-group 3 mode active
DeviceB(cfg-if-GigabitEthernet 0/3)# exit
DeviceB(cfg)# interface gigabitethernet 0/4
DeviceB(cfg-if-GigabitEthernet 0/4)# descriptive-name link-to-DeviceA-G0/4-peerlink
DeviceB(cfg-if-GigabitEthernet 0/4)# port-group 3 mode active
DeviceB(cfg-if-GigabitEthernet 0/4)# exit
```

Configure the IP address of the Layer 3 SVI of the Peer Link as 20.20.20.2/24.

```
DeviceB(cfg)# interface vlan 1
DeviceB(cfg-if-VLAN 1)# ip address 20.20.20.1/24
DeviceB(cfg-if-VLAN 1)# exit
```

Configure the IP address of the local GigabitEthernet 0/5 of the dual-active detection channel as 192.168.3.2/24.

```
DeviceB(cfg)# interface gigabitethernet 0/5
DeviceB(cfg-if-GigabitEthernet 0/5)# descriptive-name link-to-DeviceA-G0/5
DeviceB(cfg-if-GigabitEthernet 0/5)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/5)# ip address 192.168.3.2/24
```

Configure the VAP Domain ID as 1 and enter the M-LAG configuration mode. Configure the local IP address of the data synchronization channel as 20.20.20.2, and the peer IP address as 20.20.20.1. Configure the local IP address of the dual-active detection channel as 192.168.3.2, and the peer IP address as 192.168.3.1.

```
DeviceB(cfg)# vap domain 1
DeviceB(cfg-vap-domain)# data-sync local 20.20.20.2 peer 20.20.20.1
DeviceB(cfg-vap-domain)# peer-keepalive local 192.168.3.2 peer 192.168.3.1
DeviceB(cfg-vap-domain)# exit
```

Configure VAP 1: Create AggregatePort 1 and join VAP 1. Configure GigabitEthernet 0/1 of the downstream Device C as a member port of AggregatePort 1.

```
DeviceB(cfg)# interface aggregateport 1
DeviceB(cfg-if-AggregatePort 1)# switch-port
DeviceB(cfg-if-AggregatePort 1)# switch-port mode trunk
DeviceB(cfg-if-AggregatePort 1)# switch-port trunk pvid vlan 1
DeviceB(cfg-if-AggregatePort 1)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-AggregatePort 1)# vap 1
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# descriptive-name link-to-DeviceC-G0/2
DeviceB(cfg-if-GigabitEthernet 0/1)# port-group 1 mode active
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Configure VAP 2: Create AggregatePort 2 and join VAP 2. Configure GigabitEthernet 0/2 of the downstream Device D as a member port of AggregatePort 2.

```
DeviceB(cfg)# interface aggregateport 2
DeviceB(cfg-if-AggregatePort 2)# switch-port
DeviceB(cfg-if-AggregatePort 2)# switch-port mode trunk
DeviceB(cfg-if-AggregatePort 2)# switch-port trunk pvid vlan 1
DeviceB(cfg-if-AggregatePort 2)# switch-port trunk allow-pass vlan all
DeviceB(cfg-if-AggregatePort 2)# vap 2
DeviceB(cfg)# interface gigabitethernet 0/2
DeviceB(cfg-if-GigabitEthernet 0/2)# descriptive-name link-to-DeviceD-G0/1
DeviceB(cfg-if-GigabitEthernet 0/2)# port-group 2 mode active
DeviceB(cfg-if-GigabitEthernet 0/2)# exit
```

(4) Create AggregatePort 1 on Device C, and configure the uplink interfaces GigabitEthernet 0/1 and 0/2 as member ports of AggregatePort 1.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface aggregateport 1
DeviceC(cfg-if-AggregatePort 1)# switch-port
DeviceC(cfg-if-AggregatePort 1)# switch-port mode trunk
DeviceC(cfg-if-AggregatePort 1)# switch-port trunk pvid vlan 1
DeviceC(cfg-if-AggregatePort 1)# switch-port trunk allow-pass vlan all
DeviceC(cfg-if-AggregatePort 1)# exit
DeviceC(cfg)# interface gigabitethernet 0/1
DeviceC(cfg-if-GigabitEthernet 0/1)# descriptive-name link-to-DeviceA-G0/1
DeviceC(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/1)# port-group 1 mode active
DeviceC(cfg-if-GigabitEthernet 0/1)# exit
DeviceC(cfg)# interface gigabitethernet 0/2
DeviceC(cfg-if-GigabitEthernet 0/2)# descriptive-name link-to-DeviceB-G0/1
DeviceC(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceC(cfg-if-GigabitEthernet 0/2)# port-group 1 mode active
DeviceC(cfg-if-GigabitEthernet 0/2)# exit
```

(5) Create AggregatePort 2 on Device D, and configure the uplink interfaces GigabitEthernet 0/1 and 0/2 as member ports of AggregatePort 2.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# interface aggregateport 2
DeviceD(cfg-if-AggregatePort 2)# switch-port
DeviceD(cfg-if-AggregatePort 2)# switch-port mode trunk
DeviceD(cfg-if-AggregatePort 2)# switch-port trunk pvid vlan 1
DeviceD(cfg-if-AggregatePort 2)# switch-port trunk allow-pass vlan all
DeviceD(cfg-if-AggregatePort 2)# exit
DeviceD(cfg)# interface gigabitethernet 0/1
DeviceD(cfg-if-GigabitEthernet 0/1)# descriptive-name link-to-DeviceB-G0/2
DeviceD(cfg-if-GigabitEthernet 0/1)# switch-port
DeviceD(cfg-if-GigabitEthernet 0/1)# port-group 2 mode active
DeviceD(cfg-if-GigabitEthernet 0/1)# exit
DeviceD(cfg)# interface gigabitethernet 0/2
DeviceD(cfg-if-GigabitEthernet 0/2)# descriptive-name link-to-DeviceA-G0/2
DeviceD(cfg-if-GigabitEthernet 0/2)# switch-port
DeviceD(cfg-if-GigabitEthernet 0/2)# port-group 2 mode active
DeviceD(cfg-if-GigabitEthernet 0/2)# exit
```

(6) Configure a virtual spanning tree on Device A and Device B. The default MSTP mode is adopted, and the spanning tree configuration on Device A and Device B must be consistent.

On Decide A, configure the bridge priority as 4096, and enable the V-STP spanning tree function.

```
DeviceA(cfg)# stp priority 4096
DeviceA(cfg)# stp
DeviceA(cfg)# stp v-stp enable
```

On Decide B, configure the bridge priority as 4096, and enable the V-STP spanning tree function.

```
DeviceB(cfg)# stp priority 4096
DeviceB(cfg)# stp
DeviceB(cfg)# stp v-stp enable
```

(7) Enable the normal spanning tree function on Device C and Device D.

Enable the spanning tree function on Device C.

```
DeviceC(cfg)# stp
```

Enable the spanning tree function on Device D.

```
DeviceD(cfg)# stp
```

4. Verification

(1) Run the **display vap summary** command to check whether M-LAG has been established successfully.

Check the establishment of M-LAG on Device A. Device A is the Master, AggregatePort 3 of the peer link is Up, and AggregatePort 1 and AggregatePort 2 of the downstream channel are Up.

```
DeviceA# display vap summary

M-LAG domain                    : 1
```

```
M-LAG role                          : Master
Local address                       : 20.20.20.1
Peer address                        : 20.20.20.2
Negotiation status                  : ok
Peer keepalive status               : alive
Consistency check mode              : disable

Peer-link           Status
-----------------------------
Ag3                 UP

Group ID   VAP Port   Local-Status   Peer-Status
-------------------------------------------------
1          Ag1        UP             UP
2          Ag2        UP             UP
```

Check the establishment of M-LAG on Device B. Device B is the Slave, AggregatePort 3 of the peer link is Up, and AggregatePort 1 and AggregatePort 2 of the downstream channel are Up.

```
DeviceB# display vap summary
M-LAG domain                        : 1
M-LAG role                          : Slave
Local address                       : 20.20.20.2
Peer address                        : 20.20.20.1
Negotiation status                  : ok
Peer keepalive status               : alive
Consistency check mode              : disable

Peer-link           Status
-----------------------------
Ag3                 UP

Group ID   VAP Port   Local-Status   Peer-Status
-------------------------------------------------
1          Ag1        UP             UP
2          Ag2        UP             UP
```

(2) Run the **display stp v-stp information** command to check whether the V-STP function is enabled.

Check whether the V-STP function is enabled on Device A. The value in Local bridge mac field indicates that the MAC address of local bridge is 00d0.f822.363c, and the value in Selected bridge mac field indicates that the MAC address of the virtual bridge is 00d0.f822.363c.

```
DeviceA# display stp v-stp information

V-STP status                : enable
Local bridge mac            : 00d0.f822.363c
Selected bridge mac         : 00d0.f822.363c
Peerlink Port               : AggregatePort3
Calculate Virtual Index     : 0
```

```
Mlag Remote device connected : Y

MST 0 Root Port              : None
```

Check whether the V-STP function is enabled on Device B. The value in Local bridge mac field indicates that the MAC address of local bridge is 8005.8822.b170, and the value in Selected bridge mac field indicates that the MAC address of the virtual bridge is 00d0.f822.363c.

```
DeviceB# display stp v-stp information

V-STP status                  : enable
Local bridge mac              : 8005.8822.b170
Selected bridge mac           : 00d0.f822.363c
Peerlink Port                 : AggregatePort 3
Calculate Virtual Index       : 4095
Mlag Remote device connected : Y

MST 0 Root Port              : None
```

(3) Run the **display stp summary** command to check the STP calculation results.

Check the STP calculation results of Device A. Device A (00d0.f822.363c) is the root. MAC address in the Bridge ID of the device is 00d0.f822.363c. AggregatePort 1 and AggregatePort 2 of Devcie A are Designated ports and are in Forwarding state.

```
DeviceA# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : ALL
  Root ID      Priority    4096
               MAC Address      00d0.f822.363c
               this bridge is root
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID    Priority    4096
               MAC Address      00d0.f822.363c
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
Ag1                 Desg FWD 19000        128        False      P2p
Ag2                 Desg FWD 19000        128        False      P2p
```

Check the STP calculation results of Device B. Device A (00d0.f822.363c) is the root. MAC address in the Bridge ID of the device is 00d0.f822.363c, not 8005.8822.b467, which is the MAC address of Device B, indicating that the V-STP function virtualizes the M-LAG active and standby devices into a bridge device M-LAG Domain 1 (00d0. f822.363c) for calculation, and the MAC address of Device A is the Bridge ID of M-LAG Domain 1. AggregatePort 1 and AggregatePort 2 of Devcie B are Designated ports and are in Forwarding state.

```
DeviceB#display stp summary
```

```
Spanning tree enabled protocol mstp
MST 0 vlans map : ALL
  Root ID      Priority    4096
               MAC Address      00d0.f822.363c
               this bridge is root
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID   Priority    4096
               MAC Address      00d0.f822.363c
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
Ag1                 Desg FWD 19000        128        True       P2p
Ag2                 Desg FWD 19000        128        False      P2p
```

Check the STP calculation results of Device C. M-LAG Domain 1 (00d0.f822.363c) formed by Device A and Device B is the root. AggregatePort 1 of Devcie C is the Root port, and GigabitEthernet 0/3 is the Designated port, and both ports are in Forwarding state.

```
DeviceC# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : ALL
  Root ID      Priority    4096
               MAC Address      00d0.f822.363c
               this bridge is root
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

  Bridge ID   Priority    32768
               MAC Address      0074.9cee.53ca
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
Gi0/3               Desg FWD 20000        128        False      P2p
Ag1                 Root FWD 19000        128        False      P2p
```

Check the STP calculation results of Device D. M-LAG Domain 1 (00d0.f822.363c) formed by Device A and Device B is the root. AggregatePort 2 of Devcie D is the Root port and is in Forwarding state. GigabitEthernet 0/3 is the Alternate port, and is in Blocking state.

```
DeviceD# display stp summary
Spanning tree enabled protocol mstp
MST 0 vlans map : ALL
  Root ID      Priority    4096
               MAC Address      00d0.f822.363c
               this bridge is root
               Hello Time    2 sec   Forward Delay 15 sec   Max Age 20 sec
```

```
  Bridge ID   Priority      32768
                MAC Address       00d0.f822.358c
                Hello Time     2 sec   Forward Delay 15 sec   Max Age 20 sec

Interface           Role Sts Cost          Prio       OperEdge Type
----------------- ---- --- ---------- -------- -------- ----------------
Gi0/3               Altn BLK 20000         128        False      P2p
Ag2                  Root FWD 19000         128        False      P2p
```

5. Configuration Files

- Device A configuration file

```
hostname DeviceA
stp mst 0 priority 4096
stp v-stp enable
stp
sysmac 00d0.f822.363c
vlan 1
!
interface GigabitEthernet0/1
 descriptive-name link-to-DeviceC-G0/1
 switch-port
 port-group 1 mode active
!
interface GigabitEthernet0/2
 descriptive-name link-to-DeviceD-G0/2
 switch-port
 port-group 2 mode active
!
interface GigabitEthernet0/3
 descriptive-name link-to-DeviceB-G0/3-peerlink
 switch-port
 port-group 3 mode active
!
interface GigabitEthernet0/4
 descriptive-name link-to-DeviceB-G0/4-peerlink
 switch-port
 port-group 3 mode active
!
interface GigabitEthernet0/5
 unset switch-port
 descriptive-name link-to-DeviceB-G0/5
 ip address 192.168.3.1 255.255.255.0
!
interface AggregatePort1
 switch-port
 switch-port mode trunk
```

```
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 vap 1
!
interface AggregatePort2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 vap 2
!
interface AggregatePort3
 unset mac-address-learning
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 peer-link
!
interface VLAN1
 ip address 20.20.20.1 255.255.255.0
!
vap domain 1
 data-sync local 20.20.20.1 peer 20.20.20.2
 peer-keepalive local 192.168.3.1 peer 192.168.3.2
```

- Device B configuration file

```
hostname DeviceB
stp mst 0 priority 4096
stp v-stp enable
stp
sysmac 8005.8822.b467
vlan 1
!
interface GigabitEthernet 0/1
 descriptive-name link-to-DeviceC-G0/2
 switch-port
 port-group 1 mode active
!
interface GigabitEthernet 0/2
 descriptive-name link-to-DeviceD-G0/1
 switch-port
 port-group 2 mode active
!
interface GigabitEthernet 0/3
 descriptive-name link-to-DeviceA-G0/3-peerlink
 switch-port
```

```
 port-group 3 mode active
!
interface GigabitEthernet 0/4
 descriptive-name link-to-DeviceA-G0/4-peerlink
 switch-port
 port-group 3 mode active
!
interface GigabitEthernet 0/5
 unset switch-port
 descriptive-name link-to-DeviceA-G0/5
 ip address 192.168.3.2 255.255.255.0
!
interface AggregatePort 1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 vap 1
!
interface AggregatePort 2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 vap 2
!
interface AggregatePort 3
 unset mac-address-learning
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
 peer-link
!
interface VLAN 1
 ip address 20.20.20.2 255.255.255.0
!
vap domain 1
 data-sync local 20.20.20.2 peer 20.20.20.1
 peer-keepalive local 192.168.3.2 peer 192.168.3.1
```

- Device C configuration file

```
hostname DeviceC
stp
sysmac 0074.9cee.53ca
vlan 1
!
```

```
interface GigabitEthernet 0/1
 descriptive-name link-to-DeviceA-G0/1
 switch-port
 port-group 1 mode active
!
interface GigabitEthernet 0/2
 descriptive-name link-to-DeviceB-G0/1
 switch-port
 port-group 1 mode active
!
interface AggregatePort 1
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
```

- Device D configuration file

```
hostname DeviceD
stp
sysmac 00d0.f822.358c
vlan 1
!
interface GigabitEthernet 0/1
 descriptive-name link-to-DeviceB-G0/2
 switch-port
 port-group 2 mode active
!
interface GigabitEthernet 0/2
 descriptive-name link-to-DeviceA-G0/2
 switch-port
 port-group 2 mode active
!
interface AggregatePort 2
 switch-port
 switch-port mode trunk
 switch-port trunk pvid vlan 1
 switch-port trunk allow-pass vlan only 1-4094
```

6. Common Errors

- If the M-LAG member devices are not connected successfully, the spanning tree cannot be virtualized successfully. In this case, check whether the state of the M-LAG aggregate interface is correct and whether the interface configuration is correct.
- V-STP is disabled on an M-LAG member device in the M-LAG scenario, leading to a V-STP calculation error.
- M-LAG member devices have different bridge priorities and MSTP region configurations, and APs added to the same VAP group have different port costs and port priorities, which result in a V-STP calculation error.

Contents

1 Configuring LLDP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 LLDP Packet 1
1.1.3 LLDPDU 2
1.1.4 TLV 3
1.1.5 LLDP Packet Transmission Mechanism 6
1.1.6 LLDP Packet Receiving Mechanism 6
1.1.7 Protocols and Standards 7
1.2 Restrictions and Guidelines 7
1.3 Configuration Task Summary 7
1.4 Enabling the LLDP Function 8
1.4.1 Overview 8
1.4.2 Restrictions and Guidelines 8
1.4.3 Procedure 8
1.5 Configuring the LLDP Packet Encapsulation Format 9
1.5.1 Overview 9
1.5.2 Procedure 9
1.6 Configuring LLDP Work Mode 9
1.6.1 Overview 9
1.6.2 Restrictions and Guidelines 9
1.6.3 Procedure 10

1.7 Configuring Interface Initialization Delay Time 10
1.7.1 Overview 10
1.7.2 Procedure 10
1.8 Configuring Packet Sending Interval 11
1.8.1 Overview 11
1.8.2 Restrictions and Guidelines 11
1.8.3 Procedure 11
1.9 Configuring TTL Multiplier 11
1.9.1 Overview 11
1.9.2 Restrictions and Guidelines 11
1.9.3 Procedure 11
1.10 Configuring LLDP Packet Sending Delay Time 12
1.10.1 Overview 12
1.10.2 Restrictions and Guidelines 12
1.10.3 Procedure 12
1.11 Configuring LLDP Packets Fast Sending Count 12
1.11.1 Overview 12
1.11.2 Procedure 12
1.12 Configuring LLDP Management Address 13
1.12.1 Overview 13
1.12.2 Restrictions and Guidelines 13
1.12.3 Procedure 13
1.13 Configuring Optional TLVs that an Interface can Publish 14
1.13.1 Overview 14

1.13.2 Restrictions and Guidelines 14
1.13.3 Procedure 16
1.14 Configuring Interface to Publish City Location 17
1.14.1 Overview 17
1.14.2 Restrictions and Guidelines 17
1.14.3 Procedure 18
1.15 Configuring Interface to Publish Emergency Contact Number 19
1.15.1 Overview 19
1.15.2 Restrictions and Guidelines 19
1.15.3 Procedure 19
1.16 Configuring Interface to Publish Network Policy TLV 20
1.16.1 Overview 20
1.16.2 Restrictions and Guidelines 20
1.16.3 Procedure 21
1.17 Configuring Ignore PVID Detection 22
1.17.1 Overview 22
1.17.2 Procedure 22
1.18 Configuring Compatible with Competitor Device Discovery 22
1.18.1 Overview 22
1.18.2 Procedure 22
1.19 Configuring LLDP Alarm Function 23
1.19.1 Overview 23
1.19.2 Procedure 23
1.20 Configuring LLDP Error Detection Function 24

1.20.1 Overview 24
1.20.2 Restrictions and Guidelines 24
1.20.3 Procedure 24
1.21 Configuring the Function of Changing the Value of the Port ID Field in LLDP Packets to the Short Port ID 24
1.21.1 Overview 24
1.21.2 Restrictions and Guidelines 24
1.21.3 Procedure 25
1.22 Monitoring 25
1.23 Configuration Examples 26
1.23.1 Checking the Neighbors Connected to the Device by Using the LLDP Function 26
1.23.2 Configuring the LLDP Error Detection Function 28

1 Configuring LLDP

1.1 Introduction

1.1.1 Overview

The Link Layer Discovery Protocol (LLDP) is a L2 discovery protocol defined in the IEEE 802.1AB standard. It is used to discover a topology and identify topological changes. LLDP encapsulates local information of a device into LLDP data units (LLDPDUs) in the type length value (TLV) format and then sends the LLDPDUs to neighbors. It also stores LLDPDUs from neighbors in the management information base (MIB) to be accessed by the network management system (NMS).

With LLDP, the NMS can learn about topology connections, for example, interfaces of a device connected to other devices, as well as the rates and duplex modes of interfaces at both ends of a link. Administrators can quickly locate and eliminate a fault based on the information. An LLDP-compliant device is capable of discovering neighbors when its peer is either an LLDP-compliant device or a device supporting the Link Layer Discovery Protocol-Media Endpoint Discovery (LLDP-MED).

1.1.2 LLDP Packet

An LLDP-compliant device advertises to neighbors its own information by sending LLDP packets, and obtains the information of neighbors by receiving LLDP packets. LLDP packets can be encapsulated in two formats: Ethernet II and SNAP (Sub Network Access Protocols).Figure 1-1 shows the format of an LLDP packet encapsulated in Ethernet frame format.

Figure 1-1 LLDP Packet Encapsulated in Ethernet Frame Format

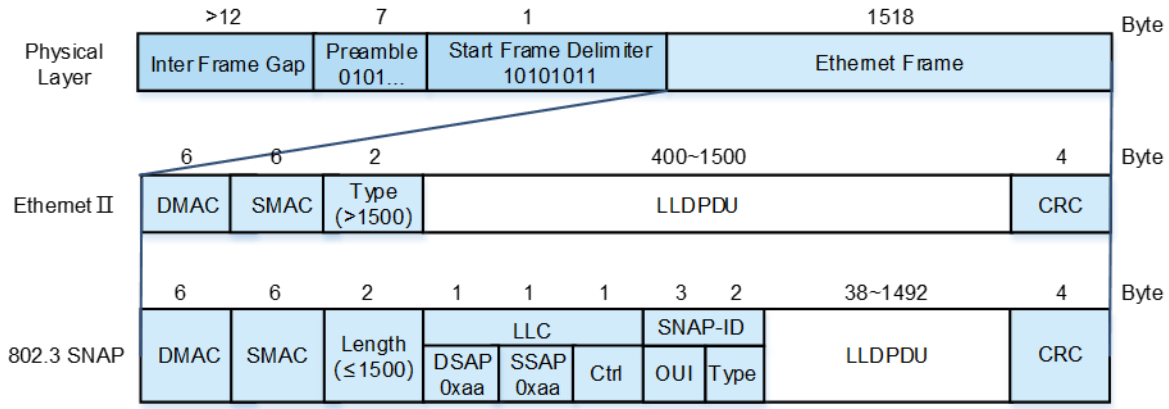


Table 1-1 describes the related fields of an LLDP packet encapsulated in Ethernet II format.

Table 1-1 Description of Fields in an LLDP Packet Encapsulated in Ethernet II Format

Field	Byte	Description
DMAC	6	Destination Address, which is a fixed multicast address 0180.c200.000e.
SMAC	6	Source Address, which is the MAC address of the device's interface.

Field	Byte	Description
Type	2	Ethernet type, which is 0x88CC.
LLDPDU	400-1500	LLDP data unit.
CRC	4	Frame check sequence.

Table 1-2 describes the related fields of an LLDP packet encapsulated in SNAP format.

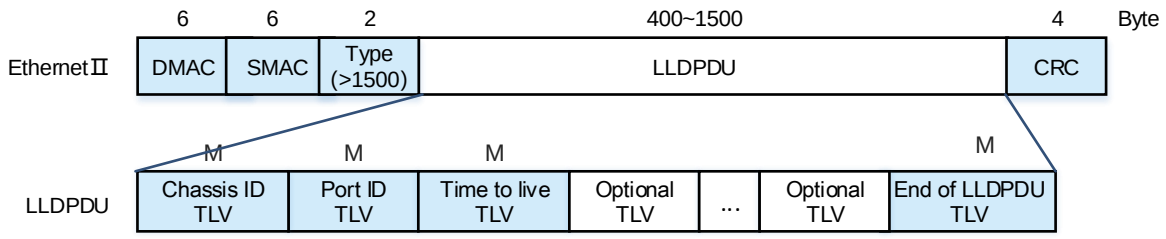
Table 1-2 Description of Fields in an LLDP Packet Encapsulated in SNAP Format

Field	Byte	Description
DMAC	6	Destination Address, which is a fixed multicast address 0180.c200.000e.
SMAC	6	Source Address, which is the MAC address of the device's interface.
DSAP	1	Destination Service Access Point. The value is 0xaa.
SSAP	1	Source Service Access Point. The value is 0xaa.
Control	1	Control information, for storing various control information and identify the type of frame. The value is 0x03.
OUI	3	Organizationally Unique Identifier, for distinguishing between basic management TLVs and TLVs defined by different organizations.
Type	2	Ethernet type, which is 0x88CC.
LLDPDU	38-1492	LLDP data unit.
CRC	4	Frame check sequence.

1.1.3 LLDPDU

The data unit in the LLDP packet is called LLDPDU (short for Link Layer Discovery Protocol Data Unit). According to the LLDP, each LLDPDU can contain 28 TLVs, starting from the Mandatory TLV - Chassis ID TLV, Port ID TLV and Time To Live TLV, Optional TLV in the middle, and LLDPDU TLV at the end. Figure 1-2 shows the format of LLDPDU, where M represents the mandatory TLV.

Figure 1-2 Format of LLDPDU



1.1.4 TLV

The data unit field in the LLDP frame is LLDPDU, which is composed of a series of variable length information elements. Each information element is called a TLV. The device supporting LLDP function records the status information of the device to the TLV to and advertise it to the neighbor, and obtains the status information of the neighbor through the TLV.

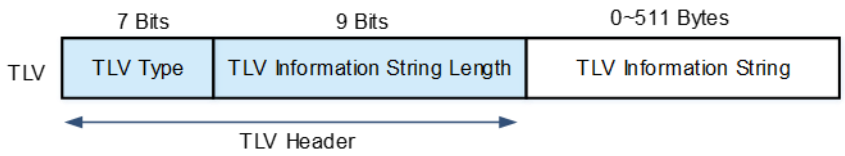
TLVs encapsulated in LLDPDUs are classified into basic management TLVs and organizationally specific TLVs.

Organizationally specific TLVs are defined by standard organizations or other institutions, for example, the IEEE 802.1 working group, the IEEE 802.3 working group, Internet Engineering Task Force (IETF), and equipment suppliers define their own TLV collections.

Organizationally specific TLVs are optional and are advertised in an LLDPDU selectively. Currently, there are three types of common organizationally specific TLVs: IEEE 802.1 organizationally specific TLVs, IEEE 802.3 organizationally specific TLVs, and LLDP-MED TLVs.

Figure 1-3 shows the structure of a TLV.

Figure 1-3 Structure of a TLV



- TLV Type: TLV type number. Different organizations advertise private information of devices by defining specific TLV types, and different TLV types are represented by numerical numbers in the TLV packet. Different organizations may define the same TLV Type number, so it is necessary to distinguish TLVs of different organizations based on the OUI (Organizationally Unique Identifier) field in the LLDP packet, and determine the TLV type based on the TLV Type value.
- TLV Information String Length: The length of TLV information. The 9-bit binary number can represent 0-511 bytes, so the length of TLV information is 0-511 bytes.
- TLV Information String: TLV information value.

1. Basic Management TLVs

Basic management TLVs are a collection of basic TLVs used for network management. This collection consists of two types of TLVs: mandatory TLVs and optional TLVs. A mandatory TLV must be contained in an LLDPDU for advertisement and an optional TLV is contained selectively. Table 1-3 describes the types of basic management TLVs.

Table 1-3 Basic Management TLVs

TLV Type	Description	Mandatory/Optional
End Of LLDPDU TLV	End flag of an LLDPDU, occupying two bytes.	Mandatory
Chassis ID TLV	Device ID, represented using a MAC address.	Mandatory
Port ID TLV	Interface sending an LLDPDU.	Mandatory

Time To Live TLV	TTL of local information on a neighbor. When a device receives a TLV with TTL of 0, it deletes the neighbor information.	Mandatory
Port Description TLV	Descriptor of the interface sending an LLDPDU.	Optional
System Name TLV	Device name.	Optional
System Description TLV	Device description, including the hardware version, software version, and operating system.	Optional
System Capabilities TLV	Main functions supported by the device, such as the bridge, routing, and relay functions.	Optional
Management Address TLV	Management address, which contains the interface ID and object identifier (OID).	Optional

Note

This LLDP-compliant device supports advertisement of basic management TLVs.

2. 802.1 TLV

- Table 1-4 describes IEEE 802.1 organizationally specific TLVs. All TLV types are optional.

Table 1-4 IEEE 802.1 Organizationally Specific TLVs

TLV Type	Description
Port VLAN ID TLV	TLV Type is 1. Virtual local area network (VLAN) identifier of an interface.
Port And Protocol VLAN ID TLV	TLV Type is 2. Protocol VLAN identifier of an interface.
VLAN Name TLV	TLV Type is 3. VLAN name of an interface.
Protocol Identity TLV	TLV Type is 4. Protocol type supported by an interface.

Note

This LLDP-compliant device does not send the protocol identity TLV but receives this TLV.

3. 802.3 TLV

- Table 1-5 describes IEEE 802.3 organizationally specific TLVs. All TLV types are optional.

Table 1-5 IEEE 802.3 Organizationally Specific TLVs

TLV Type	Description

TLV Type	Description
MAC/PHY Configuration//Status TLV	TLV Type is 1. Rate and duplex mode of an interface, and whether auto-negotiation is supported and enabled.
Power Via MDI TLV	TLV Type is 2. Power supply capacity of an interface.
Link Aggregation TLV	TLV Type is 3. Link aggregation capacity of an interface and the current aggregation state.
Maximum Frame Size TLV	TLV Type is 4. Maximum size of the frame that can be transmitted by an interface.

Note

This LLDP-compliant device supports advertisement of IEEE 802.3 organizationally specific TLVs.

- LLDP-MED TLV

LLDP-MED (Link Layer Discovery Protocol-Media Endpoint Discovery) is an extension to LLDP based on IEEE 802.1AB LLDP. It enables users to conveniently deploy the Voice over IP (VoIP) network-advertise and detect faults. It provides applications including the network configuration policies, device discovery, Power over Ethernet (PoE) management, and inventory management, which can meet requirements for low cost, effective management, and easy deployment and help simplify deployment of audio devices. Table 1-6 describes LLDP-MED TLVs. All TLV types are optional.

Table 1-6 LLDP-MED TLV

TLV Type	Description
LLDP-MED Capabilities TLV	Whether the device supports LLDP-MED, the type of the LLDP-MED TLV encapsulated into an LLDPDU, and device type (network connectivity device or endpoint device).
Network Policy TLV	Interface VLAN configuration, supported application type (such as voice or video services), and L2 priority information.
Location Identification TLV	Contains city location (common address and device type) and emergency contact number, which can be used to accurately locate endpoint devices in applications such as network topology collection.
Extended Power-via-MDI TLV	Extended power supply capacity.
Inventory – Hardware Revision TLV	Hardware version of a MED device.
Inventory – Firmware Revision TLV	Firmware version of a MED device.
Inventory – Software Revision TLV	Software version of a MED device.
Inventory – Serial Number TLV	Serial number of a MED device.

TLV Type	Description
Inventory – Manufacturer Name TLV	Name of the manufacturer of a MED device.
Inventory – Model Name TLV	Module name of a MED device.
Inventory – Asset ID TLV	Asset identifier of a MED device, used for inventory management and asset tracking.

Note

This LLDP-compliant device supports advertisement of LLDP-MED TLVs.

1.1.5 LLDP Packet Transmission Mechanism

LLDP-compliant devices can send LLDP packets to peers. LLDP packets are used to inform peers of their neighbors. When the LLDP transmission mode is canceled or the interface is disabled, LLDP packets can inform peers that their neighbors are no longer valid.

LLDP supports three work modes.

- TxRx: Transmits and receives LLDPDUs.
- Rx Only: Only receives LLDPDUs.
- Tx Only: Only transmits LLDPDUs.

LLDP periodically transmits LLDP packets when working in TxRx or Tx Only mode. When information about the local device changes, LLDP immediately transmits LLDP packets. You can configure an LLDP packet transmission delay to avoid frequent transmission of LLDP packets caused by frequent changes of local information. This delay can be manually configured.

LLDP provides two types of packets:

- Standard LLDP packet: It contains management and configuration information of the local device.
- Shutdown packet: When the LLDP work mode is disabled or the interface is shut down, LLDP shutdown packets will be transmitted. A shutdown packet consists of the chassis ID TLV, port ID TLV, TTL TLV, and end of LLDP TLV. TTL in the TTL TLV is 0. When a device receives an LLDP shutdown packet, it considers the neighbor information invalid and immediately deletes it.

When the LLDP work mode is changed from disabled or Rx to TxRx or Tx, or LLDP discovers a new neighbor (that is, a device receives a new LLDP packet and the neighbor information is not stored locally), the fast transmission mechanism is started so that the neighbor quickly learns the device information. The fast transmission mechanism enables a device to transmit a certain number of LLDP packets at an interval of 1 second.

1.1.6 LLDP Packet Receiving Mechanism

LLDP-compliant devices can receive LLDP packets from the peers. A device can discover its neighbors and determine when to age the neighbor information according to received LLDP packets.

A device can receive LLDP packets when working in TxRx or Rx Only mode. After receiving an LLDP packet, a device conducts a validity check. After the packet passes the check, the device checks whether the packet

provides information about a new neighbor or information update of an existing neighbor and stores the neighbor information locally. The device sets the TTL of neighbor information according to the value of TTL TLV in the packet. If the value of TTL TLV is 0, the neighbor information needs to be aged immediately.

1.1.7 Protocols and Standards

- IEEE 802.1AB 2005: IEEE Standard for Local and metropolitan area networks - Station and Media Access Control Connectivity Discovery
- ANSI/TIA-1057: Link Layer Discovery Protocol for Media Endpoint Devices

1.2 Restrictions and Guidelines

- LLDP can only be enabled on a physical interface, including Layer 2 Ethernet interface, Layer 3 Ethernet interface and management interface.
- For the aggregate interface, LLDP can be enabled on the member ports of the aggregate interface. The stack port and VSL (Virtual Switching Link) port also do not support LLDP.

1.3 Configuration Task Summary

The following tasks are optional:

(1) Enabling the LLDP Function

(2) Configure LLDP transmission mechanism

- Configuring the LLDP Packet Encapsulation Format
- Configuring LLDP Work Mode

(3) Configuring LLDP performance parameters

- Configuring Interface Initialization Delay Time
- Configuring Packet Sending Interval
- Configuring TTL Multiplier
- Configuring LLDP Packet Sending Delay Time
- Configuring LLDP Packets Fast Sending Count

(4) Configuring TLV Published in LLDPDU

- Configuring LLDP Management Address
- Configuring Optional TLVs that an Interface can Publish
- Configuring Interface to Publish City Location
- Configuring Interface to Publish Emergency Contact Number
- Configuring Interface to Publish Network Policy TLV

(5) Configuring Ignore PVID Detection

(6) Configuring Compatible with Competitor Device Discovery

(7) Configuring LLDP Alarm Function

(8) Configuring LLDP Error Detection Function

(9) Configuring the Function of Changing the Value of the Port ID Field in LLDP Packets to the Short Port ID

1.4 Enabling the LLDP Function

1.4.1 Overview

This section describes how to enable or disable the LLDP function globally and on an interface. The LLDP function is enabled globally and on an interface by default. The function configuration is optional.

1.4.2 Restrictions and Guidelines

- You can enable the LLDP function on a Layer 2 Ethernet interface, Layer 3 Ethernet interface and management interface.
- To make the LLDP function take effect on an interface, the LLDP function must be enabled globally and on the interface.
- You can run the **lldp enable** command in global configuration mode to enable the LLDP function globally, and run the **lldp enable** command in interface configuration mode to enable LLDP function on an interface.
- LLDP function is enabled by default. You can use the **unset lldp enable** command to disable the LLDP function.
- When a neighbor does not support LLDP but it is connected to an LLDP-supported device in downlink direction, an interface may learn information about the device that is not directly connected to the interface because the neighbor may forward LLDP packets.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the LLDP function globally.

lldp enable

The global LLDP function is enabled by default.

(4) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 3 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(5) Enable the LLDP function on the interface.

lldp enable

The LLDP function is enabled on interfaces by default.

1.5 Configuring the LLDP Packet Encapsulation Format

1.5.1 Overview

The same LLDP packet encapsulation format needs to be configured on the local device and its neighbors in interface configuration mode to ensure their normal communication.

You can use the **unset lldp encapsulation snap** command to configure the LLDP packet encapsulation format as SNAP, or use the **unset lldp encapsulation snap** command to restore the packet encapsulation format to Ethernet II.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 3 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(4) Set the LLDP packet encapsulation format to SNAP.

lldp encapsulation snap

The default LLDP packet encapsulation format is Ethernet II.

1.6 Configuring LLDP Work Mode

1.6.1 Overview

LLDP supports three work modes. This device has a default configuration, and configuration commands of this function are optional.

- TxRx: Transmits and receives LLDPDUs.
- Rx Only: Only receives LLDPDUs.
- Tx Only: Only transmits LLDPDUs.

When the LLDP work mode of the interface changes, the interface will initialize the protocol state machine. To avoid frequent initialization of the state machine caused by frequent changes in the interface work mode, the delay time of interface initialization can be configured in the global configuration mode.

1.6.2 Restrictions and Guidelines

- By default, the work mode of LLDP is **txrx**, indicating that the interface sends and receives LLDPDU. You can configure the work mode as **tx** or **rx** according to actual needs.

- When the LLDP work mode of the interface is **tx**, the interface can only send packets but not receive packets.
- When the LLDP work mode of the interface is **rx**, the interface can only receive packets but not send packets.

You can use the **unset lldp mode** command to disable LLDP work mode of the interface, and the interface will not send or receive LLDP packets.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 3 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(4) Configure the LLDP work mode.

lldp mode { **rx** | **tx** | **txrx** }

By default, the LLDP work mode is sending and receiving LLDPDU.

1.7 Configuring Interface Initialization Delay Time

1.7.1 Overview

To avoid frequent initialization of the state machine caused by frequent changes in the interface work mode, the delay time of interface initialization can be configured in the global configuration mode.

1.7.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the interface initialization delay time.

lldp timer reinit-delay *reinit-delay*

By default, the LLDP interface initialization delay time is 2 seconds.

1.8 Configuring Packet Sending Interval

1.8.1 Overview

The time interval for sending LLDP packet refers to the period for sending LLDP packet. A short period indicates high LLDP transmission frequency, increasing the system resource consumption. If the period is too long, the peer device may not discover the local device in time. According to the actual needs of the network, the time interval for sending LLDP packets can be configured to change the network topology discovery frequency.

1.8.2 Restrictions and Guidelines

- For standard MIB, the sending interval of LLDP packet is 5-32768 seconds.
- The value range of *tx-interval* is 1-32768, in seconds. The range is larger than that of the standard MIB to meet the requirements of more scenarios.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the sending interval of LLDP packets.

lldp timer tx-interval *tx-interval*

By default, the LLDP packet sending interval is 30 seconds.

1.9 Configuring TTL Multiplier

1.9.1 Overview

In the LLDP packet, TLV indicates the lifetime of the device's information on the neighbor device.

Time To Live TLV = TTL Multiplier × Packet Sending Interval + 1. TLV value can be modified by configuring TTL multiplier and LLDP packet sending interval. For example, if the TTL multiplier is 2, and the packet sending interval is 30 seconds, then the time to live of the LLDP packet of this device on the neighbor device is 2 × 30+1=61 seconds.

1.9.2 Restrictions and Guidelines

- You can use the **lldp hold-multiplier** *ttl-value* command to adjust the TTL multiplier in order to control the time to live of the device information on the neighbor device.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the TTL multiplier.

lldp hold-multiplier *ttl-value*

By default, the TTL multiplier of LLDP packet is 4.

1.10 Configuring LLDP Packet Sending Delay Time

1.10.1 Overview

When the status information of a local device changes, the local device will send a LLDP packet to neighbor devices immediately. To avoid frequently sending LLDP packets caused by frequent changes in local status information, you can configure the LLDP packet sending delay time to restrict the frequent sending of LLDP packets.

1.10.2 Restrictions and Guidelines

- The LLDP packet sending delay time is 2 seconds by default. *tx-delay* is the LLDP packet sending delay time configuration parameter, with a value range of 1-8192 seconds.
- If the value of delay time is too small, LLDP packets will be frequently sent due to the change of local status information. If the value of delay time is too large, the LLDP packet cannot be sent in time after a change occurred in the local status information. You are advised to configure the delay time according to the actual situation.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the LLDP packet sending delay time.

lldp timer tx-delay *tx-delay*

By default, the LLDP packet sending delay time is 2 seconds.

1.11 Configuring LLDP Packets Fast Sending Count

1.11.1 Overview

When a new neighbor is discovered, or the LLDP work mode is changed from Disabled or Rx to TxRx or Tx, the device will start the fast sending mechanism to enable the neighbor device to learn the information of the device as soon as possible. The fast sending mechanism will shorten the sending period of LLDP packet to 1 second, and continuously send a certain number of LLDP packets, and then resume the normal sending interval. You can use this function to configure the number of LLDP packets sent under the fast sending mechanism.

1.11.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the fast sending count of LLDP packets.

lldp fast-count *fast-count-value*

By default, the fast sending count of LLDP packets is 3.

1.12 Configuring LLDP Management Address

1.12.1 Overview

The LLDP management address is used to identify a device on the network and is transmitted in the Management Address TLV field in the basic management TLV.

The Management Address TLV is an optional basic management TLV, as shown in Table 1-3. However, in this device, the Management Address TLV is included in the LLDPDU by default. You can use this function to configure the content to be included in the Management Address TLV. The Management Address TLV may change if you configure the LLDP packet to publish the management address. If no management address is configured, the default management address will be used in LLDP packet.

1.12.2 Restrictions and Guidelines

- In the global configuration mode or interface configuration mode, the LLDP packet management address can be configured. The management address configured in the global configuration mode has priority over that configured in the interface configuration mode. When there is a global management address, the interface management address cannot be configured.

1.12.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the management address to be advertised in LLDP packets.

lldp management-address-tlv { *ipv4-address* | *ipv6-address* }

By default, the management address is the IPv4 address of the minimum VLAN supported by an interface. If no IPv4 address is configured for the VLAN, LLDP keeps searching for the next allowed minimum VLAN until a qualified IP address is found. If no IPv4 address is found, LLDP searches for the IPv6 address of the minimum VLAN supported by an interface. If no IPv6 address is found, the local address 127.0.0.1 is used as the management address.

(4) Configure the type for obtaining an LLDP management address. The configuration can change the management address selection sequence.

lldp management-address-tlv global-ip type {interface | loopback}

After this command is run to set the type for obtaining the global LLDP management address to **Interface**, the management address is obtained based on the following priority in descending order: configured LLDP management address, SVI IP address (a routed port uses its own IP address), local IP address 127.0.0.1, and system MAC address.

After this command is run to set the type for obtaining the global LLDP management address to **Loopback**, the management address is obtained based on the following priority in descending order: configured LLDP management address, IP address of the loopback interface, SVI IP address (a routed port uses its own IP address), local IP address 127.0.0.1, and system MAC address.

1.13 Configuring Optional TLVs that an Interface can Publish

1.13.1 Overview

You can use this function to configure types of optional TLVs that an interface can publish or not publish, so as to change the LLDPDU content in the LLDP packet sent by the interface.

- Table 1-3 shows basic management TLVs, where Chassis ID TLV, Port ID TLV, Time To Live TLV and End Of LLDPDU TLV are mandatory and do not need to be configured. For optional TLVs (except Management Address TLV), you can use this function to configure whether to publish these optional TLVs in the LLDPDU.
- Table 1-4 shows TLVs defined by IEEE 802.1. TLV types are optional. You can use this function to configure whether to publish Port VLAN ID TLV, Port And Protocol VLAN ID TLV and VLAN Name TLV in the LLDPDU. The LLDP protocol of this product does not support sending Protocol Identity TLVs, but supports receiving TLVs of this type.
- Table 1-5 shows TLVs defined by IEEE 802.3. TLV types are optional. You can use this function to configure whether to publish these TLVs in the LLDPDU.
- Table 1-6 shows TLVs defined by LLDP-MED. TLV types are optional. You can use this function to configure whether to publish these TLVs in the LLDPDU.

1.13.2 Restrictions and Guidelines

- Run the **lldp tlv-enable basic-tlv** command to configure optional basic management TLVs that an interface can publish. Run the **unset lldp tlv-enable basic-tlv** command to configure the interface not to publish optional basic management TLVs.
  - **all**: Port Description TLV, System Capabilities TLV, System Description TLV and System Name TLV in basic management TLVs.
  - **port-description**: Port Description TLV, indicating that the interface sends the interface description information of LLDPDU.
  - **system-capability**: System Capabilities TLV, which describes the main functions of the device, such as bridging, routing and relay.
  - **system-description**: System Description TLV, which describes the device hardware version, software version and operating system, etc.
  - **system-name**: System Name TLV, which describes the name of the device.
- Run the **lldp tlv-enable dot1-tlv** command to configure the interface to publish 802.1 TLV. Run the **unset lldp tlv-enable dot**1**-tlv** command to configure the interface not to publish 802.1 TLVs.
  - **all**: Port VLAN ID TLV, Port And Protocol VLAN ID TLV and VLAN Name TLV in 802.1 TLVs.
  - **port-vlan-id**: Port VLAN ID TLV, indicating port VLAN identifier.
  - **protocol-vlan-id** [*vlan-id*]: Port And Protocol VLAN ID TLV, indicating port and VLAN identifiers. *vlan-id* indicates the protocol VLAN ID of the interface. The value range is 1-4094.

- **vlan-name** [*vlan-id*]: VLAN Name TLV, indicating VLAN name of the interface. *vlan-id* indicates the protocol VLAN ID of the interface. The value range is 1-4094.

- Run the **lldp tlv-enable dot3-tlv** command to configure the interface to publish 802.3 TLVs. Run the **unset lldp tlv-enable dot3-tlv** command to configure the interface not to publish 802.3 TLVs.
  - **all**: All optional TLVs in 802.3 TLVs.
  - **link-aggregation**: Link Aggregation TLV, which describes the link aggregation capability and current aggregation state of the interface.
  - **mac-physic**: MAC/PHY Configuration/Status TLV, which describes the duplex status of the interface, the support of the automatic negotiation function and the disable or enable state of the function.
  - **max-frame-size**: Maximum Frame Size TLV, which describes the maximum size of frames that can be transmitted by the interface.
  - **power**: Power Via MDI TLV, which describes the power supply capability of the interface.
- Run the **lldp tlv-enable med-tlv** command to configure the interface to publish LLDP-MED TLVs. Run the **unset lldp tlv-enable med-tlv** command to configure the interface not to publish LLDP-MED TLVs.
  - **all**: The **all** parameter in the **lldp tlv-enable med-tlv** command indicates that the interface can publish all LLDP-MED TLVs except the Location Identification TLV. The **all** parameter in the **unset lldp tlv-enable med-tlv** indicates that the interface does not publish all LLDP-MED TLVs, including the Location Identification TLV.
  - **capability**: LLDP-MED Capabilities TLV, which describes whether the device supports LLDP-MED, LLDP-MED TLV type encapsulated in LLDPDU, and current device type (network connecting device or client).
  - **inventory**: Describes the media device directory management TLV, including hardware version, firmware version, software version, serial number, vendor name, module name, asset identifier, etc.
  - **power-over-ethernet**: Extended Power-via-MDI TLV, which indicates extended power supply capacity of the interface.
  - **location-id civic-location identifier** *id*: City location information in Location Identification TLV, including general address and device type. **identifier** *id* is the configured city location policy ID, and the value range is 1-1024. The LLDP city location information must be configured before the TLV can be published. See Configuring Interface to Publish City Location for details.
  - **location-id elin identifier** *id*: Emergency contact number information in Location Identification TLV. **identifier** *id* is the configured emergency contact number policy ID, and the value range is 1-1024. LLDP emergency contact number information must be configured first before the TLV can be published. See Configuring Interface to Publish Emergency Contact Number for details.
  - **network-policy profile** [*profile-number*]: Network Policy TLV, which advertises VLAN configuration, supported application types (such as voice or video), Layer 2/Layer 3 priority information of the interface. *profile-number* indicates the network policy ID, with a value range of 1-1024. When the device is connected to an IP phone, if the IP phone supports LLDP-MED, you can configure the Network Policy TLV to issue policies to the IP phone. You must first configure the Network Policy TLV policy, and then configure to allow the interface to publish this TLV. See Configuring Interface to Publish Network Policy TLV for details.

You can run the **display lldp tlv-config** [**interface** *interface-type interface-number*] command to view the TLV publish status on the interface, where the value of the STATUS field indicates the publish status. The value of the DEFAULT field indicates the default publish status.

- Follow the following steps to configure the interface to publish TLVs:
  a Allow the interface to publish LLDP 802.3 MAC/PHY Configuration/Status TLV.
  b Allow the interface to publish LLDP-MED Capabilities TLV.
  c Allow the interface to publish other types of LLDP-MED TLVs except Network Policy TLV.
- Follow the following steps to configure the interface not to publish TLVs:
  a Configure the interface not to publish other TLVs in LLDP-MED except the Network Policy TLV.
  b Configure the interface not to publish LLDP-MED Capabilities TLV.
  c Configure the interface not to publish MAC/PHY Configuration/Status TLV.

1.13.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 3 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(4) Configure the interface to publish TLVs. Configure at least one of following items.

- Configure the interface to publish optional basic management TLVs except Management Address TLV.

  lldp tlv-enable basic-tlv { **all** | **port-description** | **system-capability** | **system-description** | **system-name** }

  By default, the interface publishes Port Description TLV, System Capabilities TLV, System Description TLV, and System Name TLV.

- Configure the interface to publish 802.1 TLVs except Protocol Identity TLV.

  lldp tlv-enable dot1-tlv { **all** | **port-vlan-id** | **protocol-vlan-id** [*vlan-id*] | **vlan-name** [*vlan-id*] }

  By default, the interface publishes Port VLAN ID TLV, Port And Protocol VLAN ID TLV, and VLAN Name TLV.

- Configuration the interface to publish 802.3 TLVs.

  lldp tlv-enable dot3-tlv { **all** | **link-aggregation** | **mac-physic** | **max-frame-size** | **power** }

  By default, the interface publishes all 802.3 TLVs.

- Configuration the interface to publish LLDP-MED TLVs.

lldp tlv-enable med-tlv { **all** | **capability** | **inventory** | **location-id civic-location identifier** *id* | **location-id elin identifier** *id* | **network-policy profile** [*profile-number*] | **power-over-ethernet** }

By default, the interface publishes all LLDP-MED TLVs except Location Identification TLV.

1.14 Configuring Interface to Publish City Location

1.14.1 Overview

As shown in Table 1-6, the Location Identification TLV field in LLDP-MED TLV contains city location information (common address, device type) and emergency contact number, which are used to accurately locate endpoint devices in applications such as network topology collection.

You can use this function to configure the interface to publish city location information (common address, device type). First, create an LLDP Civic Address policy, and enter the LLDP Civic Address configuration mode to configure the common address and device type. Next, enter the interface configuration mode of a specified interface, and configure the interface to publish the city location information (common address, device type) in the Location Identification TLV.

1.14.2 Restrictions and Guidelines

- Use the **lldp location-id civic-location identifier** *id* to create an LLDP Civil Address policy, and enter the LLDP Civil Address configuration mode.
- In the LLDP Civic Address configuration mode, use the **civic-location** command to configure the common address information of the device. This parameter has two parts: address type *ca-type* and address information *ca-word*. The **civic-location** keyword is not reflected in the configuration command. The configuration starts with the parameter of the address type *ca-type*. Use the *ca-type ca-word* command to configure the address of the device, and use the **unset** *ca-type* command to delete the address information.
- Optional parameters of address type *ca-type* include: **country**; **state**; **county**; **city**; **division**; **neighborhood**; **street-group**; **leading-street-dir**; **trailing-street-suffix**; **street-suffix**; **number**; **street-number-suffix**; **landmark**; **additional-location-information**; **name**; **postal-code**; **building**; **unit**; **floor**; **room**; **type-of-place**; **postal-community-name**; **post-office-box**; and **additional-code**.
- Use the Address information *ca-word* field to configure specific information. Note that when the address type is **country**, use only two characters to represent the country. For example, **country** CH represents China.
- In the LLDP Civic Address configuration mode, use the **device-type** *device-type* command to to configure the device type. The default value of *device-type* is 1. The value 0 indicates a DHCP Server, 1 indicates a switch, and 2 indicates an LLDP MED client. You can use the **unset device-type** command to restore the default configuration.
- In the interface configuration mode, use the **lldp tlv-enable med-tlv location-id civic-location identifier** *id* command to configure the interface to publish LLDP city location information (common address, device type). **med-tlv** indicates that the published information is encapsulated in LLDP MED TLV (Link Layer Discovery Protocol Media Endpoint Discovery). **location-id civic-location identifier** *id* indicates that the interface is allowed to publish the city location information in the Location Identification TLV. The policy ID of this information is id, and the value range is 1-1024. The LLDP city location information must be configured before the TLV can be published.
- Multiple city location information policies can be configured. You can use the **lldp tlv-enable med-tlv location-id civic-location identifier** *id* command on different interfaces to publish different city location

information TLVs.

- You can use the **display lldp location-id civic-location** { **identifier** *id* | **interface** *interface-type interface-number* | **static** } command to view the configuration result.

1.14.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a LLDP city location information, and enter the LLDP Civil Address configuration mode.

lldp location-id civic-location identifier *id*

By default, no LLDP city location information exists.

(4) Configure the city location information.

(5) Configure common address information.

{ **country** | **state** | **county** | **city** | **division** | **neighborhood** | **street-group** | **leading-street-dir** | **trailing-street-suffix** | **street-suffix** | **number** | **street-number-suffix** | **landmark** | **additional-location-information** | **name** | **postal-code** | **building** | **unit** | **floor** | **room** | **type-of-place** | **postal-community-name** | **post-office-box** | **additional-code** } *ca-word*

By default, no common address information exists.

(6) Configure the device type information.

device-type *device-type*

By default, no device type information exists.

(7) Exit the global configuration mode.

exit

(8) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 3 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(9) Configure the interface to publish LLDP city location information.

lldp tlv-enable med-tlv location-id civic-location identifier *id*

By default, the interface does not publish LLDP city location information.

1.15 Configuring Interface to Publish Emergency Contact Number

1.15.1 Overview

As shown in Table 1-6, the Location Identification TLV field in LLDP-MED TLV contains city location information (common address, device type) and emergency contact number, which are used to accurately locate endpoint devices in applications such as network topology collection.

You can use this function to configure the interface to publish emergency contact number. First, configure an LLDP emergency contact number policy, and configure the emergency contact number. Next, enter the interface configuration mode of a specified interface, and configure the interface to publish the emergency contact number in the Location Identification TLV.

1.15.2 Restrictions and Guidelines

- In the global configuration mode, use the **lldp location-id elin identifier** *id* **elin-location** *telephone-number* command to configure an emergency contact number policy and telephone number.
- In the interface configuration mode, use the **lldp tlv-enable med-tlv location-id elin identifier** *id* command to configure the interface to publish LLDP emergency contact number. **med-tlv** indicates that the emergency contact number is encapsulated in LLDP MED TLV (Link Layer Discovery Protocol Media Endpoint Discovery). **location-id elin identifier** *id* indicates that the interface is allowed to publish the emergency contact number in the Location Identification TLV. The value range of the policy ID is 1-1024. LLDP emergency contact number must be configured first before the TLV can be published.
- Multiple emergency contact numbers can be configured. You can use the **lldp tlv-enable med-tlv location-id elin identifier** *id* command on different interfaces to allow the interface to publish different emergency contact number TLVs.
- You can use the **display lldp location-id elin-location** { **identifier** *id* | **interface** *interface-type interface-number* | **static** } command to view the configuration result.

1.15.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the emergency contact number of the device.

lldp location-id elin identifier *id* **elin-location** *telephone-number*

By default, no emergency contact number exists.

(4) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 3 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the management interface configuration mode.

interface mgmt *interface-number*

(5) Configure the interface to publish LLDP emergency contact number.

lldp tlv-enable med-tlv location-id elin identifier *id*

By default, the interface does not publish LLDP emergency phone number.

1.16 Configuring Interface to Publish Network Policy TLV

1.16.1 Overview

If the downstream IP phone supports LLDP (Link Layer Discovery Protocol), when it goes online, it will actively send LLDP packets. If the device is configured with LLDP Network Policy TLV, it can capture LLDP packets sent by the IP phone and send Voice VLAN policies to the IP phone. If a policy is configured to allow the IP phone to send tagged frames, and the IP phone supports sending tagged packets, the IP phone will add VLAN tags when sending voice packets according to the policy. If the policy requires the IP phone to send an untagged frame, the IP phone will send untagged packets.

1.16.2 Restrictions and Guidelines

- **network-policy profile** [*profile-number*]: Allow the interface to publish Network Policy TLV and advertise the VLAN configuration of the interface, supported application types (such as voice or video), Layer 2/Layer 3 priority information, etc. *profile-number* indicates the network policy ID, with a value range of 1-1024. When the device is connected to an IP phone, if the IP phone supports LLDP-MED, you can configure the Network Policy TLV to issue policies to the IP phone. You must first configure the Network Policy TLV policy, and then configure the interface to publish this TLV.
- **voice-vlan**: Specifies the application type of Voice. **Voice-vlan-signaling**: Specifies the application type of voice signaling. One LLDP Network Policy can only have one Voice application type and one Voice Signaling application type. Voice and Voice Signaling application types can exist at the same time.
- { *vlan-id* | **dot1p-priority** } [**cos** *cos* | **dscp** *dscp*]: Configures to send tagged frames in Voice VLAN.
  - *vlan-id*: Voice VLAN ID of the voice stream in the VLAN Tag. The value range is 1-4094. The voice packet will be tagged with this ID.
  - **dot1p-priority**: VLAN ID as 0 in the VLAN Tag, that is, the frame contains only the priority information *cos* and *dscp*.
  - **cos** *cos*: CoS (Class of Service) value of Voice VLAN voice traffic priority in the VLAN Tag. The value range is 0-7, and the default value is 5. The higher the value, the higher the priority. Ordinary VLAN packets has a CoS value of 0, with the lowest priority. By default, the CoS value of voice traffic packets entering Voice VLAN is increased to 6, which is higher than the priority of ordinary VLAN packets. The CoS value represents the Layer 2 priority, which is stored in the Layer 2 header of the packet and filled in the PRI field of the 802.1Q VLAN Tag.
  - **dscp** *dscp*: DSCP (Differentiated Services Code Point) value of Voice VLAN voice traffic priority. The value range is 0-63, and the default value is 46. The higher the value, the higher the priority. The DSCP value of ordinary IP packets is 0, with the lowest priority. The default value of Voice VLAN voice traffic packets has a DSCP value of 46, which is higher than that of ordinary IP packets. The DSCP value represents the IP priority (IP PRE) and is stored in the Layer 3 header of the packet. For IPv4 packets, the DSCP value is filled in the first six bits (bit0-bit5) of the ToS domain of the IPv4 packet header.

For IPv6 packets, the DSCP value is filled in the first 6 bits of Traffic Class field of the IPv6 packet header.

- **untagged**: Configures the VoIP device to send untagged frames. No VLAN ID or CoS values exist in the frame. The DSCP value exists in the IP header and cannot be configured using this command.
- **none**: Indicates that the LLDP network policy is not issued (that is, the default case), and the VoIP device determines the frame to be sent according to its own configuration.

In addition to creating LLDP Network Policy on the device, you also need to:

- Enable the Voice VLAN function, and statically add the interface connecting the IP phone to the Voice VLAN. If the downstream IP phone does not support LLDP-MED, you need to manually configure the MAC address of the IP phone to the Voice VLAN OUI list. See *Configuring Voice VLA*N for details.
- Configure QoS trusted interface priority modification. See *Configuring QoS* in the *ACL and QoS Configuration Guide* for details.
- Configure a secure channel if 802.1x Authentication is also enabled on the interface, in order to allow packets in Voice VLAN to pass through. See *Configuring ACL* in the *ACL and QoS Configuration Guide* for details.
- You can use the **display lldp tlv-config** [**interface** *interface-type interface-number*] command to view the TLV publish status on the interface. The STATUS field indicates the configured publish status, and the DEFAULT field indicates the default publish status.

1.16.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an LLDP network policy, and enter the LLDP Network Policy configuration mode.

lldp network-policy profile *profile-number*

By default, no LLDP network policy exists.

(4) Configure the Voice VLAN policy. The following configuration tasks are exclusive. Configure one of the following items.

- Configure the Voice VLAN policy to instruct IP phone to send tagged frames.

  { **voice-vlan** | **voice-vlan-signaling** } **vlan** { *voice-vlan-id* | **dot1p-priority** } [**cos** *cos* | **dscp** *dscp*]

  By default, no Voice VLAN policy exists.

- Configure the Voice VLAN policy to instruct the IP phone to send untagged frames. In this case, the interface can only be manually added to the Voice VLAN.

  { **voice-vlan** | **voice-vlan-signaling** } **vlan untagged**

  By default, no Voice VLAN policy exists.

- Cancel the Voice VLAN policy, and the IP phone determines the frame to be sent according to its own configuration.

  { **voice-vlan** | **voice-vlan-signaling** } **vlan none**

  By default, no Voice VLAN policy exists.

(5) Exit the global configuration mode.

exit

(6) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 3 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(7) Configure the interface to publish the Network Policy TLV policy.

lldp tlv-enable med-tlv network-policy profile *profile-number*

By default, the interface publishes the Network Policy TLV.

1.17 Configuring Ignore PVID Detection

1.17.1 Overview

You can configure the ignore PVID detection function on the device.

1.17.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the ignore PVID detection function.

lldp ignore pvid-error-detect

By default, the Ignore PVID detection function is disabled.

1.18 Configuring Compatible with Competitor Device Discovery

Note

This function is not supported by the S6920-4C, S6510-32CQ, S6520-64CQ, S6510-4C, or S6510-48VS8CQ.

1.18.1 Overview

You can configure the compatible with competitive device discovery function on the device.

1.18.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the compatible with competitive device discovery function.

lldp compliance vendor

By default, the compatible with competitive device discovery function is disabled.

1.19 Configuring LLDP Alarm Function

1.19.1 Overview

LLDP alarm function, namely LLDP Trap. By configuring the LLDP alarm function, you can send the LLDP information of the local device (such as discovering a new neighbor, detecting a communication link failure with a neighbor, and so on) to the network management server. The administrator can monitor the network operation according to this information.

To prevent the LLDP alarm information from being sent too frequently, the interval for sending LLDP alarms can be configured in the global configuration mode. During this interval, if LLDP information changes are detected, an alarm will be sent to the network management server.

1.19.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the interval for sending LLDP alarms.

lldp timer notification-interval *trap*

By default, the interval for sending LLDP alarm is 5 seconds.

(4) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 3 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(5) Enable the LLDP alarm function.

lldp notification remote-change enable

By default, the LLDP alarm function is disabled.

1.20 Configuring LLDP Error Detection Function

1.20.1 Overview

After this function is enabled, when an error is detected by the LLDP, a system log will be printed out and sent to the administrator.

Errors to be detected include VLAN configuration detection, interface status detection, interface aggregation configuration detection, MTU configuration detection and loop detection at both ends of the link.

1.20.2 Restrictions and Guidelines

- The LLDP error detection function depends on the devices at both ends of the link to send and receive specific TLVs in the LLDP packet. For this function to work properly, correct TLVs must be published by the device.

1.20.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the Layer 3 Ethernet interface configuration mode.

  interface { *ethernet-type interface-number* | **range** *ethernet-type interface-range* }

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(4) Enable the LLDP error detection function.

lldp error-detect

By default, the LLDP error detection function is enabled.

1.21 Configuring the Function of Changing the Value of the Port ID Field in LLDP Packets to the Short Port ID

1.21.1 Overview

After the function is enabled, the value of the port ID field carried in LLDP packets sent locally is changed to the short port ID.

1.21.2 Restrictions and Guidelines

- By default, the value of the port ID field carried in LLDP packets is the full name of the port. After this command is configured, the value of the port ID field is changed to the short port ID of the port.
- When this command is configured, the LLDP function is disabled and then enabled. Therefore, the following

message is displayed:

Warning: Configuring this command will reset lldp. Continue? [Y/N]:
The LLDP function will be enabled or disabled only after **Y** or **N** is entered.

1.21.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the function of changing the value of the port ID field in LLDP packets to the short port ID.

lldp short-port-id enable

By default, the value of the port ID field carried in LLDP packets is the full name of the port.

1.22 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** commands to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** commands to clear information.

Caution

Running the clear commands may lose vital information and thus interrupt services.

Table 1-7 Monitoring

Command	Purpose
cleanup lldp statistics [**interface** *interface-type interface-number*]	Clears LLDP statistics.
cleanup lldp table [**interface** *interface-type interface-number*]	Clears LLDP neighbor information.
display lldp local-information [**global** | **interface** interface-type interface-number]	Displays LLDP information on the local device, which will be organized as TLVs and sent to neighbors.
display lldp location-id { **civic-location | elin-location** } { **identifier** *id* **| interface** *interface-type interface-number* **| static** }	Displays the LLDP civic address or emergency contact number of the local device.

Command	Purpose
display lldp neighbors [interface *interface-type interface-number*] [**verbose** | **enhance**]	Displays the LLDP information of a neighbor.
display lldp network-policy { **profile** [*profile-num*] | **interface** *interface-type interface-number* }	Displays the LLDP network policy configuration of the local device.
display lldp statistics [**global | interface** *interface-type interface-number*]	Displays LLDP statistics.
display lldp status [**interface** *interface-type interface-number*]	Displays the LLDP status information.
display lldp tlv-config [**interface** *interface-type interface-number*]	Displays the configuration of TLVs to be advertised by an interface.
debugging lldp error	Debugs LLDP error processing.
debugging lldp event	Debugs LLDP event processing.
debugging lldp ha	Debugs LLDP hot backup processing.
debugging lldp packet	Debugs the LLDP packet receiving.
debugging lldp stm	Debugs the LLDP state machine.

1.23 Configuration Examples

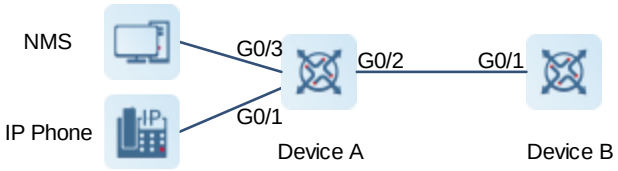
1.23.1 Checking the Neighbors Connected to the Device by Using the LLDP Function

1. Requirements

As shown in Figure 1-4, GigabitEthernet 0/1 of Device A is connected to the neighbor IP phone (MED device), and GigabitEthernet 0/2 is connected to the neighbor Device B. One NMS is connected to GigabitEthernet 0/3 of Device A. Check neighbor information on Device A by using the LLDP function.

2. Topology

Figure 1-4 Checking the Neighbors Connected to the Device by Using the LLDP Function



3. Notes

- The LLDP function is enabled by default and no configuration is required.

- The default LLDP work mode of an interface is TxRx and no configuration is required.
- The default LLDP packet transmission interval is 30 seconds, the default LLDP packet transmission delay is 2 seconds, and no configuration is required.

4. Verification

Check information about the neighbor connected to port GigabitEthernet 0/2 of Device A. It can be seen that GigabitEthernet 0/2 of Device A is connected to GigabitEthernet 0/1 of Device B, and the neighbor supports the bridging (B) and routing (R) functions.

```
DeviceA# display lldp neighbors interface gigabitethernet 0/2
Capability codes:
    (R) Router, (B) Bridge, (T) Telephone, (C) DOCSIS Cable Device
    (W) WLAN Access Point, (P) Repeater, (S) Station, (O) Other
System Name        Local Intf        Port ID        Capability        Aging-time
DeviceB            Gi0/2             Gi0/1          B, R              1minutes 49seconds

Total entries displayed: 1
```

Check details about the neighbor connected to GigabitEthernet 0/2 of Device A.

```
DeviceA# display lldp neighbors interface gigabitethernet 0/2 verbose
------------------------------------------------------------------------------
LLDP neighbor-information of port [GigabitEthernet 0/2]
------------------------------------------------------------------------------
  Neighbor index                    : 1
  Device type                       : LLDP Device
  Update time                       : 18minutes 14seconds
  Aging time                        : 1minutes 44seconds

  Chassis ID type                   : MAC address
  Chassis ID                        : 0074.9cee.f49e
  System name                       : DeviceB
  System description                : System description
  System capabilities supported     : Bridge, Router
  System capabilities enabled       : Bridge, Router

  Management address subtype        : ipv4
  Management address                : 172.26.147.182
  Interface numbering subtype       : ifIndex
  Interface number                  : 8193
  Object identifier                 :

  Management address subtype        : ipv6
  Management address                : 1010::1
  Interface numbering subtype       : ifIndex
  Interface number                  : 4197
  Object identifier                 :
```

```
Port ID type                    : Interface name
Port ID                         : Gi0/1
Port description                :

802.1 organizationally information
Port VLAN ID                    : 101
Port and protocol VLAN ID(PPVID)  : 0
    PPVID Supported             : NO
    PPVID Enabled               : NO
VLAN name of VLAN 101           : VLAN0101
Protocol Identity               :

802.3 organizationally information
Auto-negotiation supported      : YES
Auto-negotiation enabled        : YES
PMD auto-negotiation advertised :
Operational MAU type            : Speed(1000)/Duplex(Full)
PoE support                     : NO
Link aggregation supported      : YES
Link aggregation enabled        : NO
Aggregation port ID             : 0
Maximum frame Size              : 1500
```

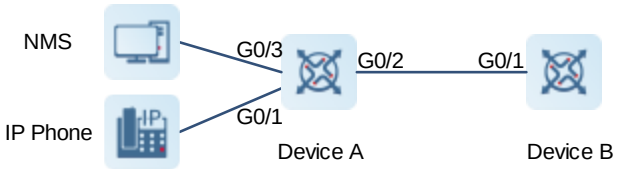
1.23.2 Configuring the LLDP Error Detection Function

1. Requirements

As shown in Figure 1-5, Device A and Device B are connected through Gigabit Ethernet (GE) interfaces. When an administrator configures the VLAN, port rate, duplex mode, aggregate interface, and port MTU, an error will be prompted if the configurations do not match those on the connected neighbor.

2. Topology

Figure 1-5 Configuring the LLDP Error Detection Function



3. Notes

- The LLDP function is enabled by default and no configuration is required.
- The default LLDP work mode of an interface is TxRx and no configuration is required.
- The LLDP error detection function is enabled by default and no configuration is required.
- The default LLDP packet transmission interval is 30 seconds, the default LLDP packet transmission delay is

2 seconds, and no configuration is required.

4. Verification

Forcibly set the rate of GigabitEthernet 0/2 of Device A to 100 Mbps. When the system prompts a mismatch with the rate and duplex mode of the connected port on the neighbor, the LLDP error detection function takes effect.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# speed 100
%Warning: the speed/duplex of port GigabitEthernet 0/2 may not match with it's neighbor.
```

IP Service Configuration

1. Configuring ARP
2. Configuring IPv4 Basics
3. Configuring DHCP
4. Configuring DHCP Client
5. Configuring DHCP Snooping
6. Configuring DNS
7. Configuring IPv6 Basics
8. Configuring DHCPv6
9. Configuring DHCPv6 Client
10. Configuring TCP
11. Configuring IP REF

Contents

1 Configuring ARP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Concepts 1
1.1.3 Principles 2
1.1.4 Protocols and Standards 5
1.2 Configuration Task Summary 6
1.3 Notes 6
1.4 Configuring a Static ARP Entry 7
1.4.1 Overview 7
1.4.2 Restrictions and Guidelines 7
1.4.3 Procedure 7
1.5 Configuring Dynamic ARP Learning Attributes 7
1.5.1 Overview 7
1.5.2 Configuration Tasks 7
1.5.3 Configuring the ARP Timeout Time 8
1.5.4 Configuring ARP Request Retransmission 9
1.5.5 Disabling Dynamic ARP Learning 9
1.5.6 Configuring Strict Dynamic ARP Learning 10
1.5.7 Configuring ARP Scanning 11
1.5.8 Configuring Scheduled Automatic ARP Scanning 12
1.6 Configuring ARP Entry Management 13

1.6.1 Overview 13
1.6.2 Configuration Tasks 13
1.6.3 Configuring a Limit on the Number of Unresolved ARP Entries 13
1.6.4 Configuring an ARP Learning Limit for an Interface 13
1.6.5 Configuring ARP Link Tracking 14
1.6.6 Configuring Fast ARP Entry Aging for an Interface 15
1.6.7 Configuring ARP Packet Rate Statistics Collection 16
1.6.8 Configuring the ARP Alarm Rate Limit 16
1.7 Enabling Trusted ARP 17
1.7.1 Overview 17
1.7.2 Procedure 17
1.8 Enabling Gratuitous ARP 17
1.8.1 Overview 17
1.8.2 Restrictions and Guidelines 18
1.8.3 Procedure 18
1.9 Enabling Proxy ARP 18
1.9.1 Overview 18
1.9.2 Procedure 19
1.10 Configuring Local Proxy ARP 19
1.10.1 Overview 19
1.10.2 Restrictions and Guidelines 19
1.10.3 Procedure 19
1.11 Configuring Any IP ARP 20
1.11.1 Overview 20

1.11.2 Restrictions and Guidelines 20
1.11.3 Procedure 20
1.12 Enabling ARP Trust Detection 21
1.12.1 Overview 21
1.12.2 Restrictions and Guidelines 21
1.12.3 Procedure 21
1.13 Enabling ARP-based IP Guard 22
1.13.1 Overview 22
1.13.2 Procedure 22
1.14 Configuring ARP Packet Filtering 23
1.14.1 Overview 23
1.14.2 Procedure 23
1.15 Restraining the Device from Sending ARP Requests to Authenticated VLANs 23
1.15.1 Overview 23
1.15.2 Restrictions and Guidelines 24
1.15.3 Procedure 24
1.16 Configuring ARP to Support VXLAN Functions 24
1.16.1 Overview 24
1.16.2 Restrictions and Guidelines 25
1.16.3 Procedure 25
1.17 Configuring Host Existence Judgment Prior to Proxy ARP Service 26
1.17.1 Overview 26
1.17.2 Restrictions and Guidelines 26
1.17.3 Procedure 26

1.18 Configuring ARP Entry Update During Active/Standby Switchover 26
1.18.1 Overview 26
1.18.2 Procedure 27
1.19 Sending ARP Requests to a Specific Sub VLAN 27
1.19.1 Overview 27
1.19.2 Procedure 27
1.20 Configuring the Function of Forbidding Sending Passerby Packets to the CPU 28
1.20.1 Overview 28
1.20.2 Restrictions and Guidelines 28
1.20.3 Procedure 28
1.21 Monitoring 28
1.22 Configuration Examples 30
1.22.1 Configuring Proxy ARP 30

1 Configuring ARP

1.1 Introduction

1.1.1 Overview

If two IP-enabled hosts need to communicate, a sender must learn the IP address and media access control (MAC) address of the receiver. With the MAC address, the sending host encapsulates IP datagrams into the data link layer (DLL) frames and transmits them over the physical network. The process of converting an IP address to a MAC address is called address resolution. There are two types of address resolution: Address Resolution Protocol (ARP) and proxy ARP.

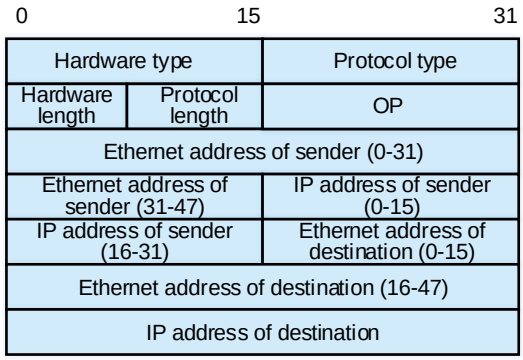
ARP maps an IP address into a MAC address. The mappings of IP and MAC addresses are stored in the ARP cache of the network device.

1.1.2 Basic Concepts

1. ARP Messages

An ARP message can be an ARP request or ARP reply. The format of an ARP message is as follows:

Figure 1-1 ARP Message Format



The description of fields in the ARP message is as follows:

- Hardware type: Indicates the type of a hardware address. For an Ethernet network, this value is **1**.
- Protocol type: Indicates the type of the protocol address to be mapped. For the IP protocol, this value is **0x0800**.
- Hardware length: Indicates the length of a hardware address, in bytes. For an Ethernet network, this value is **6**.
- Protocol length: Indicates the length of a protocol address, in bytes. For ARP requests or replies, this value is **4**.
- OP: Indicates the operation type. **1** indicates ARP requests, and **2** indicates ARP replies.
- Ethernet address of sender: Indicates a source MAC address, that is, the hardware address of a sender device.

- IP address of sender: Indicates a source IP address, that is, the IP address of a sender device.
- Ethernet address of destination: Indicates a destination MAC address, that is, the hardware address of a destination device. The value of this field is 0x00-00-00-00-00-00 for ARP requests.
- IP address of destination: Indicates a destination IP address, that is, the IP address of a destination device.

2. ARP Table

Each host maintains an ARP table that contains the mappings between IP addresses and MAC addresses. The mappings are called ARP entries.

Before sending an IP packet, a host first searches the ARP table for the MAC address corresponding to the destination IP address. If there is a corresponding MAC address in the ARP table, the host sends the IP packet directly to the MAC address instead of sending an ARP request message. If there is no corresponding MAC address in the ARP table, the host broadcasts an ARP request message to request the MAC address of the destination host. ARP entries can help greatly reduce the network traffic.

ARP entries can be classified as dynamic ARP entries or static ARP entries. The difference between them is as follows:

- Dynamic ARP entries are automatically generated and updated by the ARP protocol that exchanges ARP messages among devices, and will be aged and overwritten by new dynamic ARP entries.
- Static ARP entries are manually configured and maintained by the network administrator, and will not be aged or overwritten by dynamic ARP entries.

1.1.3 Principles

1. Static ARP Entries

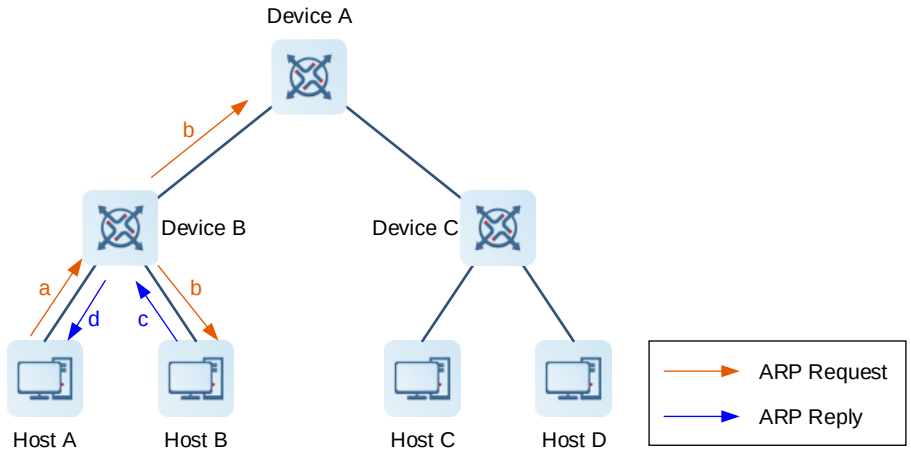
Static ARP entries are fixed mappings of IP and MAC addresses and they are created manually by the network administrator. The device cannot update the ARP entries. Static ARP entries can be deployed when communication security is a priority and network resources are sufficient.

2. Dynamic ARP Entries

Dynamic ARP allows devices to dynamically learn and update the mappings between IP and MAC addresses by exchanging ARP messages. Dynamic ARP entries can be deployed when real-time communication is priority or network resources are insufficient.

- **ARP implementation in the same network segment**

Figure 1-2 ARP Implementation in the Same Network Segment



As illustrated in Figure 1-2, hosts A and B are in the same network, host A has learned the IP address of host B and wants to send IP packets to host B. The address resolution process is as follows:

a Host A looks up in its ARP table and does not find the mapping between the IP and MAC addresses of host B. Host A then broadcasts an ARP request to request the MAC address of host B.

b After switch A receives the ARP request, Device B broadcasts the ARP request on the local network segment.

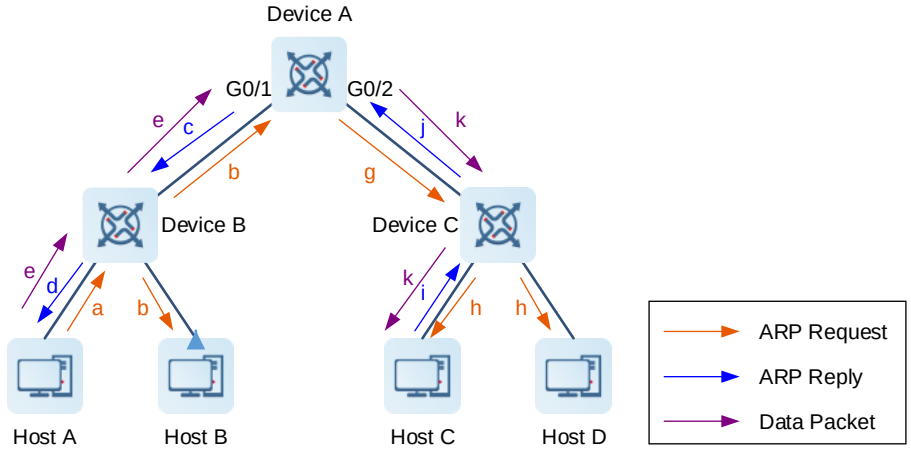
c After host B receives the ARP request, host B adds the MAC address of host A to its ARP table, and sends an ARP reply to host A. Since the destination IP address of the ARP request is not the IP address of router A, Device A directly discards the ARP request after receiving it.

d Device B forwards the ARP reply to the destination host A after receiving it.

e After host A receives the ARP reply, host A adds the MAC address of host B to its ARP table and sends the IP packets to host B.

- **ARP implementation in different network segments**

Figure 1-3 ARP Implementation in Different Network Segments



As illustrated in Figure 1-3, hosts A and B are in the same network, host A has learned the IP address of host C. For example, when host A sends an IP packet to host C, the address resolution process is as follows:

a Host A looks up in its ARP table and does not find the mapping between IP and MAC addresses of port GigabitEthernet 0/1 on the default gateway Device A that is reachable to host C. Host A then sends an ARP broadcast request to request the MAC address of port GigabitEthernet 0/1 of Device A.

f After receiving the ARP request, Device B broadcasts the packet on the local network segment.

g After receiving the ARP request, Device A adds the MAC address of host A into its ARP table, and sends an ARP reply to host A. Since the destination IP address of the ARP request is not the IP address of host B, host B directly discards the ARP request after receiving it.

h Device B forwards the ARP reply to the destination host A after receiving it.

i After receiving the ARP reply, host A adds the MAC address of port GigabitEthernet 0/1 of Device A into its ARP table and sends the IP packet to Device A.

j Device A looks up in the routing table and sends the IP packet from port GigabitEthernet 0/1 to port GigabitEthernet 0/2.

k Device A looks up in its ARP table and finds that there is no MAC address of host C. Device A then sends an ARP broadcast request to request the MAC address of host C.

l After Device C receives the ARP request, Device C broadcasts the packet on the local network segment.

m After host C receives the ARP request, host C adds the MAC address of port GigabitEthernet 0/2 of Device A into its ARP table, and sends an ARP reply to DeviceA. Since the destination IP address of the ARP request is not the IP address of host D, host D directly discards the ARP request after receiving it.

n Device C forwards the ARP reply to the destination Device A after receiving it.

o After Device A receives the ARP reply, Device A adds the MAC address of host C into its ARP table and sends the IP packet to host C.

3. Trusted ARP

When a user goes online on a GPRS support node (GSN) client, the authentication server obtains the user's real the mapping between IP and MAC addresses through the access device, and adds trusted ARP entries on the user's gateway device. This process is transparent to network administrators and does not require their extra work.

Trusted ARP entries have characteristics of both static and dynamic ARP entries, with a priority higher than that of dynamic ARP entries and lower than that of static ARP entries. Trusted ARP entries have an aging mechanism similar to that of dynamic ARP entries. When an ARP entry ages, the device actively sends an ARP request to detect whether the peer host exists. If the host sends a reply, the device regards the host active and updates the aging time of the ARP entry. Otherwise, the device deletes the ARP entry. Trusted ARP entries have characteristics of static ARP entries. The device cannot dynamically update information about the MAC addresses and interfaces in the trusted ARP entries by learning ARP messages.

Since trusted ARP entries come from authentic sources and will not be updated, they can efficiently prevent ARP spoofing targeting the gateway.

4. Gratuitous ARP

A gratuitous ARP message is a special ARP message. In a gratuitous ARP message, both the source and the destination IP addresses carried are the IP address of the local device.

Gratuitous ARP packets have the following purposes:

- Detecting IP address conflict If the device receives a gratuitous ARP packet and finds that the IP address in the packet is the same as its own IP address, it sends an ARP reply to notify the peer of the IP address conflict.
- Sending a gratuitous ARP packet to notify other devices to update ARP entries when the interface MAC address of a device changes. For example, in dual-node hot standby mode, when services are switched from the master device to the slave device, the MAC address of the previously slave device in the ARP entries on the connected device needs to be updated as that of the current master device. Generally, when services are switched from the master device to the slave device, the new master device after switchover sends a gratuitous ARP request to the network to notify other devices to update their ARP entries.
- Updating ARP entries based on gratuitous packets After receiving a gratuitous ARP message, the device checks whether a dynamic ARP entry corresponding to the source IP address in the gratuitous ARP packet exists. If yes, the device updates the ARP entry based on the information carried in the gratuitous ARP packet.

5. Proxy ARP

Proxy ARP allows devices on different subnets to resolve IP addresses into MAC addresses. If a device receives an ARP request whose source IP address is on a different network segment from the destination IP address and knows the route to the destination IP address, the device sends an ARP reply containing its own MAC address.

6. Local Proxy ARP

Local proxy ARP means that a device acts as a proxy ARP in the same virtual local area network (VLAN) that can be a common VLAN or sub VLAN.

After local proxy ARP is enabled on a device, the device can help hosts obtain the MAC addresses of other hosts in the VLAN. When port protection is enabled on the device, hosts connected to different ports are isolated at Layer 2. After receiving an ARP request, the device with local proxy ARP enabled serves as a proxy to send an ARP reply with its own Ethernet MAC address. In this case, hosts communicate with each other through Layer 3 routes.

7. Any IP ARP

Any IP ARP allows a device to access the network with any IP address. For example, a user is using a laptop in a hotel and wants to access the Internet without changing the configured IP address and gateway. After the gateway receives an ARP request and finds that the source IP address of the sender is in a different subnet and the destination address may not be its own address, the gateway still sends an ARP reply, with the MAC address of the gateway's Ethernet MAC address, and creates a direct route for the host.

1.1.4 Protocols and Standards

- RFC 826: An Ethernet Address Resolution Protocol
- RFC 1027: Using ARP to implement transparent subnet gateways

1.2 Configuration Task Summary

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring Dynamic ARP Learning Attributes
- Configuring ARP Entry Management
- Enabling Trusted ARP
- Enabling Gratuitous ARP
- Enabling Proxy ARP
- Configuring Local Proxy ARP
- Configuring Any IP ARP
- Configuring other advanced ARP functions
  - Enabling ARP Trust Detection
  - Enabling ARP-based IP Guard
  - Configuring ARP Packet Filtering
  - Restraining the Device from Sending ARP Requests to Authenticated VLANs
  - Configuring ARP to Support VXLAN Functions
  - Configuring Host Existence Judgment Prior to Proxy ARP Service
  - Configuring ARP Entry Update During Active/Standby Switchover
  - Sending ARP Requests to a Specific Sub VLAN

1.3 Notes

This section describes the precautions for configuring the ARP function.

Whether the ARP function is supported depends on devices.

Table 1-2 Product Supporting the ARP Function

Function	Product Supporting the Function
Restraining the Device from Sending ARP Requests to Authenticated VLANs Function	S6920-4C S6520-64CQ S6510-32CQ S6510-4C S6510-48VS8CQ S6250-48XS8CQ
Configuring ARP to Support VXLAN Functions Function	S6510-32CQ S6520-64CQ S6510-4C S6510-48VS8CQ S6250-48XS8CQ

Sending ARP Requests to a Specific Sub VLAN Function	S6920-4C S6520-64CQ S6510-32CQ S6510-4C S6510-48VS8CQ S6250-48XS8CQ

1.4 Configuring a Static ARP Entry

1.4.1 Overview

Users can manually configure static ARP entries to prevent the device from learning incorrect ARP entries, thus affecting the network connectivity.

1.4.2 Restrictions and Guidelines

After a static ARP entry is configured on a Layer 3 switch, the Layer 3 switch performs routing only after learning the physical port corresponding to MAC address carried in the static ARP entry.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static ARP entry.

arp [**vrf** *vrf-name* | **oob** [*mgmt-name*]] *ipv4-address mac-address arp-type*

No static ARP entry is configured by default.

1.5 Configuring Dynamic ARP Learning Attributes

1.5.1 Overview

The dynamic ARP learning function is enabled by default. You can configure ARP learning attributes as needed, for example, specify ARP entry timeout time, the number of times and interval that an ARP request can be transmitted consecutively, strict dynamic ARP learning function, ARP scanning, and scheduled automatic ARP scanning.

1.5.2 Configuration Tasks

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring the ARP Timeout Time
- Configuring ARP Request Retransmission
- Disabling Dynamic ARP Learning

- Configuring Strict Dynamic ARP Learning
- Configuring ARP Scanning
- Configuring Scheduled Automatic ARP Scanning

1.5.3 Configuring the ARP Timeout Time

1. Overview

Before an ARP entry times out, the device sends a unicast ARP request to detect whether the peer is online. If the device receives an ARP reply from the peer, it means that the peer is still online and the ARP entry is reserved. Otherwise, this ARP entry will be deleted. The ARP timeout time can be set as required.

2. Restrictions and Guidelines

The ARP timeout configuration only applies to dynamic mappings between IP and MAC addresses. When the ARP timeout time is set to a smaller value, the mapping table stored in the ARP cache is more accurate but ARP consumes more network bandwidth. In general, the ARP timeout time does not need to be set.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the switch virtual interface (SVI) configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure the ARP timeout time.

arp timeout *time*

The default timeout time of dynamic ARP entries in the ARP cache is **3600** seconds.

1.5.4 Configuring ARP Request Retransmission

1. Overview

The device sends a certain number of ARP requests at a certain time interval during address resolution until the device receives an ARP reply. This function is used to configure the interval and number of times that an ARP request can be transmitted consecutively.

2. Restrictions and Guidelines

The shorter the retransmission interval is, the faster the resolution is. The more times the ARP request can be transmitted consecutively, the more likely the resolution will succeed and the more bandwidth ARP will be consumed.

If the network resources are insufficient, you are advised to set the ARP request retransmission interval to a larger value and the number of retransmission times to a smaller value, to reduce bandwidth consumption.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the ARP request retransmission interval.

arp retry interval *interval*

The default ARP request retransmission interval is 1 second.

(4) Configure the number of times that an ARP request can be transmitted consecutively.

arp retry times *times*

The default number of times that an ARP request can be transmitted consecutively is 5. That is, if no ARP reply packet is received for the first ARP request, the ARP request packet will be retransmitted for four consecutive times.

1.5.5 Disabling Dynamic ARP Learning

1. Overview

The dynamic ARP learning function can be disabled on an interface. Then, the interface will not perform dynamic ARP learning.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Disable dynamic ARP learning.

unset arp-learning enable

The ARP learning function is enabled by default.

1.5.6 Configuring Strict Dynamic ARP Learning

1. Overview

After this function is enabled on a device, only the ARP replies in response to the ARP requests actively sent by the device can trigger the device to learn ARP entries.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable strict dynamic ARP learning.

arp strict-learning enable

Strict dynamic ARP learning is disabled by default.

1.5.7 Configuring ARP Scanning

1. Overview

This function is usually used together with the Web-based dynamic-to-static ARP entry conversion function. After ARP scanning is enabled on an interface, the device scans neighbors in the specified range and learns the ARP entries of these neighbors.

2. Restrictions and Guidelines

The device can scan the neighbors in the network segments, to which the primary IP address and secondary IP addresses belong, on an interface or scans only the neighbors in the specified address range. ARP scanning takes effect once after it is configured, and can be executed when the Layer 3 interface is up (that is, the link is up and an IP address is configured for the interface). However, ARP scanning configuration cannot be saved and will lose effect the next time.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable ARP scanning.

arp scan [*start-ipv4-address end-ipv4-address*]

ARP scanning is disabled by default.

1.5.8 Configuring Scheduled Automatic ARP Scanning

1. Overview

After scheduled automatic ARP scanning is enabled on an interface, the device scans neighbors in the specified range and learns the ARP entries of these neighbors. If ARP entries are not learned from neighbors, after the current scanning ends, the device waits for the scanning interval (5 minutes by default) and then starts next scanning until the entries are learned.

2. Restrictions and Guidelines

- Since this function adds a strict confirmation procedure in the ARP learning process, it affects the efficiency of ARP learning.
- The scanning interval ranges from 1 to 30, in minutes.
- The scanning rate ranges from 1 to 100, in IP addresses per second.
- Up to 1000 instances can be configured.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable scheduled automatic ARP scanning.

arp scan auto *start-ipv4-address end-ipv4-address*

The scheduled automatic ARP scanning function is disabled by default.

(5) (Optional) Configure the scheduled automatic ARP scanning interval.

arp scan interval *time*

The default interval of scheduled automatic ARP scanning is 5 minutes.

(6) (Optional) Configure the scheduled automatic ARP scanning rate.

arp scan rate *rate-value*

The default scheduled automatic ARP scanning rate is **20** IP addresses per second.

1.6 Configuring ARP Entry Management

1.6.1 Overview

Entries in the ARP cache can be managed as required, such as limiting the capacity of ARP entries and configuring fast aging of ARP entries.

1.6.2 Configuration Tasks

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring a Limit on the Number of Unresolved ARP Entries
- Configuring an ARP Learning Limit for an Interface
- Configuring ARP Link Tracking
- Configuring Fast ARP Entry Aging for an Interface
- Configuring ARP Packet Rate Statistics Collection
- Configuring the ARP Alarm Rate Limit

1.6.3 Configuring a Limit on the Number of Unresolved ARP Entries

1. Overview

In a local area network (LAN), ARP attacks and scanning may cause a large number of unresolved ARP entries generated on the gateway. As a result, the gateway fails to learn the MAC addresses of the hosts. To prevent this situation, if a large number of unresolved entries exist in the ARP cache and remain in the table after a while, you are advised to use this command to limit the number of unresolved ARP entries.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a limit on the number of unresolved ARP entries.

arp unresolve *unresolved-number*

The maximum number of unresolved ARP entries that can be held in an ARP cache is the total capacity of ARP entries by default.

1.6.4 Configuring an ARP Learning Limit for an Interface

1. Overview

Limiting the number of ARP entries that can be learned by an interface can flexibly control the on-demand assignment of ARP entry resources, to prevent resource waste. In addition, this can prevent malicious ARP

attacks. When the device is attacked maliciously and the ARP entry learning number is not limited, the device will generate a large number of ARP entries, which occupy excessive entry resources.

2. Restrictions and Guidelines

The configured value must be equal to or greater than the number of the ARP entries learned by the interface. Otherwise, the configuration does not take effect. The configuration is subject to the ARP entry capacity supported by the device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure an ARP learning limit for the interface.

arp interface cache-limit *limit*

The number of ARP entries that can be learned by an interface is not limited by default.

1.6.5 Configuring ARP Link Tracking

1. Overview

ARP link tracking is used to delete the ARP entry corresponding to the IP address of a tracked object on the device when the tracked object becomes unreachable. If the device needs to send a packet destined to this IP address later, it performs ARP learning again. After active/standby switchover of downlink devices, if the master device does not actively send gratuitous ARP packets, the uplink device fails to update its MAC address table timely and discards the packets destined to the IP address of the master device. When this function is configured, the device deletes the corresponding ARP entry if the destination IP address becomes unreachable. Then, the downlink master device is triggered to send an ARP reply through ARP learning, so as to help the uplink device update the MAC address table about the downlink master device.

2. Restrictions and Guidelines

- This function can be configured only on switch virtual interfaces (SVIs).
- In gateway authentication mode, the ARP entry of a tracked object in an authenticated VLAN cannot be deleted when the tracked object is unreachable.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable ARP link tracking.

arp [**vrf** *vrf-name*] *ipv4-address* **track** *object-number*

ARP link tracking is disabled by default.

1.6.6 Configuring Fast ARP Entry Aging for an Interface

1. Overview

If a host goes offline or encounters an abnormal condition, the device fails to learn the ARP entry. When the MAC address corresponding to the ARP entry ages, the dynamic ARP entry starts aging one hour after the aging time by default. If this feature is enabled on an interface, after the MAC address ages, the dynamic ARP entry ages rapidly to realize fast route convergence of the host.

2. Restrictions and Guidelines

- This function can be configured only on SVIs and OVXLAN routed interfaces.
- When the conversion of ARP entries into host routes is enabled on the device, you are advised to also enable this function to help achieve fast route convergence.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable fast ARP entry aging.

arp fast-aging enable

The fast ARP entry aging function is disabled by default.

1.6.7 Configuring ARP Packet Rate Statistics Collection

1. Overview

After this function is enabled, the device collects statistics on ARP messages received by each interface in a fixed period of time by replies, requests, and unknown packets.

2. Restrictions and Guidelines

The default statistics collection interval is 5 seconds. If there are many interfaces and the rate is not as expected, appropriately adjust the statistics collection interval. The minimum statistics collection interval can be set to 1 second.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable ARP packet rate statistics collection.

arp rate-statistic enable

The ARP packet rate statistics collection is disabled by default.

(4) (Optional) Configure the interval for ARP packet rate statistics collection.

arp rate-statistic compute interval *interval*

The default interval for collecting ARP packet rate statistics is 5 seconds.

1.6.8 Configuring the ARP Alarm Rate Limit

1. Overview

You can configure the ARP alarm rate limit to adjust the printing rate of ARP syslog alarms.

2. Restrictions and Guidelines

The actual ARP alarm rate may be lower than the configured rate, depending on system performance.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the ARP alarm rate limit.

arp warning-limit interval *interval* **times** *time*

The default ARP alarm rate limit interval is 50 seconds and the default upper limit of alarms allowed within this interval is 10.

1.7 Enabling Trusted ARP

1.7.1 Overview

Since trusted ARP entries come from authentic sources and will not be updated, they can efficiently prevent ARP spoofing targeting the gateway.

1.7.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable trusted ARP.

service trustedarp

Trusted ARP is disabled by default.

(4) (Optional) Enable VLAN translation when a trusted ARP entry is added.

arp trusted user-vlan *vlan-id* **translated-vlan** *vlan-id*

The VLAN translation is disabled when a trusted ARP entry is added by default.

Configure this command only when the VLAN delivered by the server differs from the valid VLAN in the trusted ARP entry.

(5) (Optional) Enable trusted ARP aging.

arp trusted aging

Trusted ARP entries are not aged by default.

After you configure this command, trusted ARP entries begin to age, and the aging time is the same as that of dynamic ARP entries. You can run the **arp timeout** command in interface configuration mode to set the aging time.

(6) (Optional) Adjust the capacity of trusted ARP entries.

arp trusted *number*

The maximum number of trusted ARP entries is half of the capacity of the ARP table by default.

Trusted ARP entries and other entries share the memory. If trusted ARP entries occupy much space, dynamic ARP entries may not have sufficient space. Set the capacity based on the actual requirement. Do not set it to an excessively large value.

1.8 Enabling Gratuitous ARP

1.8.1 Overview

Gratuitous ARP packets are special ARP messages. In a gratuitous ARP message, the source and destination IP addresses are the IP address of the local device. This function enables an interface to periodically send gratuitous ARP packets.

1.8.2 Restrictions and Guidelines

Generally, you need to enable this function on an interface that serves as the gateway of downlink devices so that the MAC address of the gateway of the downlink devices are updated periodically to prevent others from faking the gateway.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of learning gratuitous ARPs.

(4) The function of learning gratuitous ARPs is enabled by default. Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(5) Enable the function of sending gratuitous ARP requests.

arp gratuitous-send enable

The function of sending gratuitous ARP requests is enabled by default.

(6) Configure the interval for sending gratuitous ARP requests.

arp gratuitous-send interval *interval* [*number*]

The function of sending gratuitous ARP requests at intervals is disabled by default.

1.9 Enabling Proxy ARP

1.9.1 Overview

The device acts as a proxy to reply to ARP requests from other hosts. If a device receiving an ARP request finds the source IP address in a different network from the destination IP address and knows the route to the destination address, the device sends an ARP reply containing its own Ethernet MAC address.

1.9.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enables proxy ARP.

ip proxy-arp

Proxy ARP is disabled on an interface by default.

1.10 Configuring Local Proxy ARP

1.10.1 Overview

After local proxy ARP is enabled, the device can help hosts obtain the MAC addresses of other hosts in the same subnet.

1.10.2 Restrictions and Guidelines

This function can be configured only on SVIs and VXLAN routed interfaces.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the SVI configuration mode.

interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enables local proxy ARP.

local-proxy-arp

Local proxy ARP is disabled by default.

1.11 Configuring Any IP ARP

1.11.1 Overview

Users can access the network with any IP address. This applies when a user uses a laptop in a hotel and wants to access the Internet without changing the configured IP address and gateway.

1.11.2 Restrictions and Guidelines

This function is not applicable in the following two scenarios, in which a user must modify the configuration before accessing the Internet.

- The user's IP address is in the same network segment as the interface directly connected to the device. However, the configured gateway IP address is not the IP address configured for the interface directly connected to the device.
- The user's IP address is not in the same network segment as the interface directly connected to the device, but in the network segment of another interface. That means an IP address conflict occurs.

  As the user's IP address is not in the same network segment as the interface directly connected to the device, the dynamic ARP entry and direct route are generated only when the user initiates an ARP request. Therefore, in some scenarios (including but not limited to the following ones), the user will not be able to access the Internet unless the ARP entry is cleared and the gateway address is relearned on the user host.

  - The device acts as a proxy to respond to ARP requests. After the user host learns the MAC address of the device, the administrator deletes the dynamic ARP entry on the device. As a result, the user's dynamic ARP entry and direct route are removed and the user cannot receive the reply.
  - The device acts as a proxy to respond to ARP requests. After the user host learns the MAC address of the device, if any IP ARP is disabled and then enabled again on the interface, the user cannot receive the reply. The cause is that the user's dynamic ARP entry and direct route are immediately deleted when the any IP ARP function is disabled on the interface.

⚠ Caution

If static ARP entries or the ARP entries involving the Virtual Router Redundancy Protocol (VRRP) IP addresses exist, dynamic ARP entries generated by any IP ARP will be overwritten or fail to be added, and any IP ARP does not take effect.

1.11.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable any IP ARP.

arp any-ip [**dip-route**]

The any IP ARP function is disabled by default.

1.12 Enabling ARP Trust Detection

1.12.1 Overview

ARP trust detection is used to prevent excessive useless ARP entries generated due to ARP spoofing from occupying device resources. After ARP trust detection is enabled on an Layer 3 interface and the device receives an ARP request from this interface:

(2) If the corresponding entry does not exist, the device creates a dynamic ARP entry and performs neighbor unreachability detection (NUD) after 1 to 5 seconds. That is, the device ages the newly learned ARP entry and sends a unicast ARP request. If the device receives an ARP update packet from the peer within the aging time, it stores the entry. Otherwise, it deletes the entry.

(3) If the corresponding ARP entry exists, the device does not perform NUD.

(4) If the MAC address in the existing dynamic ARP entry is updated, the device performs NUD.

1.12.2 Restrictions and Guidelines

- Since this function adds a strict confirmation procedure in the ARP learning process, it affects the efficiency of ARP learning.
- After this function is disabled, the device no longer performs NUD for learning or updating ARP entries.

1.12.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable ARP trust detection.

arp trust-monitor enable

ARP trust detection is disabled by default.

1.13 Enabling ARP-based IP Guard

1.13.1 Overview

When receiving unresolved IP packets, the device cannot forward them through the hardware but sends them to the CPU for address resolution, that is, ARP learning. If a large number of such packets are sent to the CPU, the CPU will be congested, affecting other services on the device.

After ARP-based IP guard is enabled, the device will count the number of received ARP packets based on the destination IP address. When the number of packets with the same destination IP address exceeds a certain threshold, the device deems it as a CPU attack and will send a drop entry to the hardware so that the hardware will not send subsequent ARP packets with this destination IP address to the CPU. After the address resolution is complete, the device updates the entry to the forwarding state and continues to forward the packets with this destination IP address.

1.13.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable ARP-based IP guard.

arp anti-ip-attack *attack-num*

The default number of IP packets for triggering the discarding of ARP entries is 3.

1.14 Configuring ARP Packet Filtering

1.14.1 Overview

- ARP packet filtering supports the following functions:
  Filters out received gratuitous APR packets.
- Filters out ARP packets that hit ACL rules.
- Filters out ARP packets whose source MAC addresses is not consistent with the Ethernet source MAC address.
- Filters out ARP packets whose destination MAC addresses is not consistent with the Ethernet destination MAC address.

1.14.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure ARP packet filtering. Perform at least one of the following configuration steps:

- Enable gratuitous ARP filtering.

  arp filter gratuitous

  Gratuitous ARP filtering is disabled by default.
- Enable ARP-based ACL filtering.

  arp filter acl *acl-number*

  ARP-based ACL filtering is disabled by default.
- Enables the function of checking the source MAC addresses of ARP packets.

  arp filter smac-illegal

  The function of checking the source MAC addresses of ARP packets is disabled by default.
- Enable the function of checking the destination MAC addresses of ARP packets.

  arp filter dmac-illegal

  The function of checking the destination MAC addresses of ARP packets is disabled by default.

1.15 Restraining the Device from Sending ARP Requests to Authenticated VLANs

Note

This function is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.15.1 Overview

In gateway authentication mode, all sub VLANs in a super VLAN are authenticated VLANs by default. Users in an authenticated VLAN have to pass authentication to access the network. After authentication, a static ARP entry is generated on the device. Therefore, when accessing an authenticated user, the device does not need to send ARP requests to the authenticated VLAN. If the device attempts to access users in an authentication-exempt VLAN, it only needs to send ARP requests to the authentication-exempt VLAN.

1.15.2 Restrictions and Guidelines

- This function can be configured only on SVIs.
- In gateway authentication mode, the device does not send ARP requests to the authenticated VLANs by default. If the device needs to access authentication-exempt users in an authenticated VLAN, disable this function.

1.15.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the SVI configuration mode.

  interface vlan *interface-number*

(4) Configure the device not to send ARP requests to authenticated VLANs.

arp suppress-auth-vlan-req

ARP requests are not sent to authenticated VLANs by default.

1.16 Configuring ARP to Support VXLAN Functions

Note

This function is not supported by the S6980-64QC, S6920-4C, S6580-48CQ8QC, or S6580-24DC8QC.

1.16.1 Overview

In a network environment deployed with both the virtual extensible LAN (VXLAN) and software-defined networking (SDN), the ARP module provides the following functions to adapt to the SDN controller:

- Automatic ARP entry learning on OverlayRouter interfaces When the SDN controller is used to deliver ARP entries to the device in a unified manner, the function can be disabled. In this case, the device does not learn or update ARP entries based on the received ARP packets.
- Responding to ARP requests by the SDN controller When the SDN controller needs to manage the ARP control protocol in a unified manner, the function can be enabled. In this case, ARP requests are copied to the SDN controller, which will respond to the ARP requests.

- Restraining OverlayRouter interfaces from sending gratuitous ARP requests This function needs to be enabled to prevent an OverlayRouter interface serving as a gateway from actively sending gratuitous ARP requests when the gratuitous ARP requests will cause some terminals no longer to send ARP requests to request the MAC address of the gateway. If the terminals do not send ARP message after they learn the ARP entry of the gateway, the SDN controller will not receive ARP requests from the terminals.
- Actively sending gratuitous ARP request packets by OverlayRouter interfaces. The aging of the virtual MAC address of the OverlayRouter interface serving as a gateway can cause VXLAN-related traffic flooding and affect the device performance. In this case, this function needs to be enabled.
- Making remote neighbor entries synchronized by the Ethernet VPN (EVPN) function configured on OverlayRouter interfaces take effect. Remote neighbor entries synchronized by EVPN on OverlayRouter interfaces do not take effect by default, that is, remote entries are not used for forwarding L3 routing packets. After this function is configured, remote entries can be used for forwarding L3 routing packets.

1.16.2 Restrictions and Guidelines

For details about the VXLAN, see *Configuring VXLAN*.

1.16.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable automatic ARP entry learning on OverlayRouter interfaces.

unset arp vxlan learning-enable

The automatic ARP entry learning is enabled on all the OverlayRouter interfaces by default.

(4) Enable the SDN controller to respond to ARP requests on OverlayRouter interfaces.

arp vxlan sdn-reply

The SDN controller is not configured to respond to ARP requests on OverlayRouter interfaces by default.

(5) Restrain OverlayRouter interfaces from sending gratuitous ARP packets.

unset arp vxlan learning-enable

An OverlayRouter interface can send gratuitous ARP requests by default.

(6) This command is used to configure the interval for an OverlayRouter interface to actively send gratuitous ARP request packets.

arp vxlan gratuitous-send interval *interval*

The interval is not specified for an OverlayRouter interface to actively send gratuitous ARP request packets by default.

(7) Enter the VXLAN routing interface configuration mode.

interface overlayrouter *interface-number*

(8) Make remote neighbor entries synchronized by EVPN take effect.

unset arp vxlan remote-arp inactive

Remote neighbor entries synchronized by EVPN do not take effect by default.

1.17 Configuring Host Existence Judgment Prior to Proxy ARP Service

1.17.1 Overview

When two devices form a VRRP network and local proxy ARP is enabled on them, if the slave VRRP device sends an ARP request to a terminal, the master VRRP device will act as a proxy of the terminal and send an ARP reply to the slave VRRP device regardless of whether the terminal exists. As a result, the slave VRRP device learns a large number of proxy ARP entries.

After the host existence judgment prior to proxy ARP service is enabled, the master VRRP device, upon receiving an ARP request, first judges whether the ARP entry corresponding to the destination IP address exists. If yes, the master VRRP device acts as a proxy ARP. If no, the master VRRP device does not act as a proxy ARP. In addition, the gateway automatically broadcasts an ARP request for the ARP entry corresponding to the destination IP address. This prevents the case that the gateway fails to act as a proxy to respond to an ARP request of the destination ARP address if it does not have an ARP entry corresponding to the destination IP address.

After the host existence judgment prior to proxy ARP service is disabled, if the proxy conditions are met, the master ARP device directly acts as a proxy upon receiving an ARP request, without judging whether the ARP entry corresponding to the destination IP address has been resolved.

1.17.2 Restrictions and Guidelines

The **arp proxy-resolved** command is enabled on devices by default. That is, the master VRRP device responds to an ARP request as a proxy only after the destination IP address has been resolved.

1.17.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the master VRRP device to forcibly respond to ARP requests as a proxy.

unset arp proxy-resolved

By default, the master VRRP device judges the existence of the ARP entry corresponding to a destination IP address before the device responds to an ARP request as a proxy ARP.

1.18 Configuring ARP Entry Update During Active/Standby Switchover

1.18.1 Overview

This function can be enabled to quickly update ARP entries of a downlink device after VSU active/standby switchover, especially when the downlink device is similar to a server with dual network interface cards. When the slave device becomes the master, it will actively send ARP requests to SVIs (that are not in a super VLAN) of up to 1000 downlink terminals. After the terminals reply to these ARP requests, the device can update the ARP and MAC tables.

1.18.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of actively sending ARP requests to SVIs of downlink terminals during active/standby switchover.

arp switch-over resolve

ARP requests are not actively sent to downlink terminals during VSU active/standby switchover by default.

1.19 Sending ARP Requests to a Specific Sub VLAN

Note

This function is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.19.1 Overview

In a super VLAN scenario, the device actively broadcasts ARP resolution requests to the entire super VLAN by default. If there are many sub VLANs in the super VLAN, the ARP messages will be replicated in large quantities, which will affect device performance.

Most terminals (such as PCs and servers) request ARP information of the gateway before accessing the network. Therefore, there is no need to actively broadcast ARP resolution requests to the sub VLANs where these terminals reside. For dumb terminals (that do not actively send gratuitous ARP packets), this command can be deployed in a specified *vlan-list* to enable the device to actively send ARP resolution requests to these VLANs.

1.19.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable ARP to broadcast resolution requests to a specific sub VLAN in a super VLAN.

arp resolve vlan { *vlan-list* | **none** }

ARP broadcasts resolution requests to all sub VLANs in a super VLAN by default.

1.20 Configuring the Function of Forbidding Sending Passerby Packets to the CPU

1.20.1 Overview

When an interface receives a large number of passerby ARP packets not destined for the interface and sends them to the CPU, the CPU usage is very high. After the function of forbidding sending passerby ARP packets to the CPU is configured, an interface directly forwards the received passerby ARP packets and sends only broadcast ARP packets and ARP packets whose destination MAC addresses are L3 MAC addresses to the CPU. L3 MAC addresses include local MAC addresses, VRRP MAC addresses, global MAC address, and anycast MAC addresses.

1.20.2 Restrictions and Guidelines

- After this function is enabled, ARP packets whose destination MAC addresses are private MAC addresses of interfaces will not be sent to the CPU. This function needs to be disabled if ARP packets whose destination MAC addresses are private MAC addresses of L3 interfaces need to be sent to the CPU.
- This function is used to configure the allowlist of pass-through ARP packets allowed to be sent to the CPU.

1.20.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of forbidding sending passerby packets to the CPU.

arp-passby optimization enable

The function of forbidding sending passerby packets to the CPU is disabled by default.

1.21 Monitoring

Run the **display** command to check the configuration.

Run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** command to clear information.

⚠ Caution

Inappropriate use of the **cleanup** command may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
display arp-info trusted [*ipv4-address* [*mask*]]	Displays trusted ARP entries.
cleanup arp-cache-table trusted [*ipv4-address* [*mask*]]	Clears trusted ARP entries.
cleanup arp-cache-table [[**vrf** *vrf-name* | **oob**] [*ipv4-address* [*mask*]] | [**interface** *interface-type interface-number* | **vxlan** [**force**]]]	Clears dynamic ARP entries.
cleanup arp-cache-table packet count [*interface-type interface-number*]	Clears ARP packet statistics.
display arp-info [[**vrf** *vrf-name*] [*ipv4-address* [*mask*] | *interface-type interface-number* | *mac-address* | **complete** | **incomplete** | **static** | **trusted** [*ipv4-address* [*mask*]] | **vxlan**]]	Displays the ARP table.
display arp-info detail [*interface-type interface-number* | [**oob**] [*ipv4-address* [*mask*] | *mac-address* | **complete** | **incomplete** | **static** | **trusted** [*ipv4-address* [*mask*]] | **vxlan**] | **subvlan** { **min-max** *min-value max-value* | *subvlan-number* }]	Displays the detailed ARP table.
display arp-info oob [*ipv4-address* [*mask*] | **static** | **complete** | **incomplete** | *mac-address*]	Displays the ARP table on the management interface.
display ip arp [**vrf** *vrf-name*]	Displays the ARP table.
display arp-info [**detail**] **trusted** [*ipv4-address* [mask]]	Displays the trusted ARP table.
display arp-info counter	Displays the count of ARP entries.
display arp-info timeout	Displays the aging time of dynamic ARP entries.
display arp-info packet statistics [*interface-type interface-number*]	Displays ARP packet statistics.
display arp-info rate-statistic [*interface-type interface-number*]	Displays ARP packet rate statistics.
display arp-info timeout	Displays the aging time of dynamic ARP entries.

display arp-info flapping record [*ipv4-address*]	Displays ARP flapping records.
display arp-info send-disable table [**ip** *ipv4-address*]	Displays the ARP suppression table.
display arp-info anti-attack count	Displays the statistics on illegal ARP packets.
display arp-info dynamic-entry-limit	Displays the statistics on the ARP capacity limit of a line card.
debugging arp	Displays the statistics on ARP packets sent and received.
debugging arp event	Displays the creation and deletion status of ARP entries.

1.22 Configuration Examples

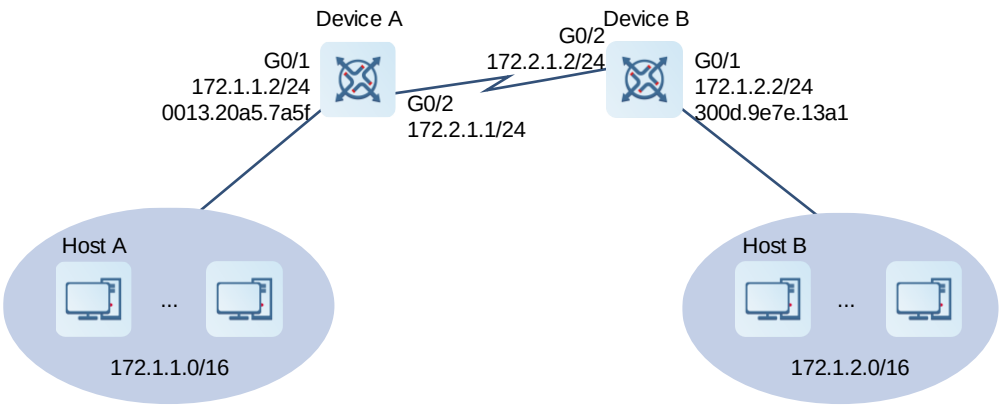
1.22.1 Configuring Proxy ARP

1. Requirements

As shown in Figure 1-4, the IP addresses of host A and host B are on the same network segment, but are isolated in two LANs by routers. Neither of the two hosts is configured with a default gateway, and the two routers are connected over a serial link. Enable proxy ARP on the routers to achieve direct communication between hosts in different LANs.

2. Topology

Figure 1-4 Configuring Proxy ARP



3. Notes

- Enable proxy ARP on interfaces of device A and device B that are connected to hosts.
- Configure default routes on device A and device B.

4. Procedure

(1) Configure device A.

Configure an IP address for a Layer 3 Ethernet port GigabitEthernet 0/1.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 172.1.1.2 24
```

Enable proxy ARP on a Layer 3 Ethernet port GigabitEthernet 0/1.

```
DeviceA(cfg-if-GigabitEthernet 0/1)# ip proxy-arp
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Configure an IP address for a Layer 3 Ethernet port GigabitEthernet 0/2.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# ip address 172.2.1.1 24
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

Configure the static route to 172.1.2.0/16.

```
DeviceA(cfg)# ip route-config 172.1.2.0 255.255.0.0 gigabitethernet 0/2
```

(2) Configure device B.

Configure an IP address for a Layer 3 Ethernet port GigabitEthernet 0/1.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 172.1.2.2 24
```

Enable proxy ARP on a Layer 3 Ethernet port GigabitEthernet 0/1.

```
DeviceB(cfg-if-GigabitEthernet 0/1)# ip proxy-arp
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Configure an IP address for a Layer 3 Ethernet port GigabitEthernet 0/2.

```
DeviceB(cfg)# interface gigabitethernet 0/2
DeviceB(cfg-if-GigabitEthernet 0/2)# ip address 172.2.1.2 24
DeviceB(cfg-if-GigabitEthernet 0/2)# exit
```

Configure the static route to 172.1.1.0/16.

```
DeviceB(cfg)# ip route-config 172.1.1.0 255.255.0.0 gigabitethernet 0/2
```

(3) Configure hosts.

Set the IP address of host A to 172.1.1.10/16.

Set the IP address of host B to 172.1.2.10/16.

5. Verification

Host A can ping host B successfully.

You can check the ARP table of host A and find that the MAC address corresponding to host B is the MAC address of a Layer 3 Ethernet port GigabitEthernet 0/1 on device A.

6. Configuration Files

- Device A configuration file

```
!
```

```
hostname DeviceA
ip route-config 172.1.2.0 255.255.0.0 gigabitethernet 0/2
!
 interface gigabitethernet 0/1
 ip address 172.1.1.2 24
 ip proxy-arp
!
 interface gigabitethernet 0/2
 ip address 172.2.1.1 24
!
```

- Device B configuration file

```
!
hostname DeviceB
ip route-config 172.1.1.0 255.255.0.0 gigabitethernet 0/2
!
 interface gigabitethernet 0/1
 ip address 172.1.2.2 24
 ip proxy-arp
!
 interface gigabitethernet 0/2
 ip address 172.2.1.2 24
!
```

Contents

1 Configuring IPv4 Basics 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Concepts 1
1.1.3 Principles 4
1.1.4 Protocols and Standards 7
1.2 Configuration Task Summary 7
1.3 Notes 7
1.4 Configuring an IP Address for an Interface 8
1.4.1 Overview 8
1.4.2 Configuration Tasks 8
1.4.3 Configuring a Static IP Address for an Interface 8
1.4.4 Configuring an IP Address Combination for an Interface 9
1.4.5 Configuring an IP Broadcast Address 10
1.4.6 Configuring a Backup IP Address 10
1.5 Configuring the Default Gateway for a Management Interface 11
1.5.1 Overview 11
1.5.2 Restrictions and Guidelines 11
1.5.3 Procedure 11
1.6 Configuring Directed Broadcast Packets 11
1.6.1 Overview 11
1.6.2 Restrictions and Guidelines 12

1.6.3 Procedure....12
1.7 Configuring ICMP Messages....12
1.7.1 Overview....12
1.7.2 Configuration Tasks....12
1.7.3 Enabling the Function of Sending TTL Timeout Messages....13
1.7.4 Enabling the Timestamp Query....13
1.7.5 Enabling the Function of Sending ICMP Destination Unreachable Messages....13
1.7.6 Enabling the Function of Sending ICMP Redirection Messages....14
1.7.7 Enabling the Function of Sending ICMP Mask Reply Messages....14
1.8 Configuring the Transmission Rate of ICMP Error Messages....15
1.8.1 Overview....15
1.8.2 Restrictions and Guidelines....15
1.8.3 Procedure....15
1.9 Setting the IP MTU Value....15
1.9.1 Overview....15
1.9.2 Restrictions and Guidelines....15
1.9.3 Procedure....16
1.10 Setting the IP TTL Value....16
1.10.1 Overview....16
1.10.2 Procedure....16
1.11 Configuring IP Routed Port Protection....17
1.11.1 Overview....17
1.11.2 Restrictions and Guidelines....17
1.11.3 Procedure....17

1.12 Configuring the IP Source Route Option 17
1.12.1 Overview 17
1.12.2 Restrictions and Guidelines 17
1.12.3 Procedure 17
1.13 Monitoring 18
1.14 Configuration Examples 18
1.14.1 Configuring IPv4 Addresses for Network Communication 18

1 Configuring IPv4 Basics

1.1 Introduction

1.1.1 Overview

Internet Protocol (IP) is one of the most core protocols in the Transmission Control Protocol (TCP)/IP protocol suite and works at the network layer. Each IP device in the network needs a logical virtual address, which is used by the IP protocol to realize communication between devices. This address is usually called IP address.

The Internet Protocol version 4 (IPv4) is the fourth revision of the Internet Protocol, and is also the first version that has been widely used. The Internet Protocol version 6 (IPv6), the successor of IPv4, is still in its early stage of deployment in 2011 when IPv4 addresses were about to be used up. IPv4 uses the 32-bit address format, so there are only 4,294,967,296 addresses in the address space, some of which are reserved for special purposes.

Note

The following description is for IPv4 only, and IP addresses mentioned in this document refer to IPv4 addresses.

1.1.2 Basic Concepts

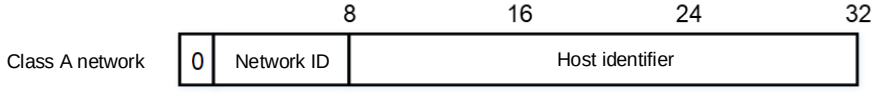
1. Classification of IP Addresses

IP addresses are used for interconnection at the IP layer. An IP address consists of 32-bit binary digits. To facilitate writing and description, an IP address is generally expressed in decimal notation. When expressed in decimal notation, an IP address is divided into four groups, with eight bits in each group. The value range of each group is from 0 to 255, and groups are separated by dots (.), which are called dotted decimal notation. For example, "192.168.1.1" is an IP address expressed in decimal notation.

A 32-bit IP address consists of two parts: network ID and local address. A network ID indicates a network identifier, and a local address identifies a host. Based on the values of the first several bits in the network part, IP addresses in use can be classified into four classes.

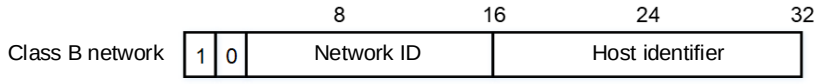
- For a class A address, the most significant bit is "0", the following 7 bits indicate a network ID, and the last 24 bits indicate a local address (host identifier). There are 2^7 = 128 class A networks in total.

Figure 1-1 Class A Network



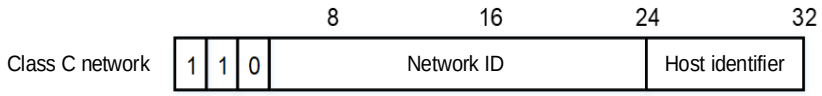
- For a class B address, the two most significant bits are "10", the following 14 bits indicate a network ID (network identifier), and the last 16 bits indicate a local address (host identifier). There are 2^{14}=16,384 class B networks in total.

Figure 1-2 Class B Network



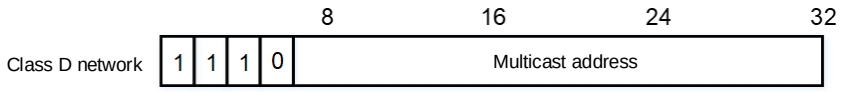
- For a class C address, the three most significant bits are "110", and the following 21 bits indicate a network ID, and the last 8 bits indicate a local address. There are 2^{21}=2,097,152 class C networks in total.

Figure 1-3 Class C Network



- For a class D address, the four most significant bits are "1110" and other bits are designated for multicast addresses.

Figure 1-4 Class D Network



Note

The addresses with the four most significant bits of "1111" cannot be assigned. These addresses are called class E addresses and are reserved for experimental use.

IP addresses must be assigned according to the nature of the network during IP address planning in the process of network construction. If the network needs to be connected to the Internet, users need to apply to relevant organizations for IP addresses. IP addresses can be applied from China Internet Network Information Center (CNNIC) in China, and the Internet Corporation for Assigned Names and Numbers (ICANN) is responsible for the final IP address assignment. If the network is built to be an internal private network, users do not need to apply for IP addresses and can be assigned dedicated private network addresses.

Table 1-1 Reserved and Available Addresses

Class	Address Range	Status
Class A network	0.0.0.0–0.255.255.255	Reserved
	1.0.0.0–126.255.255.255	Available
	127.0.0.0–127.255.255.255	Reserved
Class B network	128.0.0.0–191.254.255.255	Available
	191.255.0.0–191.255.255.255	Reserved

Class	Address Range	Status
Class C network	192.0.0.0–192.0.0.255	Reserved
	192.0.1.0–223.255.254.255	Available
	223.255.255.0–223.255.255.255	Reserved
Class D network	224.0.0.0–239.255.255.255	Multicast addresses
Class E network	240.0.0.0–255.255.255.254	Reserved
	255.255.255.255	Broadcast addresses

2. Private Network IP Addresses

Private network IP addresses are not used in the Internet. If the devices to which private addresses are assigned need to be connected to the Internet, these IP addresses need to be converted into valid Internet addresses. Table 1-2 lists private network address ranges. Private network addresses are defined in RFC 1918 and include three address blocks.

Table 1-2 Private Network Addresses

Class	Address Range	Status
Class A network	10.0.0.0–10.255.255.255	One class A network
Class B network	172.16.0.0–172.31.255.255	16 class B networks
Class C network	192.168.0.0–192.168.255.255	256 class C networks

For assignment of IP addresses, TCP/User Datagram Protocol (UDP) ports, and other codes, refer to RFC 1166.

3. Special IP Addresses

The IP addresses listed below are used for special purposes and cannot be used as host IP addresses.

- Addresses with the network ID of all zeros: Indicates the host in the network. For example, 0.0.0.16 indicates the host numbered 16 in the network.
- Addresses with the host ID of all zeros: Indicates a network address, used for identifying a network, for example, 192.168.1.0.
- Addresses with the host ID of all ones: Indicates a broadcast address of the network. For example, the packets with the destination address of 192.168.1.255 will be forwarded to all hosts in the 192.168.1.0 network.

4. Network Mask

A network mask is also a 32-bit value and identifies the bits occupied by the network part of an IP address. In a network mask, the IP address bits corresponding to the bits whose values are ones are the network part, and the IP address bits corresponding to the bits whose values are zeros are the host address. For example, for class A networks, the network mask is 255.0.0.0. Network masks can be used to divide a network into several

subnets. Some bits of the host address are used as the network ID to decrease the host capacity and increase the number of networks. In this case, network masks are called subnet masks.

5. MTU

Maximum transmission unit (MTU) is the maximum size, in bytes, of the data transmitted by the data link layer to its upper layer, such as IP. The MTU is protocol-specific, for example:

- The MTU for IEEE802.3/802.2 is 1,492 bytes.
- The MTU range of an Ethernet port is 64 to 1,518 bytes. Ethernet frames beyond this range may be dropped.

The IP protocol decides whether to fragment data based on the Layer 2 MTU value. When two hosts communicate through multiple networks, MTUs of different sizes may exist in every link. The smallest MTU determines the transmission rate.

6. TTL

Time to live (TTL) refers to the number of network segments, through which packets are allowed to pass before the packets are discarded. The TTL value is used to judge whether packets stay on the network for a long time and should be discarded.

When two hosts communicate through multiple networks, a TTL value is set at the source end and decreases by 1 each time the IP packet passes through a router. When the TTL value drops to zero, the router deems the destination unreachable and discards the packet. TTL prevents infinite transmission of useless packets and bandwidth waste.

1.1.3 Principles

1. Obtaining IP Addresses

A device must have an IP address before sending and receiving IP datagrams. Only the interface configured with an IP address can run the IP protocol.

An interface can obtain IP addresses in the following four ways. These ways are mutually exclusive. If you configure a new way to obtain an IP address, the old IP address will be overwritten.

(1) Manually configuring IP addresses

(2) Obtaining IP addresses through the Dynamic Host Configuration Protocol (DHCP): For details about the DHCP protocol, see *Configuring DHCP.*

2. Configuring Multiple IP Addresses for an Interface

An interface on a device can be configured with multiple IP addresses, of which one is the primary IP address and the others are secondary IP addresses or slave IP addresses. Theoretically, the number of secondary IP addresses is not limited. However, secondary IP addresses must belong to different networks and secondary IP addresses must be in different networks from primary IP addresses. In network construction, secondary IP addresses are often used in the following circumstances:

- A network does not have enough host addresses. For example, when the number of hosts exceeds 254 in a local area network (LAN), one class C network is not enough and another class C network is needed. In this case, two networks need to be connected. Therefore, more IP addresses are needed.
- Many old networks are based on Layer 2 bridged networks without subnetting. You can use secondary IP addresses to upgrade the network to a routing network based on IP layer. For each subnet, one device is

configured with one IP address.

- When two subnets of one network are isolated by another network, in consideration that one subnet cannot be configured on two or more interfaces of a device, you can connect the isolated subnets by creating a subnet on the isolated network and configuring a secondary address.

3. Broadcast Packet Processing

Broadcast packets refer to the packets destined for all hosts on a physical network. The device supports two types of broadcast packets. One type is limited broadcast packets, with the 32-bit destination address of all ones, that is, 255.255.255.255. The other type is directed broadcast packets, with the destination address of all hosts in a specified network, that is, host ID bits of all ones, for example, 192.168.1.255/24. Limited broadcast packets are blocked by the router, whereas directed broadcast packets can be forwarded.

If IP network devices forward limited broadcast packets, the network may be overloaded, which severely affects network performance. This circumstance is called a broadcast storm. Devices provide multiple approaches to confine broadcast storms within the local network and prevent continuous spread of broadcast storms. Layer 2 network devices such as bridges and switches still forward and spread broadcast storms.

The best way to avoid broadcast storms is to assign a broadcast address to each network, which is directed broadcast. This requires the IP protocol to use directed broadcast packets rather than limited broadcast packets to spread data. When receiving IP directed broadcast packets, the devices that is not directly connected to the destination subnet will forward them in the same way as that for unicast packets. After directed broadcast packets reach the device that is directly connected to the destination subnet, the device converts the directed broadcast mode into limited broadcast mode and broadcasts the packets to all hosts on the destination subnet at the link layer.

For details about broadcast, see RFC 919 and RFC 922.

4. ICMP Packet

Internet Control Message Protocol (ICMP) is a sub protocol of the TCP/IP protocol suite and is used to transfer control messages between IP hosts and network devices to notify certain devices of packet transmission exceptions. The following describes common ICMP packets.

- Echo request and echo reply

  Common ping operations use echo requests and echo replies. A device sends an echo request to another node. If there is no exception on the way, the destination host returns an echo reply, indicating that the host is reachable.

- Protocol unreachable message

  When a device receives a non-broadcast packet destined for itself and the packet uses the IP protocol that cannot be processed by the device, the device sends an ICMP protocol unreachable message to the source host.

- Host unreachable message

  When a device receives a non-broadcast packet destined for itself but the device does not know a route to forward the packet, it sends an ICMP host unreachable message to the source address.

- Redirection message

  Sometimes a network device sends packets from the interface that receives the packets, indicating that a route may be less than optimal. If this condition occurs, the device will send an ICMP redirection message

to the data source, informing the source that the gateway reachable to the destination address is another device on the same subnet. In this way, the data source sends subsequent packets according to the optimal path.

- Mask reply message

  Sometimes, a network device sends an ICMP mask request message to obtain the mask of a subnet. The network device that receives the ICMP mask request message sends a mask reply message.

- TTL timeout error message

  When a device forwards an IP packet, if the TTL expires, the device responds to the source end with a TTL timeout error message.

  To prevent attacks from other devices after the device is located through traceroute, you can disable the function of sending TTL timeout errors on the device. After this function is disabled, the device will no longer make a response when receiving a TTL timeout message.

- Timestamp query

  RFC 792 requires the system to return its current time after receiving an ICMP timestamp query request.

  To prevent attackers from obtaining the system time through this protocol and attacking other time-based protocols, you can disable the timestamp query function. Then, the device directly discards received ICMP timestamp query requests.

5. Limiting Transmission Rate of ICMP Error Packets

This function limits the transmission rate of ICMP error packets by using the token bucket algorithm to prevent denial of service (DoS) attacks.

If an IP packet needs to be fragmented but the Don't Fragment (DF) bit in the header is set to 1, the device returns an ICMP destination unreachable message to the source host. This ICMP error message is used to discover the path MTU. When there are too many other ICMP error messages, the ICMP destination unreachable message may not be sent. As a result, the path MTU discovery function fails. To avoid this problem, you should limit the transmission rate of ICMP destination unreachable messages and other ICMP error messages respectively.

6. IP MTU

The IP protocol at the network layer checks the size of each packet from the upper layer protocol and determines whether it should be "fragmented" based on the MTU size of the local device.

If an IP packet exceeds the IP MTU size, the device fragments the packet. For all devices on the same physical network segment, the IP MTU of interconnected interfaces must be the same. After the link MTU of interfaces is changed, the IP MTU of interfaces will be changed accordingly. The IP MTU of interfaces automatically keeps consistent with the link MTU of interfaces. However, if the IP MTU of interfaces is adjusted, the link MTU of interfaces will not be changed.

7. IP Source Routing

When a device receives an IP packet, it checks the options such as the strict source route, loose source route, and record route in the IP packet header. These options are detailed in RFC 791. If the device detects that the packet enables one option, it performs an action accordingly. If the device detects an invalid option, it returns an ICMP parameter error message to the source and then discards the packet.

After the function of processing IP source routing information is enabled on a device, the source route option is added to an IP packet to test the throughput of a specific network or help the packet bypass the failed network. However, this may cause network attacks such as source address spoofing and IP spoofing.

1.1.4 Protocols and Standards

- RFC 1918: Address Allocation for Private Internet
- RFC 1166: Internet Numbers
- RFC 919: Broadcasting Internet Datagrams
- RFC 922: Broadcasting Internet Datagrams Subnets
- RFC 791: Internet Protocol
- RFC 792: Internet Control Message Protocol

1.2 Configuration Task Summary

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring an IP Address for an Interface
- Configuring the Default Gateway for a Management Interface
- Configuring Directed Broadcast Packets
- Configuring ICMP Messages
- Configuring the Transmission Rate of ICMP Error Messages
- Setting the IP MTU Value
- Setting the IP TTL Value
- Configuring IP Routed Port Protection
- Configuring the IP Source Route Option

1.3 Notes

This section describes the precautions for configuring IPv4 basic functions.

Whether IPv4 basic functions are supported depends on devices.

Table 1-1 Product Supporting IPv4 Basic Functions

Function	Product Supporting the Function
Configuring IP Routed Port Protection Function	S6980-64QC S6920-4C S6580-48CQ8QC S6580-24DC8QC S6520-64CQ S6510-32CQ S6510-4C S6510-48VS8CQ S6250-48XS8CQ

1.4 Configuring an IP Address for an Interface

1.4.1 Overview

This section describes how to configure an IP address for an interface for communication over the IP network. IP addresses can be manually configured or obtained through DHCP. In general, an interface only needs to be configured with one IP address to achieve communication with other hosts. Configuring a backup IP address for a management interface can prevent an IP address conflict occurring on the management interface of a virtual switching unit (VSU) in dual-active mode.

1.4.2 Configuration Tasks

(1) Configure an IP address for an interface. The configuration steps below are mutually exclusive. Please configure only one task.

- Configuring a Static IP Address for an Interface
- Configuring an IP Address Combination for an Interface

(2) (Optional) Configuring an IP Broadcast Address

(3) (Optional) Configuring a Backup IP Address

1.4.3 Configuring a Static IP Address for an Interface

1. Restrictions and Guidelines

Manual configuration of IP addresses and dynamic acquisition of IP addresses are mutually exclusive. If you configure a new approach to obtain an IP address, the old IP address will be overwritten.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the switch virtual interface (SVI) configuration mode.

  interface vlan *interface-number*

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(4) Configure an IP address for an interface.

ip address *ipv4-address* { *mask* | *mask-length* } [[**dual-active backup**] | [**secondary**]]

No IP address is configured for an interface by default.

1.4.4 Configuring an IP Address Combination for an Interface

1. Restrictions and Guidelines

The IP address combination configuration command can be configured only on switch virtual interfaces (SVIs) and management (MGMT) interfaces.

The IP address combination configuration command and the static IP address configuration command as well as DHCP are mutually exclusive, but the IP address combination configuration command can be used to configure both static IP addresses and DHCP.

- When the IP address combination configuration command is used to configure a static IP address, if an IP address on the same network segment is already configured, the configuration fails.
- When the IP address combination configuration command is used to configure both a static IP address and a dynamic IP address, if the IP address obtained through DHCP does not conflict with the network segment of the static IP address, the IP address obtained through DHCP is the primary IP address and the static IP address is a secondary IP address. If the IP address obtained through DHCP conflicts with the network segment of the static IP address, an IP address will be obtained again, during which the static IP address is a primary IP address.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the switch virtual interface (SVI) configuration mode.

  interface vlan *interface-number*

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(4) Configure an IP address combination for the interface.

ip address mix { *ipv4-address* { *mask* | *mask-length* } | **dhcp** }

No IP address combination is configured for an interface by default.

1.4.5 Configuring an IP Broadcast Address

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Configure an IP broadcast address.

ip broadcast *ipv4-address*

The default IP broadcast address of an interface is 255.255.255.255.

1.4.6 Configuring a Backup IP Address

1. Restrictions and Guidelines

- The backup IP address configuration command can be configured only on management interfaces.
- A backup IP address configured for a management interface of a VSU can prevent an IP address conflict on the management interfaces of the VSU in dual-active mode. When a VSU runs properly, the backup IP address does not take effect. When the VSU is split, the slave device becomes the master device. If the original master device is still running, the system enters the dual-active mode. In this case, the new master device uses the configured backup IP address as the IP address of the management interface, and the IP address of the management interface on the original master device remains unchanged. You can perform the configuration according to O&M requirements.
- If different management IP addresses are configured for all members of a VSU and a backup IP address is configured for the VSU, after the VSU is split and the system enters the dual-active mode, the backup IP

address overwrites the IP address of the management interface on the new master device and takes effect, resulting in IP address waste. Therefore, you are advised to run the **mgmt_mode unique** command to set the management interface creation mode of a VSU to **unique** (that is, only one global management interface is created) and then configure a backup IP address.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the management interface configuration mode.

interface mgmt *interface-number*

(4) Configure a backup IP address.

ip address *ipv4-address* { *mask* | *mask-length* } **dual-active backup**

1.5 Configuring the Default Gateway for a Management Interface

1.5.1 Overview

This section describes how to configure the default gateway for a management interface. After configuration, a default route is generated, with the outbound interface of the management interface and the next hop of the configured gateway.

1.5.2 Restrictions and Guidelines

This command can be configured on a management interface only.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the management interface configuration mode.

interface mgmt *interface-number*

(4) Configure the default gateway for the management interface.

gateway *ipv4-address*

No default gateway is configured for a management interface by default.

1.6 Configuring Directed Broadcast Packets

1.6.1 Overview

You can configure a specified interface to forward directed broadcast packets. Then, the interface can forward directed broadcast packets destined for the directly connected network.

1.6.2 Restrictions and Guidelines

- This function applies to the transmission of directed broadcast packets within the destination subnet and will not affect the forwarding of other directed broadcast packets.
- On an interface, you can define an access control list (ACL) to control the forwarding of certain directed broadcast packets. After an ACL is defined, only directed broadcast packets that match the ACL are forwarded.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Configure the interface to allow forwarding directed broadcast packets.

ip enable-directed-broadcast [*acl-number*]

The function of translating the IP directed broadcast mode to the physical broadcast mode is disabled by default.

1.7 Configuring ICMP Messages

1.7.1 Overview

The ICMP protocol is used to transfer control messages between IP hosts and network devices to notify network exceptions. You can enable the functions of sending TTL timeout messages, ICMP destination unreachable messages, ICMP redirection messages, and ICMP mask reply messages, and timestamp query function.

1.7.2 Configuration Tasks

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Enabling the Function of Sending TTL Timeout Messages
- Enabling the Timestamp Query
- Enabling the Function of Sending ICMP Destination Unreachable Messages

- Enabling the Function of Sending ICMP Redirection Messages
- Enabling the Function of Sending ICMP Mask Reply Messages

1.7.3 Enabling the Function of Sending TTL Timeout Messages

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of sending TTL timeout messages.

ip ttl-expires enable

The function of sending TTL timeout messages is enabled by default.

1.7.4 Enabling the Timestamp Query

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the timestamp query function.

ip icmp timestamp

The timestamp query function is enabled by default.

1.7.5 Enabling the Function of Sending ICMP Destination Unreachable Messages

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Enable the function of sending ICMP destination unreachable messages.

ip unreachables

The function of sending ICMP destination unreachable messages is enabled by default.

1.7.6 Enabling the Function of Sending ICMP Redirection Messages

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Enable the function of sending ICMP redirection messages.

ip redirects

The function of sending ICMP redirection messages is enabled by default.

1.7.7 Enabling the Function of Sending ICMP Mask Reply Messages

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Enable the function of sending ICMP mask reply messages.

ip icmp-mask-reply

The function of sending ICMP mask reply messages is disabled by default.

1.8 Configuring the Transmission Rate of ICMP Error Messages

1.8.1 Overview

This function limits the transmission rate of ICMP error packets by using the token bucket algorithm to prevent denial of service (DoS) attacks.

If an IP packet needs to be fragmented but the Don't Fragment (DF) bit in the header is set to 1, the device sends an ICMP destination unreachable message to the source host. This ICMP error message is used to discover the path MTU. If there are too many other ICMP error messages, the ICMP destination unreachable packet may not be sent. As a result, the path MTU discovery function fails. To avoid this problem, you should limit the transmission rate of ICMP destination unreachable messages and other ICMP error messages respectively.

1.8.2 Restrictions and Guidelines

Since the precision of the timer is 10 milliseconds, you are advised to set the refresh cycle of a token bucket to an integer multiple of 10 milliseconds. If the refresh cycle is set to a value greater than 0 and smaller than 10 milliseconds, the refresh cycle that actually takes effect is 10 milliseconds. For example, if the transmission rate is set to 1 packet per 5 milliseconds, two ICMP error packets are actually sent per 10 milliseconds. If the refresh cycle is not an integral multiple of 10 milliseconds, the refresh cycle that actually takes effect is automatically converted into an integral multiple of 10 milliseconds. For example, if the transmission rate is set to 3 packets per 15 milliseconds, two ICMP error packets are actually sent per 10 milliseconds.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the transmission rate of ICMP error packets.

ip icmp error-rate-limit [**df**] *interval* [*bucket-size*]

Ten ICMP error packets are transmitted within 100 milliseconds by default.

1.9 Setting the IP MTU Value

1.9.1 Overview

This section describes how to set the MTU of an IP packet. If the size of an IP packet exceeds the IP MTU size, the packet will be fragmented.

1.9.2 Restrictions and Guidelines

For all devices on the same physical network segment, the IP MTU of interconnected interfaces must be the same. The IP MTU of interfaces automatically keeps consistent with the link MTU of interfaces. If the link MTU of interfaces is changed, the IP MTU of interfaces will be changed accordingly. However, if the IP MTU of interfaces is adjusted, the link MTU of interfaces will not be changed.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Set the IP MTU.

ip mtu *mtu*

The default MTU of an IP packet is 1500 bytes.

1.10 Setting the IP TTL Value

1.10.1 Overview

When an IP packet is transmitted from the source address to the destination address through routers, a TTL value is set at the source end and decreases by 1 every time that the IP packet passes through a router. When the TTL value drops to zero, the router discards the packet. TTL prevents infinite transmission of useless packets and waste of bandwidth.

1.10.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Set the IP TTL.

ip ttl

The default TTL value of unicast packets sent by the device is 64.

1.11 Configuring IP Routed Port Protection

1.11.1 Overview

This section describes how to enable routed port protection to prevent packets from being transmitted through the same interface that receives the packets.

1.11.2 Restrictions and Guidelines

This command can only be used for Layer 3 routed ports.

1.11.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure the IP routed port protection.

ip redirect-drop

The IP routed port protection function is disabled by default.

1.12 Configuring the IP Source Route Option

1.12.1 Overview

The IP source route option is used to test the throughput of a specific network and help the packet bypass a failed network.

1.12.2 Restrictions and Guidelines

Enabling the IP source route option may cause network attacks such as source address spoofing and IP spoofing.

1.12.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IP source route option.

ip source-routing

The IP source route option is enabled by default.

1.13 Monitoring

This section describes the display commands used for checking the running status of a configured function to verify the configuration effect.

Table 1-3 Monitoring

Command	Purpose
display ip interface-status [*interface-type interface-number* | **brief** [**oob**]]	Displays the IP address of an interface.
display ip packet statistics [*interface-type interface-number* | **total**]	Displays IP packet statistics.
display ip packet queue	Displays the statistics on sent and received IP packets in the protocol stack.
display ip raw-socket [*protocol-number*]	Displays all IPv4 raw sockets.
display ip sockets	Displays all IPv4 sockets.
display ip udp [**local-port** *port-number* | **peer-port** *port-number*]	Displays all IPv4 UDP sockets.
display ip udp statistics	Displays the statistics on all IPv4 UDP sockets.

1.14 Configuration Examples

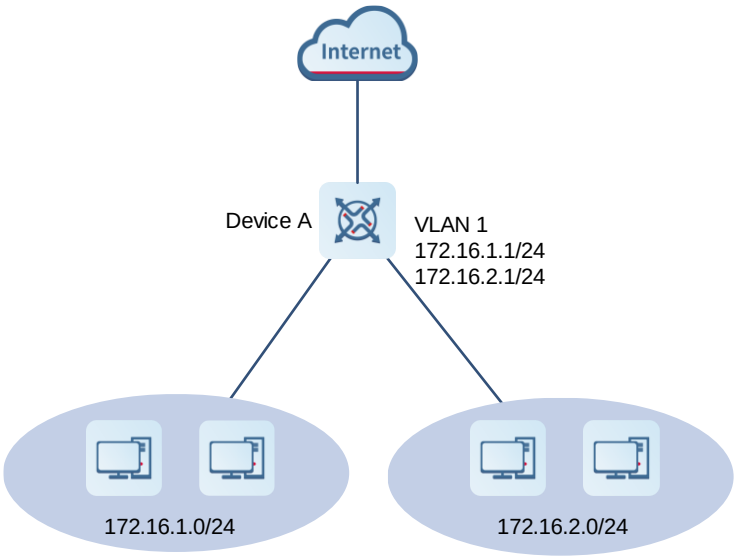
1.14.1 Configuring IPv4 Addresses for Network Communication

1. Requirements

As shown in Figure 1-5, device A is connected to a LAN (belonging to VLAN 1), which involves two network segments: 172.16.1.0/24 and 172.16.2.0/24. Computers on the two network segments are required to access the Internet through device A and hosts on the two network segments can also communicate with each other.

2. Topology

Figure 1-5 Configuring IPv4 Addresses for Network Communication



3. Notes

- Configure two IP addresses on the interface VLAN 1, of which one is a primary IP address and the other is a slave IP address.
- Configure the device as the gateway on the hosts on these two network segments.

4. Procedure

(1) Configure device A.

Configure a primary IP address and a slave IP address for the interface VLAN 1.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface vlan 1
DeviceA(cfg-if-VLAN 1)# ip address 172.16.1.1 255.255.255.0
DeviceA(cfg-if-VLAN 1)# ip address 172.16.2.1 255.255.255.0 secondary
```

(2) Configure the hosts on the 172.16.1.0/24 network segment.

Set the gateway to 172.16.1.1/24 for the hosts.

(3) Configure the hosts on the 172.16.2.0/24 network segment.

Set the gateway to 172.16.2.1/24 on the hosts.

5. Verification

Run the ping command on device A to check the connectivity with the hosts in the 172.16.1.0/24 network segment.

```
DeviceA# ping 172.16.1.2
Sending 5, 100-byte ICMP Echoes to 172.16.1.2, timeout is 2 seconds:
  < press Ctrl+C to break >
```

```
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the ping command on device A to check the connectivity with the hosts in the 172.16.2.0/24 network segment.

```
DeviceA# ping 172.16.2.1
Sending 5, 100-byte ICMP Echoes to 172.16.2.1, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the ping command to check the connectivity between the hosts in the 172.16.1.0/24 network segment and the hosts in the 172.16.2.0/24 network segments.

6. Configuration Files

Device A

```
hostname DeviceA
!
interface vlan 1
 ip address 172.16.1.1 255.255.255.0
 ip address 172.16.2.1 255.255.255.0 secondary
!
```

Contents

1 Configuring DHCP 1

1.1 Introduction 1

1.2 Principles 2

1.2.1 Basic Concepts 2

1.2.2 Message Format 3

1.2.3 DHCP Lease Process 6

1.2.4 DHCP Server 7

1.2.5 DHCP Relay Agent 8

1.2.6 DHCP Associated with VRRP Monitoring 12

1.2.7 Protocols and Standards 12

1.3 Configuring DHCP Server 12

1.3.1 Overview 12

1.3.2 Configuration Task Summary 13

1.3.3 Notes 13

1.3.4 Enabling DHCP Server 13

1.3.5 Configuring Static IP Address Assignment 14

1.3.6 Configuring Dynamic Assignment of IP Addresses and Network Parameters 15

1.3.7 Configuring DHCP Address Pool Management Function 16

1.3.8 Providing the Boot Image File Name 17

1.3.9 Assigning IP Addresses Based on Class Rules 18

1.3.10 Assigning IP Addresses Based on AM Rules 19

1.3.11 Refreshing Trusted ARP Entries 20

1.3.12 Configuring Rogue DHCP Server Detection....21
1.3.13 Configuring IP Address Exclusion....21
1.3.14 Configuring IP Address Conflict Detection....22
1.3.15 Configuring Compulsory NAK Reply....23
1.3.16 Configuring Preferential Assignment of External DNS Server Addresses....23
1.3.17 Configuring VRRP Monitoring....24
1.3.18 Configuring ARP-based Go-Offline Detection....24
1.3.19 Saving Historical Leases....25
1.4 Configuring DHCP Relay....25
1.4.1 Overview....25
1.4.2 Configuration Task Summary....25
1.4.3 Configuring Basic DHCP Relay Functions....26
1.4.4 Configuring Option 82....27
1.4.5 Configuring Server-ID Check....28
1.4.6 Enabling DHCP Relay Suppression....28
1.4.7 Configuring Multiple Gateway IP Addresses....29
1.4.8 Forcing a DHCP Relay to Send a Reply Packet....30
1.4.9 Configuring the Source IP Address of DHCP Relay Packets....30
1.4.10 Configuring the Link-Selection Address for the DHCP Relay....31
1.5 Monitoring....31
1.6 Configuration Examples....33
1.6.1 Configuring Static Address Binding....33
1.6.2 Dynamically Assigning IP Addresses....34
1.6.3 Configuring DHCP Relay....37

1.6.4 Assigning IP Addresses Based on Class Rules 39

1 Configuring DHCP

1.1 Introduction

1. Overview

Dynamic Host Configuration Protocol (DHCP) is a local area network (LAN) protocol that dynamically assigns reusable network addresses and additional configurations to hosts. DHCP uses User Datagram Protocol (UDP) as its transport protocol to send and receive DHCP messages through port 67.

DHCP works in client/server mode. A DHCP client sends a request message to a DHCP server for an IP address and other configurations. When a DHCP client and a DHCP server are not on the same network segment, they need a DHCP relay agent to forward DHCP request and reply messages.

2. Origin and Development

DHCP originates from the Bootstrap Protocol (BOOTP), which is used by diskless hosts to obtain their own IP addresses, server IP address, boot image file name, gateway IP address, and other information from a server. BOOTP is designed for static environments. It requires the hardware address of a host be manually added to the BOOTP table and mappings between hardware addresses and IP addresses be static. In other words, BOOTP is less dynamic. In an environment with limited IP addresses, one-to-one mapping of BOOTP wastes IP addresses.

As the network scale expands and network complexity increases, BOOTP designed for static hosts can no longer meet requirements in wireless network scenarios and scenarios in which the number of computers exceeds the number of available IP addresses.

Therefore, the Internet Engineering Task Force (IETF) proposes a new protocol, that is, DHCP. DHCP is an enhanced version of BOOTP and consists of a server and a client. A DHCP server centrally manages all IP network settings and processes DHCP requests from DHCP clients. DHCP clients use the IP environment data assigned by the DHCP server. Compared with BOOTP, DHCP employs the lease concept to allocate clients' TCP/IP settings effectively and dynamically. As an enhancement version of BOOTP, DHCP also meets compatibility requirements of BOOTP clients.

3. Benefits

During network construction, a DHCP server can bring the following benefits:

- Low network access cost: Generally, the static address assignment mode considers the real locations of hosts and the labor cost is high. A DHCP application mode only needs to be uniformly configured on a server, thereby reducing a network access cost.
- Simplified configuration and low network construction cost: Dynamic address assignment significantly simplifies device configuration and reduces the deployment cost if no professional technical personnel are available to deploy devices.
- Centralized management: You can modify configurations for multiple subnets by simply modifying the DHCP server configuration.

1.2 Principles

1.2.1 Basic Concepts

- DHCP server

  A DHCP server is implemented based on RFC 2131. It assigns IP addresses to hosts and manages these IP addresses.

- DHCP client

  A DHCP client enables a device to automatically obtain an IP address and other configurations from a DHCP server.

- DHCP relay agent

  When a DHCP client and a DHCP server are not in the same subnet, they need a DHCP relay agent to forward DHCP request and reply messages.

- Lease time

  Lease time is a period of time specified by a DHCP server for a client to use an assigned IP address. An IP address can be leased to a client when a lease time is active. Before a lease expires, a client needs to renew the lease through a server. When a lease expires or is deleted from a server, the lease time becomes inactive.

- Address pool

  An address pool is a collection of IP addresses that a DHCP server can assign to clients.

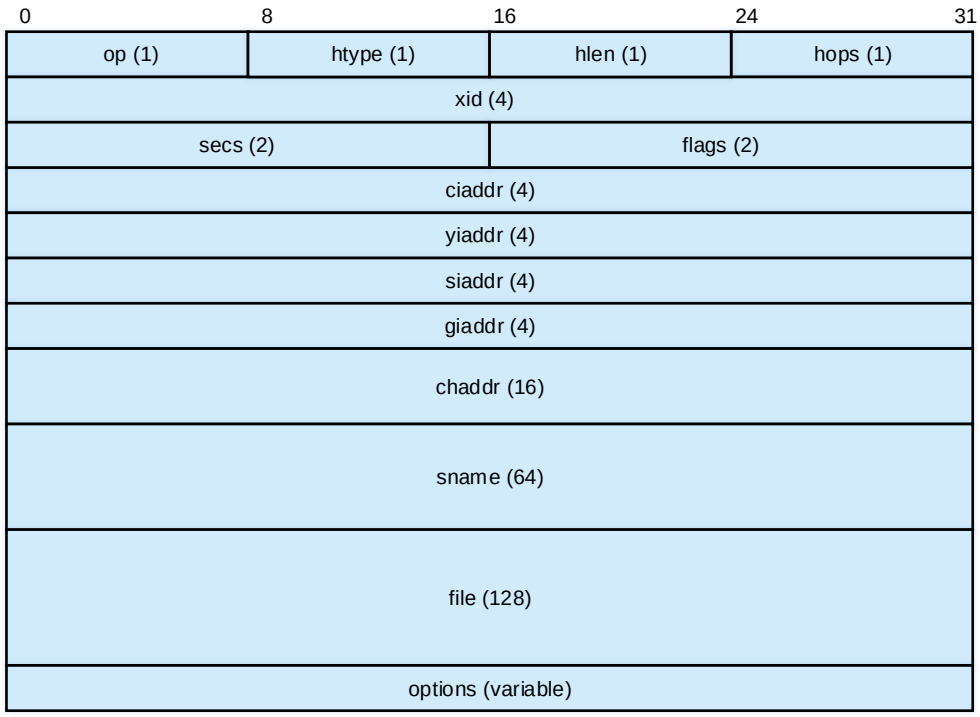
- Option type

  An option type is a parameter specified by a DHCP server when it provides the lease service to a DHCP client. For example, common options include the IP addresses of the default gateway, Windows Internet Name Service (WINS) server, and Domain Name System (DNS) server. A DHCP server also allows configuration of other options. Most options are predefined in RFC 2132. You can add user-defined options.

1.2.2 Message Format

1. DHCP Packet

Figure 1-1 DHCP Packet



The following describes fields contained in a DHCP packet:

- **op**: Packet type, including request packets and reply packets. A packet with **op** set to **1** is a request, and a packet with **op** set to **2** is a reply. The specific packet type is identified by the **options** field.
- **htype** (hardware type): Hardware address type of a DHCP client. For Ethernet clients, the field value is **1**.
- **hlen** (hardware length): Length of the hardware address of a DHCP client. For Ethernet clients, the field value is **6**.
- **hops**: Number of DHCP relay agents that a DHCP packet passes through. The field value increments by 1 each time a DHCP request packet passes through one DHCP relay agent.
- **xid**: Random number selected by a DHCP client when it sends a request. It is used to identify an address requesting process.
- **secs**: Time (in seconds) elapsed since a DHCP client starts a DHCP request.
- **flags**: The first bit is the broadcast reply flag bit and is used to identify whether a DHCP server sends a reply packet in unicast or broadcast mode. **0** indicates unicast, and **1** indicates broadcast. The other 15 bits are set to **0**.
- **ciaddr** (client IP address): IP address of a DHCP client. If a client has a valid and available IP address, add it to this field. During initialization, the client does not have an IP address, and this field is set to **0.0.0.0**. This field is not used to apply for a specific IP address by the client.
- **yiaddr** (your client IP address): IP address assigned by a DHCP server to a DHCP client.

- **siaddr** (server IP address): IP address of the server from which a DHCP client obtains boot configurations.
- **giaddr** (gateway IP address): IP address of the first DHCP relay agent that a request packet of a DHCP client passes through.
- **chaddr** (client hardware address): Hardware address of a DHCP client.
- **sname** (server name): Name of the server from which a DHCP client obtains boot configurations.
- **file**: Boot configuration file name and path specified by a DHCP server for a DHCP client.
- **options**: Optional fields with variable length, including the packet type, valid lease time, IP address of the DNS server, and IP address of the WINS server.

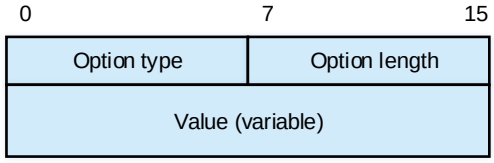
2. Message Types

- DHCP DISCOVER: first message broadcast by a DHCP client for DHCP interaction when it accesses a network for the first time.
- DHCP OFFER: message sent by a DHCP server in response to the DHCP DISCOVER message. The message carries various configuration parameters.
- DHCP REQUEST: This message has the following functions:
  - After a client is initialized, it broadcasts a DHCP REQUEST message to respond to the DHCP OFFER message of a DHCP server.
  - After a client is restarted and initialized, it broadcasts a DHCP REQUEST message to confirm the previously assigned IP address and other configurations.
  - After a client is bound to an IP address, it sends a DHCP REQUEST message to update the IP address lease.
- DHCP ACK: acknowledgment message of a DHCP server in response to a DHCP client's DHCP REQUEST message. A client obtains an IP address and related configuration parameters only after receiving this message.
- DHCP NAK: After a DHCP server receives a DHCP REQUEST message, the server cannot find the lease record or assign an IP address to the client due to some reasons. The server sends a DHCP NAK packet to inform the DHCP client that no proper IP address can be assigned.
- DHCP DECLINE: After a DHCP client receives an ACK message, the client finds that the IP address assigned by a server has been used through address conflict detection. Or the assigned IP address is unavailable due to other reasons. The client sends a DHCP DECLINE packet to inform the DHCP server that the assigned IP address is unavailable and request for a new IP address.
- DHCP REALEASE: When a DHCP client no longer uses the assigned IP address, it sends a DHCP RELEASE packet to inform the DHCP server of the assigned IP address not be used any longer and request the DHCP server to release the assigned IP address.

3. Option Format

A DHCP server uses the Options field to transfer control information and network configuration parameters to dynamically assign IP addresses and provide rich network configurations to clients. The Options field consist of the type, length, and value parts. Figure 1-2 shows the format.

Figure 1-2 Option Format



- **Option type**: option number.
- **Option length**: length of the information content.
- **Value**: information content.

4. Common DHCP Options

The type value of the **Options** field ranges from 1 to 255. Common DHCP options include:

- Option 1: subnet mask.
- Option 3: gateway address.
- Option 6: DNS address.
- Option 15: domain name.
- Option 33: static route. This option contains a group of classified static routes (that is, the mask of the destination network address is a natural mask and cannot be used for subnet division). After a client receives this option, it adds these static routes to the routing table. If Option 33 and Option 121 coexist, Option 33 is ignored.
- Option 51: IP address lease time.
- Option 53: DHCP message type.
- Option 55: parameter request list. A client uses this option to specify the network configurations that need to be obtained from a server. The option content is the values of parameters requested by a client.
- Option 121: classless static route. This option contains a group of classless static routes (that is, the mask of the destination network address is any value and subnets can be divided by mask). After a client receives this option, it adds these static routes to the routing table. If Option 33 and Option 121 coexist, Option 33 is ignored.

For details about more DHCP options, see RFC 2132 and RFC 3442.

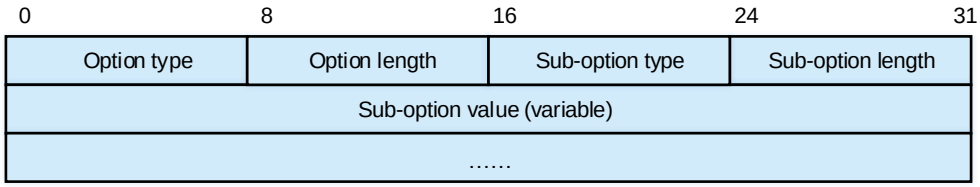
5. User-Defined DHCP Options

- Option 43: specific vendor information

  A DHCP server uses Option 43 to assign specific vendor information to a client. Option 43 uses several sub-options to assign different network configurations to users. The sub-option fields are as follows:

  - **Sub-option type**: type of a sub-option. The value can be **0x01** (ACS parameter), **0x02** (service provider identifier), or **0x80** (PXE boot server address).
  - **Sub-option length**: length of a sub-option.
  - **Sub-option value**: value of a sub-option.

Figure 1-3 Option 43

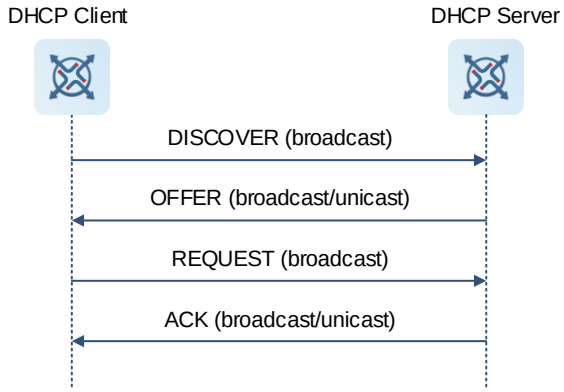


- Option 82: relay agent information

  Option 82 records the location information of a DHCP client. When a DHCP relay agent receives a request packet destined for a DHCP server from a DHCP client, the DHCP relay agent adds Option 82 to the packet and forwards the packet to the DHCP server. An administrator obtains the DHCP client location information from Option 82 to locate the DHCP client, assign an address in a specific range to the client, and control the client security and accounting.

1.2.3 DHCP Lease Process

Figure 1-4 DHCP Lease Process



A DHCP client requests an IP address as follows:

(1) The DHCP client broadcasts a DHCP DISCOVER packet to find a DHCP server in the network.

(2) The DHCP server sends a DHCP OFFER packet to the client in unicast/broadcast mode based on the broadcast bit in the **flags** field of the DISCOVER packet, with the packet containing an IP address, a Media Access Control (MAC) address, a domain name, and a lease time.

(3) The DHCP client broadcasts a DHCP REQUEST packet to formally request an IP address from the DHCP server.

(4) The DHCP server searches for the corresponding lease record based on the MAC address carried in the DHCP REQUEST packet.

  - If the lease record exists, the DHCP server sends a DHCP ACK packet to the DHCP client. After receiving the DHCP ACK packet, the DHCP client sends a free Address Resolution Protocol (ARP) packet to check whether another host uses the IP address assigned by the DHCP server. If the DHCP client does not receive a response within the specified time, the DHCP client uses the assigned IP address. If the DHCP client finds IP address conflict, it sends a DHCP DECLINE message to notify the

server and requests another IP address from the server after 10s.

- If the lease record does not exist or the DHCP server cannot assign an IP address to the client due to some reasons, the DHCP server sends a DHCP NAK packet to inform the DHCP client that no proper IP address can be assigned. The DHCP client needs to send another DHCP DISCOVER packet to apply for a new IP address.

Note

- The DHCP client may receive DHCP OFFER packets from multiple DHCP servers, but usually it accepts only the first DHCP OFFER packet.
- In addition, the address specified in the DHCP OFFER packet is not necessarily assigned to the client. Generally, the DHCP server reserves this address until the client sends a DHCP REQUEST packet.
- A client broadcasts a DHCP REQUEST packet to inform all DHCP servers that respond with DHCP OFFER packets. The servers not selected by the DHCP REQUEST message will release the previously assigned IP address.

1.2.4 DHCP Server

1. Address Pool

After a DHCP server receives a client's request, it chooses a valid address pool, determines an available IP address from the pool by using the ping mechanism, and delivers the pool configurations and IP address to the client. In addition, the server saves the lease information locally for validity check upon lease renewal.

Address pool configurations include the IP address range, gateway address, DNS address, and lease time.

2. Address Assignment Policy

A DHCP server selects an IP address for a client based on the following sequence:

(1) The DHCP server selects an IP address statically bound to the MAC address of the client from its own database.

(2) The DHCP server selects an IP address previously used by the client, that is, the address specified in the IP address option of the DHCP DISCOVER packet sent by the client.

(3) The DHCP server selects an available IP address first found from the DHCP address pool.

(4) If no available IP address is found from the DHCP address pool, the DHCP server checks whether an expired or conflicted IP address exists. If yes, it assigns an available IP address to the client. If not, it reports an error.

3. ARP Entry Check

The ARP entry check function can be used together with the ping mechanism to detect IP address conflicts. If a firewall is enabled on hosts in an actual environment, the ping operation for address conflict detection of a DHCP server becomes invalid. DHCP clients that dynamically apply for IP addresses may be assigned with the same IP address. Address conflict may occur. If the ARP entry check function is enabled, the DHCP server queries the ARP entries after detecting an IP address conflict by using the ping operation. If an ARP entry exists for the IP address to be assigned and is inconsistent with the MAC address of the client to which the IP address is assigned, the DHCP server regards that the IP address is occupied by another user.

You are not advised to enable the ARP entry check function in an environment with ARP attacks. Otherwise, the IP address assignment is affected, resulting in slow or failed IP address assignment.

4. Compulsory NAK Reply

In wireless applications, DHCP clients usually move from one network to another. When receiving a REQUEST packet from a client for lease renewal, a DHCP server replies with a NAK packet for the client if the server finds that the network segment of the client is changed or the lease expires. Such practice avoids the DHCP client from repeatedly sending REQUEST packets and avoids the DHCP client from obtaining an IP address only after the REQUEST packet times out.

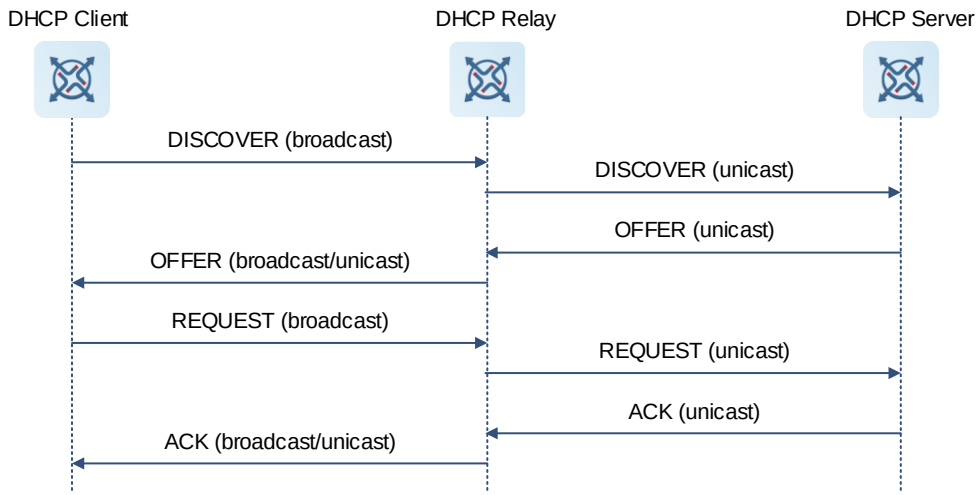
However, the DHCP server can send NAK packets only to its managed DHCP clients because it has lease records of its managed clients only. When a DHCP client moves from another network to the local network, if the DHCP server cannot find the corresponding lease record locally, the DHCP server does not reply with a NAK packet. In this case, the DHCP client needs to send REQUEST packets repeatedly. The client can obtain an IP address again only when the request times out. The time for obtaining an IP address is lengthened. This case also occurs when a client requires lease renewal but the DHCP server loses the lease record of the client due to a restart. In this case, you can configure compulsory NAK reply for the DHCP server if no lease record is found, so as to ensure that the client can obtain an IP address quickly.

1.2.5 DHCP Relay Agent

1. Principles

A DHCP relay agent helps to implement dynamic IP address assignment across network segments. The DHCP relay agent transparently forwards DHCP packets in unicast mode between a DHCP server and DHCP clients that reside on diverse network segments. It serves as a repeater. The Client-Relay-Server mode achieves dynamic management of IP addresses across multiple network segments with only one DHCP server.

Figure 1-5 Working Process

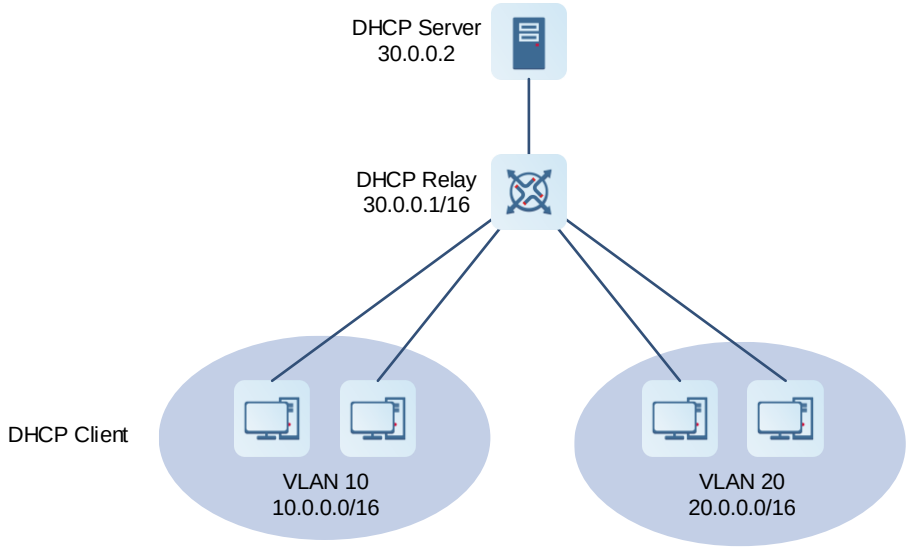


(1) After a DHCP relay agent receives a DHCP DISCOVER or DHCP REQUEST packet broadcast by a DHCP client, the DHCP relay agent adds its own IP address to the **giaddr** field of the packet and forwards the packet to a specified DHCP server in unicast mode.

(2) Based on the **giaddr** field, the DHCP server assigns an IP address and other configurations to the client, so that the DHCP relay agent forwards the configurations to the client, ensuring dynamic client configuration.

2. Typical Applications

Figure 1-6 DHCP Relay Application Scenario



VLAN 10 and VLAN 20 are on the network segments 10.0.0.1/16 and 20.0.0.1/16, respectively. A DHCP server with the IP address 30.0.0.2 is in network segment 30.0.0.1/16. To achieve dynamic management of IP addresses in VLAN 10 and VLAN 20 by the DHCP server, you only need to enable the DHCP relay function on a gateway and configure the IP address 30.0.0.2 for the DHCP server.

3. Option 82

As defined in RFC 3046, a DHCP relay agent can add the DHCP relay agent information option to indicate a DHCP client's network information, so that a DHCP server can assign IP addresses of various privileges based on more accurate information. The option is called Option 82.

Based on the physical port that receives DHCP request packets and the MAC address of the device, the DHCP relay agent constructs Option 82. Option 82 can contain a maximum of 255 sub-options. If Option 82 is defined, it should contain at least one sub-option. Currently, the device supports sub-option 1 (**Circuit ID**) and sub-option 2 (**Remote ID**). Figure 1-7 and Figure 1-8 show the formats of the two sub-options.

Figure 1-7 Option 82 Circuit ID

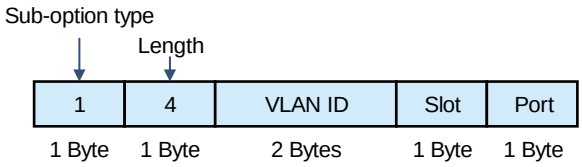
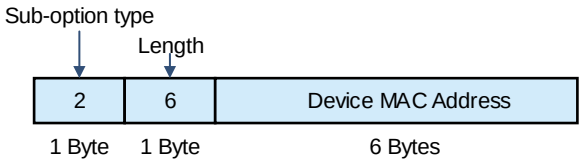


Figure 1-8 Option 82 Remote ID



4. Custom Option 82

Option 82 can be customized. A DHCP relay agent can form an Option 82 based on the physical port that receives DHCP request packets and the MAC address and the name of the device. Figure 1-9 and Figure 1-10 show the formats of the sub-options.

Figure 1-9 Custom Option 82 Circuit ID

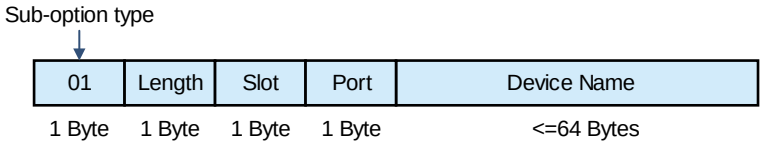
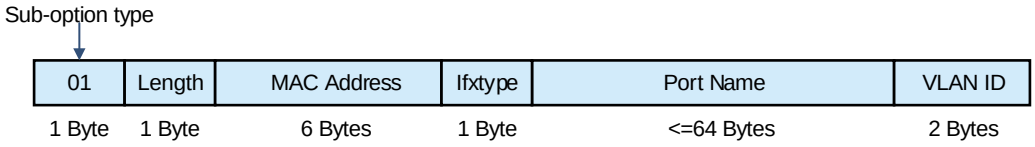


Figure 1-10 Custom Option 82 Remote ID

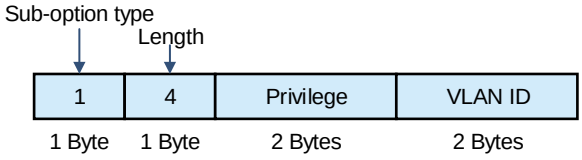


5. Option Dot1x

Option Dot1x should be used in combination with 802.1X authentication and SAM authentication and accounting management system.

A DHCP relay agent constructs the **Circuit ID** sub-option based on the IP privilege delivered during 802.1X authentication and the VLAN ID of a DHCP client. Figure 1-11 shows the option format.

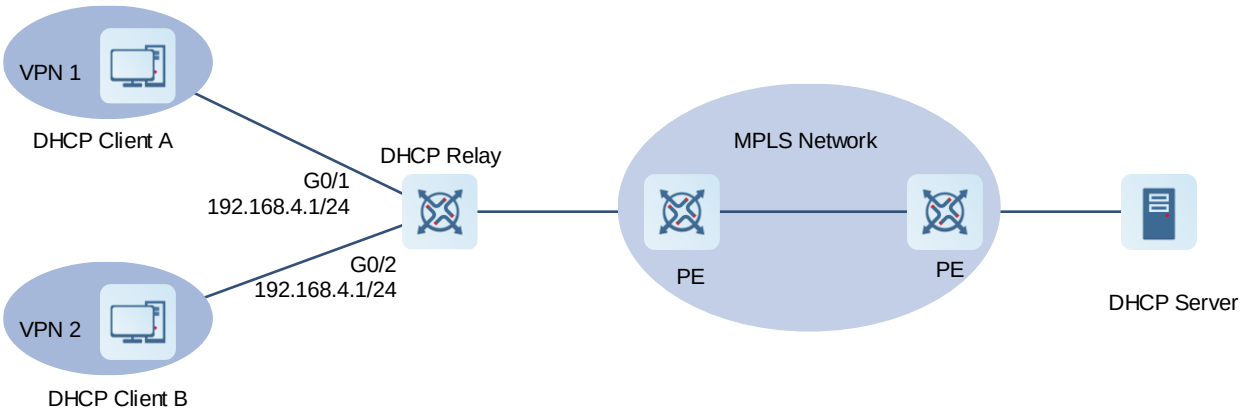
Figure 1-11 Option Dot1x Circuit ID Sub-option



6. Option VPN

Option virtual private network (VPN) should be used in combination with Multiprotocol Label Switching (MPLS) VPN functions.

Figure 1-12 Application in an MPLS VPN Environment



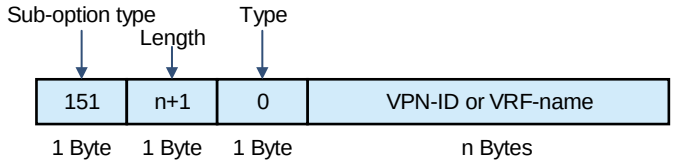
As shown in Figure 1-12, in an MPLS VPN environment, DHCP Client A is connected to GigabitEthernet 0/1 on a DHCP relay agent, and DHCP Client B is connected to GigabitEthernet 0/2 on the DHCP relay. GigabitEthernet 0/1 and GigabitEthernet 0/2 belong to different virtual routing and forwarding (VRF) instances. DHCP Client A and DHCP Client B obtain IP addresses through DHCP. According to the network plan, VPN 1 and VPN 2 share the network segment 192.168.4.0/24. In this case, traditional DHCP application cannot meet the deployment requirement.

To support DHCP Relay in the MPLS VPN environment, Option VPN is introduced, which includes three sub-options, namely, **VPN-ID**, **Subnet-Selection**, and **Server-Identifier-Override**.

- VPN-ID

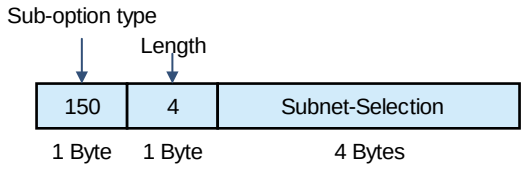
  When a DHCP relay agent receives a DHCP request packet, it adds the VPN information of the DHCP client to the packet as an option. The DHCP server retains **VPN-ID** in a reply packet, so that the DHCP relay agent forwards the packet to the correct VRF instance based on the option. Figure 1-13 shows the option format.

Figure 1-13 VPN-ID



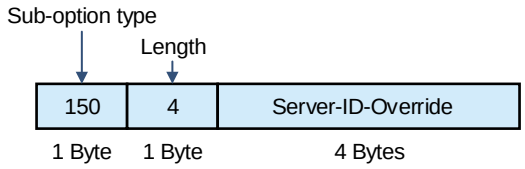
- **Subnet-Selection**: In a traditional DHCP Relay environment, the **giaddr** field is used to indicate information about a client's network and the addresses of a DHCP server and a DHCP relay agent. In an MPLS VPN environment, set **giaddr** to the IP address of an interface on a DHCP relay agent connected to a DHCP server, so that the server can communicate directly with the relay. The client's subnet must be indicated by the **Subnet-Selection** option. Figure 1-14 shows the option format.

Figure 1-14 Subnet-Selection



- **Server-ID-Override**: In an MPLS VPN environment, request packets from a DHCP client cannot be directly sent to a DHCP server. A DHCP relay agent uses this option to carry the IP address of the interface directly connected to the DHCP client. When the server sends a reply packet, this option overrides the **Server-Identifier** option. In this way, the DHCP client sends packets to the DHCP relay agent, and the DHCP relay agent forwards them to the DHCP server. Figure 1-15 shows the option format.

Figure 1-15 Server-Identifier-Override



1.2.6 DHCP Associated with VRRP Monitoring

In a Virtual Router Redundancy Protocol (VRRP) scenario, DHCP provides a command to monitor the VRRP status of the current interface. For interfaces configured with a VRRP address and VRRP monitoring, a DHCP server processes DHCP clients' request packets from an interface on a master device but discards the request packets from an interface on a backup device. If no VRRP address is configured on an interface, the DHCP server does not monitor the VRRP status and processes all DHCP request packets from this interface. VRRP monitoring is configured only on Layer 3 interfaces. It is disabled by default, namely, only the master device provides the DHCP service.

1.2.7 Protocols and Standards

- RFC 2131: Dynamic Host Configuration Protocol
- RFC 2132: DHCP Options and BOOTP Vendor Extensions
- RFC 3046: DHCP Relay Agent Information Option
- RFC 3442: Classless Static Route Option for DHCPv4

1.3 Configuring DHCP Server

1.3.1 Overview

Configure the device as a DHCP server to assign IP addresses to DHCP clients and manage these IP addresses and other configurations. You can also configure other custom functions for the DHCP server as required to customize DHCP services.

1.3.2 Configuration Task Summary

(1) Enabling DHCP Server

(2) (Optional) Configuring Static IP Address Assignment

(3) Configuring Dynamic Assignment of IP Addresses and Network Parameters

(4) (Optional) Configuring DHCP Address Pool Management Function

(5) (Optional) Providing the Boot Image File Name

(6) (Optional) Assigning IP Addresses Based on Class Rules

(7) (Optional) Assigning IP Addresses Based on AM Rules

(8) Configuring security functions of a DHCP server: All the configuration tasks below are optional. Perform the configuration tasks as required.

- Refreshing Trusted ARP Entries
- Configuring Rogue DHCP Server Detection

(9) Configuring other custom functions of the DHCP server: All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring IP Address Exclusion
- Configuring IP Address Conflict Detection
- Configuring Compulsory NAK Reply
- Configuring Preferential Assignment of External DNS Server Addresses
- Configuring VRRP Monitoring
- Configuring ARP-based Go-Offline Detection
- Saving Historical Leases

1.3.3 Notes

Function	Product Supporting the Function
Assigning IP Addresses Based on AM Rules	S6920-4C, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ

1.3.4 Enabling DHCP Server

1. Overview

For the device to function as a DHCP server and assign IP addresses and other network parameters to DHCP clients, enable the DHCP Server function on the device.

2. Restrictions and Guidelines

The **service dhcp-server** command enables both the DHCP server and DHCP relay functions. When the device is configured with a valid address pool, it acts as a server and directly processes packets. Otherwise, it serves as a relay and forwards the packets.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the DHCP server function.

service dhcp

The DHCP server function is disabled by default.

1.3.5 Configuring Static IP Address Assignment

1. Overview

Deliver specific IP addresses and configurations to specific DHCP clients, such as servers and printers.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an address pool and enter the DHCP address pool configuration mode.

ip dhcp-server pool *dhcp-pool*

(4) Configure the IP address and network mask of a client host.

host *ipv4-address* [*netmask*]

No IP address or network mask is configured for a client host by default.

(5) Configure the hardware address or unique ID of a client. Select either of the following configuration tasks.

- Configure a client hardware address.

  hardware-address *hardware-address* [*type*]

  No hardware address is configured by default.

- Configure a unique client ID.

  dhcp-client-identifier [*unique-identifier*]

  No DHCP client ID is configured by default.

  A client ID consists of the media type, MAC address, and interface name.

(6) (Optional) Configure a client name.

dhcp-client-name *client-name*

No client name is configured by default. A client name can be any characters in the American Standard Code for Information Interchange (ASCII) character set. The name should not contain the domain name.

1.3.6 Configuring Dynamic Assignment of IP Addresses and Network Parameters

1. Overview

Configure address pool attributes, such as the address lease time, DNS address, and custom options. A DHCP server can use options to send the information to clients, so as to ensure dynamic network parameter assignment.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an address pool and enter the DHCP address pool configuration mode.

ip dhcp-server pool *dhcp-pool*

(4) Configure the primary network segment for dynamic assignment in a DHCP address pool.

network *network-number mask* [*low-ipv4-address high-ipv4-address*]

No primary network segment for dynamic assignment in a DHCP address pool is configured by default.

(5) (Optional) Configure the default gateway for a client.

default-router *ipv4-address*&<1-8>

No default gateway is configured for a DHCP client by default.

The gateway address that a DHCP server specifies for a DHCP client must be in the same network segment as the address assigned to the client.

(6) (Optional) Configure the domain name for a client.

domain-name *domain-name*

No domain name is configured for a DHCP client by default.

After a DHCP client obtains a specified domain name, it can directly use its host name to access a host whose name contains the same domain name.

(7) (Optional) Configure the DNS server to be assigned to a DHCP client.

dns-server *ipv4-address*&<1-8>

No DNS server is configured for a DHCP client by default.

When multiple DNS servers are defined, the first defined DNS server has the highest priority. A DHCP client selects the next DNS server only when it fails to communicate with the first defined DNS server.

(8) (Optional) Configure the NetBIOS WINS server for a DHCP client.

netbios-ns { *ipv4-address*&<1-8> }

No NetBIOS WINS server is configured for a DHCP client by default.

When multiple NetBIOS WINS servers are defined, the first defined NetBIOS WINS server has the highest priority. A DHCP client selects the next NetBIOS WINS server only when it fails to communicate with the first defined NetBIOS WINS server.

(9) (Optional) Configure the type of a NetBIOS node to be assigned to a DHCP client.

netbios-node-type *type*

The default node type of a DHCP client running a Microsoft operating system is broadcast or hybrid. If no WINS server is configured, the client is a broadcast node. Otherwise, it is a hybrid node.

(10) (Optional) Define DHCP server options.

option *code* { **ascii** *string* | **hex** *string* | **ip** *ipv4-address* }

No DHCP server option is defined by default.

1.3.7 Configuring DHCP Address Pool Management Function

1. Configuration Tasks

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring an Address Lease Time
- Adding Trusted ARP Entries During Address Assignment
- Forcibly Disabling Gateway Assignment
- Enabling or Disabling an Address Pool
- Configuring an Alarm Threshold for an Address Pool

2. Configuring an Address Lease Time

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an address pool and enter the DHCP address pool configuration mode.

ip dhcp-server pool *dhcp-pool*

(4) Configure an address lease time.

lease { *days* [*hours*] [*minutes*] | **infinite** }

The default lease time of a static address pool is permanent by default. The default lease time of other address pools is 1 day.

3. Adding Trusted ARP Entries During Address Assignment

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an address pool and enter the DHCP address pool configuration mode.

ip dhcp-server pool *dhcp-pool*

(4) Add trusted ARP entries.

update arp

No trusted ARP entry is added during DHCP address assignment by default.

4. Forcibly Disabling Gateway Assignment

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an address pool and enter the DHCP address pool configuration mode.

ip dhcp-server pool *dhcp-pool*

(4) Forcibly disable gateway assignment.

force-no-router

A DHCP server assigns a gateway to a DHCP client by default.

5. Enabling or Disabling an Address Pool

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an address pool and enter the DHCP address pool configuration mode.

ip dhcp-server pool *dhcp-pool*

(4) Enable or disable the address pool.

pool-status { **disable** | **enable** }

A created address pool is enabled by default.

6. Configuring an Alarm Threshold for an Address Pool

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an address pool and enter the DHCP address pool configuration mode.

ip dhcp-server pool *dhcp-pool*

(4) Configure an alarm threshold for the address pool.

lease-threshold *threshold-percentage*

The default alarm threshold is 90% for the address pool.

1.3.8 Providing the Boot Image File Name

1. Overview

Some DHCP clients need to download the operating system or configuration file in the boot process. A DHCP server must provide the required image file name during boot for a DHCP client to download the file from the corresponding server.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the DHCP address pool configuration mode.

ip dhcp-server pool *dhcp-pool*

(4) Configure the list of boot servers that a DHCP client needs to access during boot.

next-server { *ipv4-address*&<1-8> }

No list of boot servers that a DHCP client needs to access during boot is configured by default.

When multiple boot servers are defined, the first defined boot server has the highest priority. A DHCP client selects the next boot server only when it fails to communicate with the first defined boot server.

(5) (Optional) Configure the boot file name for a DHCP client.

bootfile *file-name*

No boot file name is configured for a DHCP client by default.

1.3.9 Assigning IP Addresses Based on Class Rules

1. Overview

In DHCP relay scenarios, addresses in a DHCP address pool can be assigned based on Option 82. When clients apply for IP addresses through different access points (APs), Option 82 carried in packets of the clients may be different. After class rules are configured on a DHCP server, the DHCP server can assign IP addresses in different network segments to clients based on Option 82 carried in the packets of the clients.

During address assignment, a DHCP server first determines an available address pool based on the network segment of a client. Then, it determines the class of the client based on Option 82, and assigns an IP address from the network segment corresponding to the class. When a request packet matches multiple classes in the address pool, the DHCP server assigns an IP address from the network segments corresponding to the classes based on the class configuration sequence. If the number of assigned IP addresses of a class reaches the limit, the DHCP server assigns an IP address based on the next matching class.

2. Restrictions and Guidelines

A DHCP address pool can be associated with multiple classes.

Each class corresponds to one network segment. Network segments are assigned in ascending order, and the network segments of multiple classes can overlap. If a class is associated with an address pool but no network segment is configured for the class, the default network segment of the class is the same as the network segment of the address pool.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a class rule and enter the class configuration mode.

ip dhcp-server class-config *class-name*

No class rule is configured by default.

(4) (Optional) Configure identification information for a class.

comment *class-remark*

No identification information is configured for a class by default.

(5) Enter the Option 82 information configuration mode.

relay option82 class

(6) Configure an Option 82 information matching rule for a class.

relay-agent-info hex *hex-string*

No Option 82 matching rule is configured by default.

(7) Return to the global configuration mode.

exit

(8) Enter the DHCP address pool configuration mode.

ip dhcp-server pool *dhcp-pool*

(9) Associate an address pool with a class.

class-config *class-name*

No class is associated with a DHCP address pool by default.

(10) (Optional) Configure the IP address range corresponding to a class rule.

address scope *low-ipv4-address high-ipv4-address*

The IP address range corresponding to a class rule is the network segment of the associated address pool by default.

(11) Return to the global configuration mode.

exit

(12) Enable address assignment based on class rules.

ip dhcp-server enable class

Address assignment based on class rules is not configured by default.

1.3.10 Assigning IP Addresses Based on AM Rules

Note

Only S6920-4C, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ support the function of assigning IP addresses by the DHCP server based on AM rules.

1. Overview

In super VLAN scenarios, address management (AM) rules can be used to plan IP address ranges for DHCP clients in different VLANs or DHCP clients using different ports, so that IP addresses can be assigned effectively and more flexibly. After an AM rule is configured, all DHCP clients from the specified VLANs and ports can obtain IP addresses. If a DHCP client is from a VLAN or port that is not specified, there are two cases: If a default AM rule is configured, the client obtains an IP address from the default IP address range; if no default AM rule is configured, the client cannot obtain an IP address.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the AM rule configuration mode.

address-manage

(4) Configure an AM rule. Select at least one of the following configuration steps.

- Configure the default IP address range matched with an AM rule.

  match ip default *ipv4-address* { *mask* | *mask-length* }

- Configure dynamic address assignment based on the VLAN and port.

  match ip *ipv4-address* { *mask* | *mask-length* } [*interface-type interface-number*] [**add | remove**] **vlan** *vlan-id*

- Configure static address assignment based on the VLAN.

  match ip *ipv4-address* { *mask* | *mask-length* } [**add** | **remove**] **vlan** *vlan-id*

No AM matching rule is configured by default.

(5) (Optional) Configure the loose mode for an AM rule.

match ip loose

The loose mode of AM rules is disabled by default.

If the loose mode is configured, clients that match no AM rule can obtain IP addresses in the way same as the case with no AM rule configured.

1.3.11 Refreshing Trusted ARP Entries

1. Overview

Trusted ARP entries protect a gateway from ARP spoofing. Devices with DHCP enabled provide a configuration command to determine whether trusted ARP entries are delivered during address assignment. After the function of adding trusted ARP entries is configured, a DHCP server adds trusted ARP entries during address assignment to prevent ARP spoofing. A trusted ARP entry has a higher priority than a dynamic ARP entry and is not overridden by the dynamic ARP entry.

2. Restrictions and Guidelines

After the function of refreshing trusted ARP entries is configured, a DHCP server refreshes trusted ARP entries only for clients assigned with addresses from an address pool that is configured with the **update arp** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Refresh trusted ARP entries.

ip dhcp-server refresh arp

1.3.12 Configuring Rogue DHCP Server Detection

1. Overview

If a DHCP server is deployed illegally, when clients request IP addresses from this server, clients are assigned with wrong addresses. This server is a rogue server. Devices with DHCP enabled provide a command to enable rogue server detection. After rogue server detection is enabled, DHCP packets are checked for Option 54 (**Server Identifier**). If Option 54 is carried and its value is different from the actual DHCP server identifier, the IP address of the rogue server and the port receiving the packets are recorded. Rogue server detection is only a post-event detection function for security and cannot prevent an illegal DHCP server from assigning IP addresses to clients.

After rogue server detection is configured, rogue servers in a network are recorded into logs.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure rogue server detection.

ip dhcp-server server detect

Rogue server detection is disabled by default.

1.3.13 Configuring IP Address Exclusion

1. Overview

An excluded address is an IP address or address segment excluded from the address pool of a DHCP server, so that the DHCP server does not assign this IP address or address segment to a client. Excluding IP addresses can improve the address assignment efficiency.

2. Configuring Excluded IP Addresses

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure excluded IP addresses.

ip dhcp-server excluded-ip-address *low-ipv4-address* [*high-ipv4-address*]

By default, a DHCP server assigns all addresses from an IP address pool to DHCP clients.

1.3.14 Configuring IP Address Conflict Detection

1. Overview

When the DHCP server attempts to assign an IP address from the address pool, it uses the ping operation to check whether the IP address is occupied by another host. If yes, the DHCP server records the IP address. If not, the DHCP server assigns the IP address to a DHCP client.

The ARP entry check function can be used with the ping operation to detect IP address conflicts. If a client is configured with a static IP address and Layer 2 isolation is configured in an environment and the ping operation for IP address conflict detection becomes invalid (for example, a firewall is enabled on the client), this IP address may be assigned to another client that dynamically requests an address, resulting in an IP address conflict.

2. Configuration Tasks

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring the Number of Ping Operations Executed for Address Conflict Detection
- Configuring the Timeout Time of a Ping Operation Executed for Address Conflict Detection
- Configuring ARP Entry Check

3. Configuring the Number of Ping Operations Executed for Address Conflict Detection

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the number of ping operations executed by a DHCP server when it detects IP address conflicts.

ip dhcp-server ping numbers [*ping-times*]

A DHCP server pings a conflicted IP address two times by default.

4. Configuring the Timeout Time of a Ping Operation Executed for Address Conflict Detection

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the timeout time of a ping operation executed when a DHCP server detects address conflicts.

ip dhcp-server ping timeout *time*

The default timeout time of a ping operation for detecting address conflicts is 500 ms.

5. Configuring ARP Entry Check

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the ARP entry check.

ip dhcp-server arp-probe

The ARP entry check is disabled by default.

1.3.15 Configuring Compulsory NAK Reply

1. Overview

When a DHCP client moves from another network to the local network, the DHCP server does not reply with a NAK packet because it cannot find the corresponding lease record locally. In this case, the DHCP client needs to send REQUEST packets repeatedly. The client can obtain an IP address again only when the request times out, and the time for obtaining an IP address is lengthened. This case also occurs when a client requires lease renewal but the DHCP server loses the lease record of the client due to a restart. In this case, you can configure compulsory NAK reply for the DHCP server if no lease record is found, so as to ensure that the client can obtain an IP address quickly.

2. Restrictions and Guidelines

When compulsory NAK reply is enabled, enable only one DHCP server in a broadcast domain. Configuring compulsory NAK reply may cause exceptions to other servers.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure compulsory NAK reply for a DHCP server.

ip dhcp-server force-send-nak

Compulsory NAK reply is enabled by default.

1.3.16 Configuring Preferential Assignment of External DNS Server Addresses

1. Overview

When a device works in the Point-to-Point Protocol over Ethernet (PPPoE) or DHCP client mode, it can obtain DNS server addresses automatically from an external DHCP server. Configure the DNS server addresses on the DHCP server of the local device, so that users do not need to perform DNS configuration. When a device serves as a DHCP server, it preferentially assigns clients with a DNS server address obtained from an external DHCP server.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure preferential assignment of external DNS server addresses to clients when the device works in the DHCP client or PPPoE mode.

ip dhcp-server dns dynamic

By default, run the **dns-server** command to assign DNS server addresses to clients.

1.3.17 Configuring VRRP Monitoring

1. Overview

After VRRP monitoring is enabled in VRRP application scenarios, only the master server processes DHCP packets.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the switch virtual interface (SVI) configuration mode.

interface vlan *interface-number*

(4) Enable VRRP monitoring.

ip dhcp-server monitor-vrrp-state

VRRP monitoring is disabled for an interface by default.

1.3.18 Configuring ARP-based Go-Offline Detection

1. Overview

Devices with DHCP enabled provide a command to enable ARP-based go-offline detection. After this function is enabled, a DHCP server receives an ARP entry aging notification when a client goes offline, and reclaims the client's IP address. If the client does not go online within a period of time (5 minutes by default), the DHCP server reclaims the IP address and assigns it to another client. If the client goes online again within the specified period, it can still use the original IP address.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure ARP-based go-offline detection.

ip dhcp-server server arp-detect

ARP-based go-offline detection is not configured for a DHCP server by default.

1.3.19 Saving Historical Leases

1. Overview

Enable a DHCP server to save historical leases, the DHCP server saves the IP address lease of a client to the database after an IP address is assigned to a client and the client goes offline. When the client goes online again, the DHCP server assigns this address to the client again. Historical leases are saved when a DHCP process restarts or the device performs a hot backup switchover.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of saving historical leases.

ip dhcp-server save-history-enable

The function of saving historical leases is disabled by default.

1.4 Configuring DHCP Relay

1.4.1 Overview

When a DHCP client and a DHCP server are in different network segments, a DHCP relay agent is required for dynamic IP address management.

1.4.2 Configuration Task Summary

(1) Configuring Basic DHCP Relay Functions

(2) Configuring extended DHCP relay functions: All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring Option 82
- Configuring Server-ID Check
- Enabling DHCP Relay Suppression
- Configuring Multiple Gateway IP Addresses

- Forcing a DHCP Relay to Send a Reply Packet
- Configuring the Source IP Address of DHCP Relay Packets

1.4.3 Configuring Basic DHCP Relay Functions

1. Overview

Generally, the DHCP relay function is implemented by an interface on a device. The IP address of the server should be specified on the DHCP relay agent. In this case, DHCP broadcast packets received over this interface are sent to the specified server.

2. Restrictions and Guidelines

To enable DHCP relay, configure IPv4 unicast routing in a network.

- The **service dhcp-server** command is used to enable both the DHCP server and DHCP relay functions. However, a device cannot function as a DHCP server and relay at the same time. When a device is configured with a valid address pool, it acts as a server. Otherwise, it serves as a relay.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the DHCP relay function.

service dhcp-server

The DHCP relay function is disabled by default.

(4) Configure the IP address of the DHCP server.

ip dhcp-server-address { **cycle-mode** | [**vrf** *vrf-name*] *ipv4-address* }

The IP address of the DHCP server is not configured for the DHCP relay agent by default.

The DHCP server address can be globally configured or configured on a Layer 3 interface. A maximum of 20 DHCP server addresses can be globally configured or configured on each Layer 3 interface. When an interface receives a DHCP request packet, the DHCP server list on the interface prevails over that configured globally. If the interface is not configured with the DHCP server list, the global DHCP server list takes effect.

(5) (Optional) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI interface configuration mode.
  interface vlan *interface-number*

(6) (Optional) Disable DHCP relay on the interface.

ip dhcp-server relay-agent disable

DHCP relay is enabled on the interface by default.

1.4.4 Configuring Option 82

1. Overview

A DHCP relay agent can use Option 82 to identify network information of DHCP clients, so that a DHCP server can assign IP addresses of different privileges to the clients more accurately based on this option.

2. Restrictions and Guidelines

- Configure basic DHCP Relay functions.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable DHCP Option 82.

ip dhcp-server relay-agent information option82

DHCP Option 82 is disabled by default.

(4) Configure sub-options of DHCP Option 82. All the configuration steps below are optional. Select configuration steps as required.

- Customize information in **Circuit ID**.

  ip dhcp-server relay-agent information option82 user-defined circuit-id *circuit-id-text*

  No custom information is configured for **Circuit ID** of DHCP Option 82 by default.

- Customize information in **Remote ID**.

  ip dhcp-server relay-agent information option82 user-defined remote-id *remote-id-text*

  No custom information is configured for **Remote ID** of DHCP Option 82 by default.

  The circuit-ID and remote-ID of option 82 can be configured either globally or on a Layer 3 interface. The interface configuration takes precedence over the global configuration.

(5) (Optional) Customize the MAC address format.

ip dhcp-server relay-agent information option82 user-defined mac-format *mac-format-type*

The default MAC address format is H.H.H.

(6) (Optional) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI interface configuration mode.
  interface vlan *interface-number*

(7) (Optional) Enable DHCP option 82 on the specified interface.

ip dhcp-server relay-agent-info enable

DHCP option 82 is disabled by default.

(8) (Optional) Enable the sub-options of DHCP option 82 on the specified interface.

- Configure the circuit-ID.
  - **ip dhcp-server relay-agent-info layer3 circuit-id format** { **ascii** | **hex | user-defined** } *option82-circuit-id*
    - The circuit-ID of DHCP option 82 is not configured by default.
- Configure the remote-ID.
  - **ip dhcp-server relay-agent-info layer3 remote-id format** { **ascii** | **hex | user-defined** } *option82-remote-id*
    - The remote-ID of DHCP option 82 is not configured by default.
    -

1.4.5 Configuring Server-ID Check

1. Overview

In a DHCP relay application environment, multiple DHCP servers are configured in each network to provide server backup, thereby helping ensure uninterrupted network operation. When a DHCP client has selected a DHCP server to send a DHCP REQUEST packet, a **Server-ID** option is carried in the packet. To reduce server load in specific environments, enable the **Server-ID** check function on the DHCP relay agent, so as to send the DHCP REQUEST packet only to the DHCP server specified in this option.

In this case, the DHCP relay agent sends DHCP request packets only to the specified server. If this function is not configured, the DHCP relay agent sends DHCP request packets to all configured DHCP servers.

2. Restrictions and Guidelines

3. Configure basic DHCP Relay functions. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable a DHCP relay agent to check **Server-ID**.

ip dhcp-server relay-agent check server-id

By default, a DHCP relay agent forwards a DHCP Request message to all DHCP servers.

1.4.6 Enabling DHCP Relay Suppression

1. Overview

After you enable DHCP relay suppression on an interface, DHCP request packets received over the interface are filtered out, but the other DHCP requests are forwarded.

2. Restrictions and Guidelines

Configure basic DHCP Relay functions.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

(4) Enable the DHCP relay suppression function.

ip dhcp-server relay-agent suppression

The DHCP relay suppression function is disabled on all interfaces by default.

1.4.7 Configuring Multiple Gateway IP Addresses

1. Overview

After the function of configuring multiple gateway IP addresses is enabled, a DHCP relay agent can use multiple interface IP addresses to send address requests to a DHCP server. Generally, the primary IP address is used as the gateway IP address, and the DHCP server assigns a network segment based on the gateway IP address. When a client fails to apply for an IP address through the gateway by using the primary IP address, it applies for an IP address through the gateway by using a secondary IP address.

After the automatic gateway switchover function is enabled, the DHCP relay agent adds another address to the **giaddr** field if it does not receive a reply for three consecutive DISCOVER packets.

2. Restrictions and Guidelines

Configure basic DHCP Relay functions.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of configuring multiple gateway IP addresses on a DHCP relay agent.

ip dhcp-server relay-agent multiple-giaddr

The function of configuring multiple gateway IP addresses on a DHCP relay agent is disabled by default.

(4) (Optional) Enable the automatic gateway switchover function.

ip dhcp-server auto-switch-relay

The automatic gateway switchover function is disabled by default.

1.4.8 Forcing a DHCP Relay to Send a Reply Packet

1. Overview

After the function of forcing a DHCP relay agent to send a reply packet is enabled, a DHCP relay agent forcibly specifies a gateway interface to send a reply if it fails to find a MAC address egress. When the function is disabled, the DHCP relay agent discards packets if it fails to find a MAC address egress.

2. Restrictions and Guidelines

You must enable basic DHCP Relay functions.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of forcing a DHCP relay agent to send a reply packet.

ip dhcp-server relay-agent force-send-reply-pack

The function of forcing a DHCP relay agent to send a reply packet is disabled by default.

1.4.9 Configuring the Source IP Address of DHCP Relay Packets

1. Overview

You can run the **ip dhcp-server relay-agent source** command on an interface to configure the source IP address of DHCP relay packets.

2. Restrictions and Guidelines

- Configure basic DHCP Relay functions.
- Only one source IP address can be specified for DHCP relay packets on each interface.
- This function must be used with Option 82.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the switch virtual interface (SVI) configuration mode.

  interface vlan *interface-number*

(4) Configure the source IP address for DHCP relay packets.

ip dhcp-server relay-agent source *ipv4-address*

No source address is configured for DHCP relay packets by default.

1.4.10 Configuring the Link-Selection Address for the DHCP Relay

1. Overview

After the **ip dhcp-server relay-agent link-selection** command is configured on an interface, the specified unicast IP address can be filled in sub option 5 of option 82 in packets on the interface.

2. Restrictions and Guidelines

- Basic functions of DHCP relay must be configured.
- Only one link-selection address can be specified for one interface.
- This function needs to be used together with the function of specifying the source address for DHCP relay packets (by running the **ip dhcp-server relay-agent source** command) and the option 82 function (by running the **ip dhcp-server relay-agent class option82** command in global configuration mode or the **ip dhcp-server relay-agent-info enable** command in interface configuration mode). This command can take effect only after the two functions are enabled.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the L3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the L3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

(4) Configure the link-selection address for the DHCP relay.

ip dhcp-server relay-agent link-selection *ip-address*

The function of enabling the DHCP relay to fill in sub option 5 of option 82 is not configured by default, and the master IP address of an interface is filled in sub option 5 of option 82 by default.

1.5 Monitoring

Run the **display** command to check the configuration.

Run the **debugging** command to output debugging information.

Caution

The output of debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** command to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup ip dhcp-server binding-database { *ipv4-address* | ***** }	Clears DHCP address binding records.
cleanup ip dhcp-server conflict-record { *ipv4-address* | ***** }	Clears DHCP address conflict records.
cleanup ip dhcp-server history { *mac-address* | ***** }	Clears DHCP historical lease records.
cleanup ip dhcp-server server detect { *ipv4-address* | ***** }	Clears rogue DHCP server detection records.
cleanup ip dhcp-server server rate	Clears DHCP server performance statistics.
cleanup ip dhcp-server server count	Clears DHCP server statistics.
cleanup ip dhcp-server relay-agent count	Clears DHCP relay statistics.
display ip dhcp-server binding-database	Displays DHCP address bindings.
display ip dhcp-server conflict-record	Displays address conflict records of a DHCP server.
display ip dhcp-server database	Displays the running status of the database backup function of a DHCP server.
display ip dhcp-server dns dynamic	Displays DNS server addresses obtained from an external DHCP server when the device works in the PPPoE or DHCP client mode.
display ip dhcp-server history	Displays DHCP historical lease records.
display ip dhcp-server identifier	Displays the address pool ID and address usage of a DHCP server.
display ip dhcp-server pool [*pool-name*]	Displays the address pool status and utilization of a DHCP server.

display ip dhcp-server relay-statistic	Displays DHCP relay statistics.
display ip dhcp-server server detect	Displays rogue DHCP server detection information.
display ip dhcp-server server statistic	Displays DHCP server statistics.
display ip dhcp-server socket	Displays DHCP socket indexes.
debugging ip dhcp-server server option82	Debugs a DHCP server.
debugging ip dhcp-server server ha	Debugs DHCP hot backup.
debugging ip dhcp-server server pool	Debugs DHCP address pools.
debugging ip dhcp-server server vrrp	Debugs association between a DHCP server and a VRRP group.
debugging ip dhcp-server server all	Debugs all DHCP servers.
debugging ip dhcp-server relay-agent	Debugs a DHCP relay.

1.6 Configuration Examples

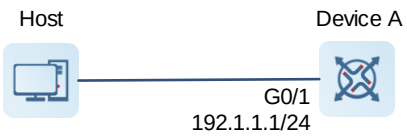
1.6.1 Configuring Static Address Binding

1. Requirements

As shown in Figure 1-16, a host that serves as a DHCP client obtains a statically bound IP address from Device A that serves as a DHCP server. The MAC address of the host is 0050.56b0.2f50.

2. Topology

Figure 1-16 Configuring Static Address Binding



3. Notes

Configure Device A:

- Configure the IP address of the interface.
- Enable the DHCP server service and configure address pool parameters.

4. Procedure

(1) Configure Device A.

Configure the IP address of the interface.

```
DeviceA> privilege
```

```
DeviceA # config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.1.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Enable the DHCP server service.

```
DeviceA(cfg)# service dhcp-server
```

Create an address pool named **pool1**.

```
DeviceA(cfg)# ip dhcp-server pool pool1
```

Assign a statically bound IP address to the host.

```
DeviceA(cfg-dhcp-pool)# host 192.1.1.99 255.255.255.0
DeviceA(cfg-dhcp-pool)# hardware-address 0050.56b0.2f50
```

(2) Configure the host.

Enable DHCP on the host to obtain an IP address. (Omitted)

5. Verification

Check that the host has obtained the IP address 192.1.1.99/24.

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
 interface gigabitethernet 0/1
 ip address 192.1.1.1 255.255.255.0
!
 service   dhcp-server
!
 ip   dhcp-server pool pool1
 host 192.1.1.99 255.255.255.0
 hardware-address 0050.56b0.2f50
!
```

7. Common Errors

- No address pool is configured.
- The DHCP Server service is not enabled.

1.6.2 Dynamically Assigning IP Addresses

1. Requirements

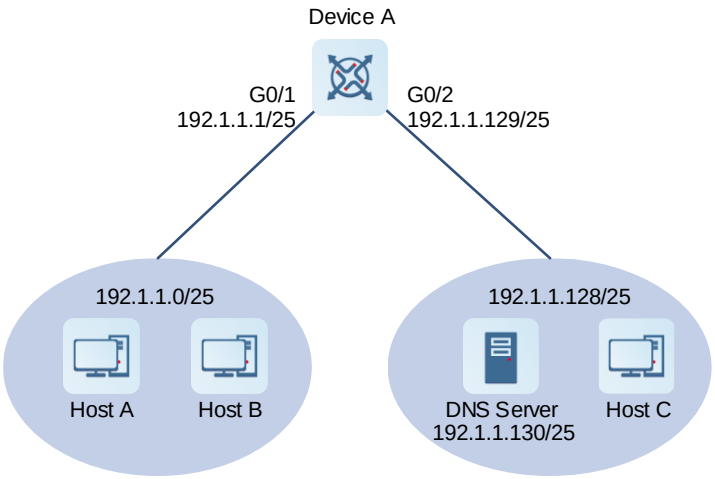
As shown in Figure 1-17, Device A serving as a DHCP server dynamically assigns IP addresses to clients in the same network segment. The address pool is divided into two network segments: 192.1.1.0/25 and 192.1.1.128/25.

The lease time of IP addresses in network segment 192.1.1.0/25 is 5 days, the domain name suffix is **test.com**, the DNS address is 192.1.1.130, and the gateway address is 192.1.1.1. The lease time of IP addresses in

network segment 10.1.1.128/25 is 3 days, the domain name suffix is **test.com**, the DNS address is 192.1.1.130, and the gateway address is 192.1.1.129.

2. Topology

Figure 1-17 Configuring Dynamic IP Address Assignment



3. Notes

Configure Device A:

- Configure the IP address of the interface for connecting to the host.
- Enable the DHCP Server service and configure parameters of two address pools.

4. Procedure

(1) Configure Device A:

Configure the IP address of the interface for connecting to the host.

```
DeviceA> privilege
DeviceA # config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.1.1.1 255.255.255.128
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# ip address 192.1.1.129 255.255.255.128
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

Enable the DHCP Server service.

```
DeviceA(cfg)# service dhcp-server
```

Configure network parameters of pool 1.

```
DeviceA(cfg)# ip dhcp-server pool pool1
DeviceA(cfg-dhcp-pool)# network 192.1.1.0 255.255.255.128
DeviceA(cfg-dhcp-pool)# default-router 192.1.1.1
DeviceA(cfg-dhcp-pool)# dns-server 192.1.1.130
DeviceA(cfg-dhcp-pool)# domain-name test.com
```

```
DeviceA(cfg-dhcp-pool)# lease 5
DeviceA(cfg-dhcp-pool)# exit
```

Configure network parameters of pool 2.

```
DeviceA(cfg)# ip dhcp-server pool pool2
DeviceA(cfg-dhcp-pool)# network 192.1.1.128 255.255.255.128
DeviceA(cfg-dhcp-pool)# default-router 192.1.1.129
DeviceA(cfg-dhcp-pool)# dns-server 192.1.1.130
DeviceA(cfg-dhcp-pool)# domain-name test.com
DeviceA(cfg-dhcp-pool)# lease 3
```

(2) Configure the DNS server:

Set the DNS server IP address to 192.1.1.130/25. (Omitted)

(3) Configure the host:

Enable DHCP on Host A, Host B, and Host C to obtain IP addresses. (Omitted)

5. Verification

Check that Host A and Host B have obtained addresses in network segment 192.1.1.0/25.

Check that Host C has obtained an address in network segment 192.1.1.128/25.

Run the **display ip dhcp-server pool** command on Device A to display address pool configurations and usage.

```
DeviceA(cfg)# display ip dhcp-server pool
Pool  name          Total       Distributed   Remained     Percentage
------------  -----------  -----------  --------  -----------
pool1             126             2              124          0.98413
pool2             126             1              125          0.99206
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
 interface gigabitethernet 0/1
 ip address 192.1.1.1 255.255.255.128
!
 interface gigabitethernet 0/2
 ip address 192.1.1.129 255.255.255.128
!
 service  dhcp-server
!
 ip  dhcp-server pool pool1
 network 192.1.1.0 255.255.255.128
 default-router 192.1.1.1
 dns-server 192.1.1.130
 domain-name test.com
 lease 5
!
 ip  dhcp-server pool pool2
```

```
network 192.1.1.128 255.255.255.128
default-router 192.1.1.129
dns-server 192.1.1.130
domain-name test.com
lease 3
!
```

7. Common Errors

- No address pool is configured.
- The DHCP Server service is not enabled.

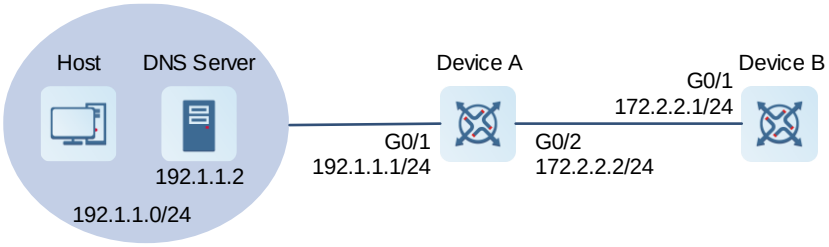
1.6.3 Configuring DHCP Relay

1. Requirements

As shown in Figure 1-18, a host serving as a DHCP client is located in network segment 192.1.1.0/24, Device A is the gateway, and Device B is a DHCP server. To apply for an IP address and other configurations from the DHCP server, the DHCP client needs to use the gateway as a DHCP relay agent, so that the DHCP relay agent forwards DHCP packets to the DHCP server.

2. Topology

Figure 1-18 Configuring DHCP Relay



3. Notes

- Configure Device A:
  - Configure the IP address of the interface.
  - Enable the DHCP relay function.
- Configure Device B:
  - Configure the IP address of the interface.
  - Configure a route from Device B to GigabitEthernet 0/1 on Device A.
  - Enable the DHCP server service and configure address pool parameters.

4. Procedure

(1) Configure Device A:

Configure the IP address of the interface.

```
DeviceA> privilege
DeviceA# config-system terminal
```

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.1.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# ip address 172.2.2.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

Enable the DHCP relay function.

```
DeviceA(cfg)# service dhcp-server
```

Add the DHCP server address to the DHCP relay agent.

```
DeviceA(cfg)# ip dhcp-server-address 172.2.2.1
```

(2) Configure Device B:

Configure the IP address of the interface.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 172.2.2.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Configure a static route to network segment 192.1.1.0/24.

```
DeviceB(cfg)# ip route-config 192.1.1.0 255.255.255.0 gigabitethernet 0/1
```

Enable the DHCP server function.

```
DeviceB(cfg)# service dhcp-server
```

Configure network parameters of pool 1.

```
DeviceB(cfg)# ip dhcp-server pool pool1
DeviceB(cfg-dhcp-pool)# network 192.1.1.0 255.255.255.0
DeviceB(cfg-dhcp-pool)# default-router 192.1.1.1
DeviceB(cfg-dhcp-pool)# dns-server 192.1.1.2
DeviceB(cfg-dhcp-pool)# exit
```

(3) Enable DHCP on the host to obtain an IP address. (Omitted)

5. Verification

Check that the host has obtained an address in network segment 192.1.1.0/24.

Run the **display ip dhcp-server pool** command on Device B to display the address pool configurations and usage.

```
DeviceB(cfg)# display ip  dhcp-server pool
Pool  name          Total       Distributed   Remained     Percentage
------------  -----------  -----------  --------  -----------
pool1             126              1                125            0.99206
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
```

```
 interface gigabitethernet 0/1
 ip address 192.1.1.1 255.255.255.0
!
 interface gigabitethernet 0/2
!
 service   dhcp-server
 ip   dhcp-server-address 172.2.2.1
!
```

- Device B configuration file

```
hostname DeviceB
!
 interface gigabitethernet 0/1
 ip address 172.2.2.1 255.255.255.0
!
 ip route-config 192.1.1.0 255.255.255.0 gigabitethernet 0/1
 service   dhcp-server
!
 ip   dhcp-server pool pool1
 network 192.1.1.0 255.255.255.0
 default-router 192.1.1.1
 dns-server 192.1.1.2
!
```

7. Common Errors

- The DHCP relay function is disabled on the DHCP relay agent.
- No route is configured between the DHCP relay agent and DHCP server.
- The IP address of the DHCP server is configured on the DHCP relay agent.

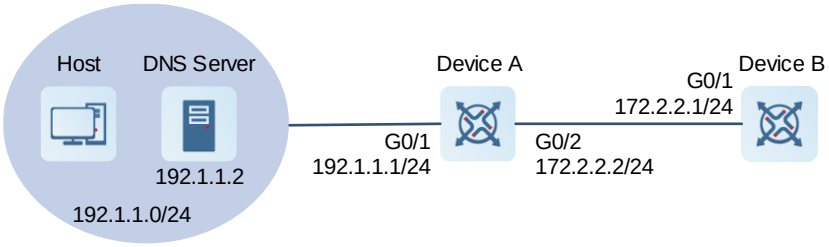
1.6.4 Assigning IP Addresses Based on Class Rules

1. Requirements

As shown in Figure 1-19, Device A serves as a DHCP server to assign IP addresses and other network configurations to clients, and Device B serves as a DHCP relay agent to forward DHCP packets between the clients and server. It is required that Device A assign an IP address in the range of 192.1.1.200 to 192.1.1.254 if a DHCP request packet carries Option 82 and assigns an IP address in the range of 192.1.1.1 to 192.1.1.199 if a DHCP request packet does not carry Option 82.

2. Topology

Figure 1-19 Configuring IP Address Assignment based on Class Rules



3. Notes

- Configure Device A:
  - Configure the IP address of the interface.
  - Enable the DHCP relay function.
  - Enable the Option 82 function.
- Configure Device B:
  - Configure the IP address of the interface.
  - Configure a route from Device B to GigabitEthernet 0/1 on Device A.
  - Enable the DHCP server service and configure address pool parameters.
  - Add a specific class rule.

4. Procedure

(1) Configure Device A:

Configure the IP address of the interface.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.1.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# ip address 172.2.2.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

Enable the DHCP relay function.

```
DeviceA(cfg)# service dhcp-server
```

Add the DHCP server address.

```
DeviceA(cfg)# ip dhcp-server-address 172.2.2.1
```

Enable Option 82.

```
DeviceA(cfg)# ip dhcp-server relay-agent class option82
```

(2) Configure Device B:

Configure the IP address of the interface.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 172.2.2.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Configure a static route to network segment 192.1.1.0/24.

```
DeviceB(cfg)# ip route-config 192.1.1.0 255.255.255.0 gigabitethernet 0/1
```

Enable the DHCP Server function.

```
DeviceB(cfg)# service dhcp-server
```

Configure a class rule.

```
DeviceB(cfg)# ip dhcp-server class-config myclass
DeviceB(cfg-dhcp-class)# relay-agent agent class
DeviceB(cfg-dhcp-class-relayinfo)#    relay-agent-info hex 060223*
```

Create an address pool and configure related network parameters.

```
DeviceB(cfg)# ip dhcp-server pool pool1
DeviceB(cfg-dhcp-pool)# network 192.1.1.0 255.255.255.0
DeviceB(cfg-dhcp-pool)# default-router 192.1.1.1
DeviceB(cfg-dhcp-pool)# dns-server 192.1.1.2
```

Configure the network segment for assigning an IP address when a class is matched.

```
DeviceB(cfg-dhcp-pool)# class-config myclass
DeviceB(cfg-dhcp-pool-class)# address scope 192.1.1.200 192.1.1.254
DeviceB(cfg-dhcp-pool-class)# exit
```

(3) Enable DHCP on the host to obtain an IP address. (Omitted)

5. Verification

Capture packets on Device B and check that the clients that match class rules obtain IP addresses in the range of 192.1.1.200 to 192.1.1.254 and other clients obtain IP addresses in the range of 192.1.1.1 to 192.1.1.199.

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
 service    dhcp-server
 ip    dhcp-server-address 172.2.2.1
 ip    dhcp-server relay-agent class option82
!
 interface gigabitethernet 0/1
 ip address 192.1.1.1 255.255.255.0
 exit
 interface gigabitethernet 0/2
 ip address 172.2.2.2 255.255.255.0
!
```

- Device B configuration file

```
hostname DeviceB
!
 ip route-config 192.1.1.0 255.255.255.0 gigabitethernet 0/1
 service    dhcp-server
!
 interface gigabitethernet 0/1
 ip address 172.2.2.1 255.255.255.0
!
 ip    dhcp-server class-config myclass
 relay-agent agent class
   relay-agent-info hex 060223*
!
 ip    dhcp-server pool pool1
 network 192.1.1.0 255.255.255.0
 default-router 192.1.1.1
 dns-server 192.1.1.2
 class-config myclass
     address scope 192.1.1.200 192.1.1.254
!
```

7. Common Errors

- The DHCP Relay function is disabled.
- No route is configured between the DHCP relay agent and DHCP server.
- No DHCP server IP address is configured.

Contents

1 Configuring DHCP Client 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.1.3 Protocols and Standards 2

1.2 Enabling DHCP Client Function 2

1.2.1 Overview 2

1.2.2 Restrictions and Guidelines 2

1.2.3 Procedure 2

1.3 Monitoring 3

1.4 Configuration Examples 3

1.4.1 Configuring DHCP Client 3

1 Configuring DHCP Client

1.1 Introduction

1.1.1 Overview

Dynamic Host Configuration Protocol (DHCP) is a local area network (LAN) protocol based on the User Datagram Protocol (UDP) for dynamically assigning reusable network resources over port 68, for example, IP addresses.

The DHCP works in client/server mode. A DHCP client sends a request message to a DHCP server to obtain an IP address and other configurations. When a DHCP client and a DHCP server are not in the same subnet, they need a DHCP relay agent to forward DHCP request and reply messages.

1.1.2 Principles

1. DHCP client

A DHCP client enables a device to automatically obtain an IP address and other configurations from a DHCP server.

For details about interaction between the DHCP client and server, see *Configuring DHCP*.

2. Lease renewal

Generally, a dynamic IP address assigned from a DHCP server to a DHCP client has a lease period (one day by default). After the lease period expires, the DHCP server will reclaim this IP address. If the DHCP client wants to continue to use this IP address, it automatically sends a DHCP REQUEST packet to the DHCP server at half of the lease period to renew the lease. If this IP address is valid, the DHCP server returns a DHCP ACK packet to inform the DHCP client that it has obtained a new IP address lease.

If the lease renewal operation performed at half of the lease period fails (for example, the DHCP server does not make a response), the DHCP client broadcasts another DHCP REQUEST packet at 7/8 of the lease period to renew the lease. The DHCP server renews the lease in the same way.

3. Network reconnection

When a DHCP client reconnects to the network after it goes offline, it does not need to send a DHCP DISCOVER packet but only broadcasts a DHCP REQUEST packet that carries the previously assigned IP address.

After receiving the DHCP REQUEST packet, the DHCP server returns a DHCP ACK packet to inform the DHCP client that it can continue to use the original IP address if this IP address is not assigned. If the original IP address has been assigned to another client, the DHCP server returns a DHCP NAK packet. Upon receiving the DHCP NAK packet, the DHCP client needs to re-send a DHCP DISCOVER packet to apply for a new IP address.

4. Lease release

When a DHCP client no longer needs to use the assigned IP address, it can send a DHCP RELEASE packet to inform the DHCP server, for the DHCP server to release the DHCP lease and reclaim the IP address. The

DHCP server saves the configuration information of the DHCP client temporarily for reuse in next IP address application by the DHCP client.

1.1.3 Protocols and Standards

- RFC 2131: Dynamic Host Configuration Protocol
- RFC 2132: DHCP Options and BOOTP Vendor Extensions
- RFC 3046: DHCP Relay Agent Information Option
- RFC 1542: Clarifications and Extensions for the Bootstrap Protocol

1.2 Enabling DHCP Client Function

1.2.1 Overview

After the DHCP Client function is enabled on a device, the device can dynamically obtain IP addresses and other configurations, such as the default gateway and Domain Name System (DNS) server.

1.2.2 Restrictions and Guidelines

The device supports DHCP client configurations on Ethernet, Frame Relay (FR), Point-to-Point Protocol (PPP), and High-Level Data Link Control (HDLC) interfaces.

The SN specified by the **ip dhcp client dhcp-client-identifier type** command indicates the SN of the chassis where the active supervisor module resides. When the active/standby switchover occurs, the SN may change.

1.2.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Enable DHCP client.

ip dhcp client dhcp-client-identifier type { **1** | **2** }

The default type of DHCP option 61 is type 1.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the switch virtual interface (SVI) configuration mode.

  interface vlan *interface-number*

(5) Enable DHCP client on an interface.

ip address from-dhcp

No interface can obtain an IP address through DHCP by default.

1.3 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** commands to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 Monitoring

Command	Purpose
display from-dhcp lease-time	Displays lease information of a DHCP client.
debugging ip from-dhcp client	Debugs DHCP packets.

1.4 Configuration Examples

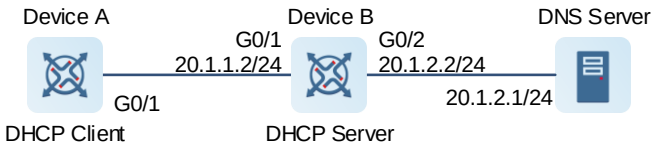
1.4.1 Configuring DHCP Client

1. Requirements

As shown in Figure 1-1, GigabitEthernet 0/1 on Device A uses DHCP to obtain an IP address and DNS server address 20.1.1.1/24 from Device B (DHCP server).

2. Topology

Figure 1-1 Configuring DHCP Client



3. Notes

- Enable the DHCP Client function on GigabitEthernet 0/1 of Device A.
- Enable the DHCP Server function on Device B and configure address pool parameters as required.

4. Procedure

(1) Configure Device A:

Enable the DHCP Client service on GigabitEthernet 0/1.

```
DeviceA> privilege
DeviceA# config-system terminal
```

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address from-dhcp
```

(2) Configure Device B:

Enable the DHCP Server service.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# service from-dhcp
```

Configure the interface IP address used to connect to the DHCP client.

```
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 20.1.1.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Configure the interface IP address used to connect to the DNS server.

```
DeviceB(cfg)# interface gigabitethernet 0/2
DeviceB(cfg-if-GigabitEthernet 0/2)# ip address 20.1.2.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/2)# exit
```

Configure a dynamic address pool.

```
DeviceB(cfg)# ip from-dhcp pool User
DeviceB(cfg-dhcp-pool)# network 20.1.1.0 255.255.255.0
DeviceB(cfg-dhcp-pool)# dns-server 20.1.2.1
DeviceB(cfg-dhcp-pool)# exit
```

Configure excluded IP addresses.

```
DeviceB(cfg)# ip from-dhcp excluded-address 20.1.1.2
```

(3) Configure the DNS server:

Set the DNS server IP address to 20.1.2.1/24. (Omitted)

5. Verification

Run the **display from-dhcp lease-time** command on Device B to display the IP address information obtained from an interface.

```
DeviceB# display from-dhcp lease-time
Temp IP addr: 20.1.1.1 for peer on Interface: GigabitEthernet 0/1
Temp sub net mask: 255.255.255.0
    DHCP Lease server: 20.1.1.2, state: 7 Bound
    DHCP transaction id: ca811331
    Lease: 14400 secs, Renewal: 7200 secs, Rebind: 12600 secs
    Next timer fires after: 4916 secs
    Retry count: 0 Client-ID: 0158696cc59347
```

Run the **display hosts** command on Device A to display DNS server information obtained from an interface.

```
DeviceA# display hosts-config
Name servers are:
20.1.2.1 dynamic from DHCP
Host                                    type    Address                 TTL(sec)
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
 interface-status gigabitethernet 0/1
 ip address from-dhcp
!
```

- Device B configuration file

```
hostname DeviceB
!
ip dhcp service
ip from-dhcp excluded-ip-address 20.1.1.2
!
 interface-status gigabitethernet 0/1
 ip address 20.1.1.2 255.255.255.0
!
 interface-status gigabitethernet 0/2
 ip address 20.1.2.2 255.255.255.0
!
 ip from-dhcp pool User
 network 20.1.1.0 255.255.255.0
 dns-server 20.1.2.1
!
```

Contents

1 Configuring DHCP Snooping 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.1.3 Applications 3

1.1.4 Protocols and Standards 7

1.2 Restrictions and Guidelines 7

1.3 Configuration Task Summary 7

1.4 Configuring Basic DHCP Snooping Functions 8

1.4.1 Overview 8

1.4.2 Procedure 8

1.5 Disabling DHCP Snooping on a VLAN 9

1.5.1 Overview 9

1.5.2 Procedure 9

1.6 Enabling Source MAC Address Verification 9

1.6.1 Overview 9

1.6.2 Procedure 9

1.7 Writing Dynamic User Information in the DHCP Snooping Binding Database to the Flash Memory in Real Time 10

1.7.1 Overview 10

1.7.2 Procedure 10

1.8 Writing Dynamic User Information in the DHCP Snooping Binding Database to a Flash Memory at a Scheduled Time 10

1.8.1 Overview 10

1.8.2 Procedure 10

1.9 Writing Dynamic User Information in the DHCP Snooping Binding Database to a Specified Backup File at a Scheduled Time 11

1.9.1 Overview 11

1.9.2 Procedure 11

1.10 Enabling DHCP Snooping to Support BOOTP Binding 11

1.10.1 Overview 11

1.10.2 Procedure 11

1.11 Enabling DHCP Snooping to Support Relay Request Packet Processing 11

1.11.1 Overview 11

1.11.2 Procedure 12

1.12 Enabling DHCP Snooping to Support Binding Entry Migration 12

1.12.1 Overview 12

1.12.2 Procedure 12

1.13 Enabling DHCP Snooping to Fast Age Client Migration Entries 12

1.13.1 Overview 12

1.13.2 Procedure 12

1.14 Configuring an Interface in the Suppression State 13

1.14.1 Overview 13

1.14.2 Procedure 13

1.15 Configuring the Maximum Number of Users Bound to a VLAN 13

1.15.1 Overview 13

1.15.2 Procedure 13

1.16 Configuring Option 82 14

1.16.1 Overview 14
1.16.2 Restrictions and Guidelines 14
1.16.3 Procedure 14
1.17 Enabling DHCP Snooping Monitoring 15
1.17.1 Overview 15
1.17.2 Restrictions and Guidelines 15
1.17.3 Procedure 15
1.18 Monitoring 16
1.19 Configuration Examples 16
1.19.1 Configuring DHCP Snooping 16

1 Configuring DHCP Snooping

Note

DHCP snooping is not supported by only the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1.1 Introduction

1.1.1 Overview

The Dynamic Host Configuration Protocol (DHCP) snooping function allows a device to snoop DHCP packets exchanged between clients and a server to record and monitor the IP address usage and filter out invalid DHCP packets, including request packets from the clients and response packets from the server. User data entries generated from DHCP Snooping records may serve security applications such as IP Source Guard.

1.1.2 Principles

1. Basic Concepts

- DHCP request packets

  Request packets are sent from a DHCP client to a DHCP server, including DHCP DISCOVER, DHCP REQUEST, DHCP DECLINE, DHCP RELEASE, and DHCP INFORM packets.

- DHCP response packets

  Response packets are sent from a DHCP server to a DHCP client, including DHCP OFFER, DHCP ACK, and DHCP NAK packets.

- VLAN-based DHCP Snooping

  DHCP Snooping is enabled on a per-VLAN basis. When DHCP Snooping is enabled on a device, it is effective to all VLANs of the device by default. You can flexibly configure the VLANs on which DHCP Snooping is enabled.

- DHCP Snooping trusted ports

  Packet for obtaining IP addresses through DHCP are exchanged via broadcast. Rough DHCP services affect clients' normal acquisition of IP addresses and lead to user information spoofing and stealing. DHCP Snooping ports are classified into trusted ports and untrusted ports to prevent rough DHCP services.

  The device only transmits DHCP response packets received on trusted ports and discards those received from untrusted ports. In this way, the ports connected to legitimate DHCP servers are configured as trusted ports and the other ports are configured as untrusted ports to shield rough DHCP servers.

  On switches, all switching ports or L2 aggregation ports (APs) are defaulted as untrusted, and trusted ports can be specified.

- DHCP Snooping binding database

  In a DHCP network, clients may set static IP addresses randomly. This increases not only the difficulty of network maintenance but also the possibility that authorized clients with IP addresses assigned by a DHCP server may fail to use the network due to address conflicts. By snooping packets between the clients and

server, DHCP Snooping summarizes user entries, including IP addresses, media access control (MAC) addresses, VLAN IDs (VIDs), ports, and lease time to build the DHCP Snooping binding database. Combined with Address Resolution Protocol (ARP) detection or ARP check, DHCP Snooping controls the reliable assignment of IP addresses to legitimate clients.

- DHCP Snooping packet suppression

  To shield all the DHCP packets on a specific client, enable DHCP Snooping packet suppression on its untrusted ports.

- DHCP Snooping rate limit

  DHCP Snooping rate limit can be configured using the rate limit command of Network Foundation Protection Policy (NFPP). For NFPP configuration, see *Configuring NFPP*.

- DHCP Option 82

  DHCP Option 82, also called the DHCP relay agent information option, is an option in DHCP packets. Option 82 is used to enhance the DHCP server security and optimize the IP address assignment policy. This option is usually configured on the DHCP relay service component, such as DHCP Relay and DHCP Snooping of a network access device. This option is transparent to clients and added or removed by the DHCP relay component.

- Invalid DHCP packets

  The device with DHCP Snooping enabled checks validity of received DHCP packets, discards invalid DHCP packets, and records information about legitimate users to form the DHCP Snooping binding database for query by other applications (for example, ARP detection). The following types of packets are considered as invalid DHCP packets:

  - DHCP response packets received on untrusted ports, including DHCP ACK, DHCP NAK, and DHCP OFFER packets
  - DHCP request packets carrying the **giaddr** field (gateway information) received on untrusted ports
  - Packets with the source MAC addresses different from the value of the **chaddr** field in DHCP packets in a case with MAC address verification enabled
  - DHCP RELEASE packets corresponding to entries in the DHCP Snooping binding database, with the untrusted port on which the packets are received being inconsistent with the untrusted port in the binding table
  - DHCP packets in incorrect formats or incomplete DHCP packets

2. Filtering Invalid DHCP Packets

The device with DHCP Snooping enabled checks validity of DHCP packets from untrusted ports to control the transmission scope of the packets and prevent spoofing.

During snooping, it checks the receiving ports and fields of packets to filter packets, and modifies the destination ports of the packets to control the packet transmission scope.

- Checking ports

  When receiving a DHCP packet, the device first checks whether the port receiving the packet is a DHCP trusted port. If yes, the device directly forwards the packet without checking the validity or generating a binding record. If not, the device checks its validity.

- Checking packet encapsulation and length

The device checks whether received packets are UDP packets, whether the destination port is port 67 or 68, and whether the packet length matches the length field defined in protocols.

- Checking packet fields and types

  According to the invalid DHCP packet types introduced in the "Basic Concepts" section, the device first checks the **giaddr** and **chaddr** fields in packets and then checks whether the restrictive conditions for this packet type are met.

3. Building a DHCP Snooping Binding Database

The device with DHCP Snooping enabled detects packets exchanged between the DHCP clients and the DHCP server and generates entries in the DHCP Snooping binding database based on information about valid DHCP packets. All these entries are used as the information table of legitimate users, and provided to other security modules of the device as the basis for filtering network packets.

The binding database is updated during snooping based on the types of DHCP packets.

- Generating binding entries

  When a DHCP ACK packet is snooped on a trusted port, the device with DHCP Snooping enabled extracts the client's IP address, MAC address, and lease time from the packet and generates a binding entry based on the recorded port ID (interface index) and VLAN ID of the client.

- Deleting binding entries

  A binding entry is deleted in the following scenarios:

  - The recorded lease time expires.
  - A valid DHCP RELEASE or DHCP DECLINE packet sent from the client is snooped.
  - A NAK packet is received from a trusted port.
  - A user runs the **cleanup** command to delete a binding entry.

1.1.3 Applications

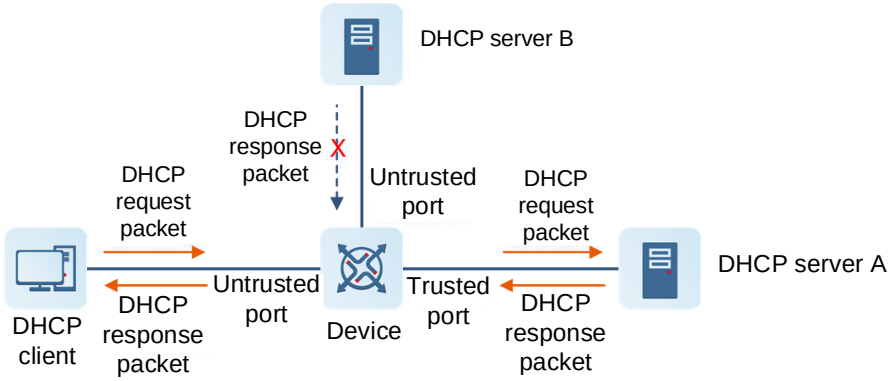
1. Guarding Against DHCP Spoofing

Multiple DHCP servers may exist in a network. It is essential to ensure that PCs can obtain network configurations only from the DHCP servers within a controlled area.

As shown in the following figure, the DHCP client can communicate with a trusted DHCP server only.

- Request packets from the DHCP client are transmitted only to the trusted DHCP server.
- Only response packets from the trusted DHCP server can be transmitted to the DHCP client.

Figure 1-1 Guarding Against DHCP Spoofing



Device is an access device, DHCP Server A is in the controlled area, and DHCP Server B is beyond the controlled area.

- DHCP Snooping is enabled on Device to monitor DHCP packets.
- The port on Device for connecting to DHCP Server A is configured as a DHCP trusted port to forward response packets.
- The rest of ports on Device are configured as DHCP untrusted ports to filter response packets.

2. Guarding Against DHCP Packet Flooding

Malicious DHCP clients in a network may send DHCP packets at a high rate. As a result, legitimate users cannot obtain IP addresses, and access devices are heavily loaded or even break down.

- To ensure network stability, the DHCP Snooping rate limit function needs to be enabled for DHCP packets. The request packets from DHCP clients are sent at a rate below the limit. Packets sent at rates beyond the limit are discarded. DHCP Snooping is correlated with the ARP module to delete entries of non-existing users.

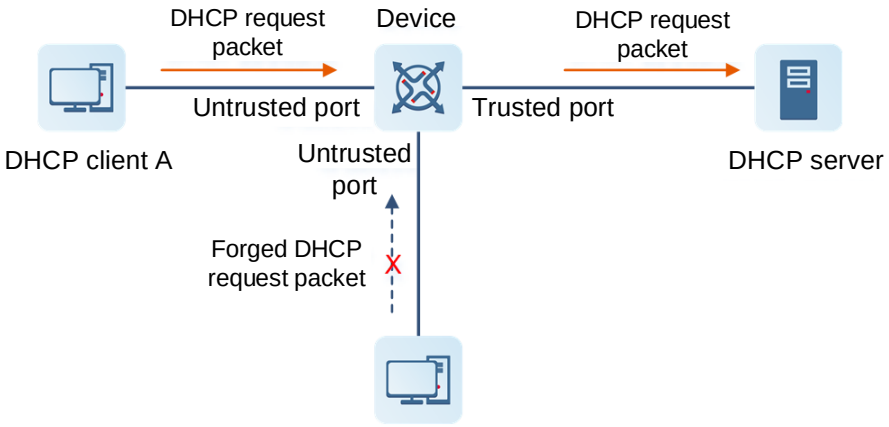
3. Guarding Against Forged DHCP Packets

Malicious DHCP clients in a network may forge DHCP request packets, which consumes applicable IP addresses from the servers and probably preempts legitimate users' IP addresses. Therefore, it is necessary to filter out invalid DHCP packets.

As shown in the following figure, the DHCP request packets sent from DHCP clients are checked.

- The source MAC address field in request packets from DHCP clients must match the client hardware address field of the DHCP packets.
- The RELEASE and DECLINE packets from clients must match the entries in the DHCPv6 Snooping binding database.

Figure 1-2 Guarding Against Forged DHCP Packets



Device is an access device, and DHCP Server is in the controlled area.

- DHCP Snooping is enabled on Device to monitor DHCP packets.
- The port on Device for connecting to DHCP Server is configured as a DHCP trusted port to forward response packets.
- The rest of ports on Device are configured as DHCP untrusted ports to filter response packets.
- DHCP source MAC addresses need to be verified on untrusted ports of Device to filter out invalid DHCP packets.

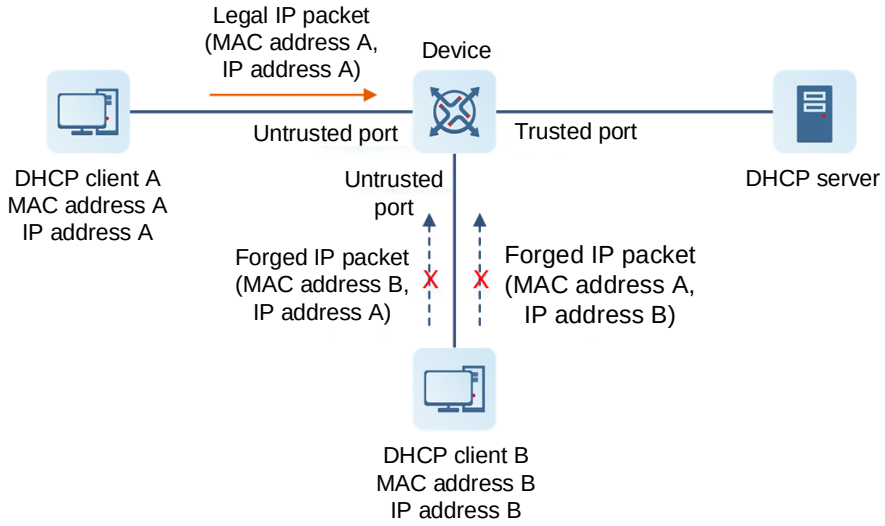
4. Guarding Against IP/MAC Address Spoofing

You can check IP packets from untrusted ports based the IP address field or IP address and MAC address fields to filter out forged IP packets.

As shown in the following figure, IP packets sent from DHCP clients are checked.

- The source address field in IP packets must match the IP address assigned by the DHCP server.
- The source MAC address field in L2 packets must match the client hardware address field in DHCP request packets from clients.

Figure 1-3 Guarding Against IP/MAC Address Spoofing



Device is an access device, and DHCP Server is in the controlled area.

- DHCP Snooping is enabled on Device to monitor DHCP packets.
- The port on Device for connecting to DHCP Server is configured as a DHCP trusted port.
- All downlink ports on Device are configured as DHCP untrusted ports.
- IP Source Guard is enabled on Device to filter IP packets.
- The IP Source Guard matching mode is set to IP+MAC on Device to check the MAC address and IP address fields in IP packets.

5. Preventing Private IP Addresses

The device with DHCP Snooping enabled checks whether the source addresses in IP packets from untrusted ports are consistent with the IP addresses assigned by the DHCP server.

If the source addresses, connected ports, and L2 source MAC addresses in IP packets do not match the assignments of the DHCP server, such packets are discarded.

The working process of the device in this scenario is the same as that in “Guarding Against IP/MAC Address Spoofing".

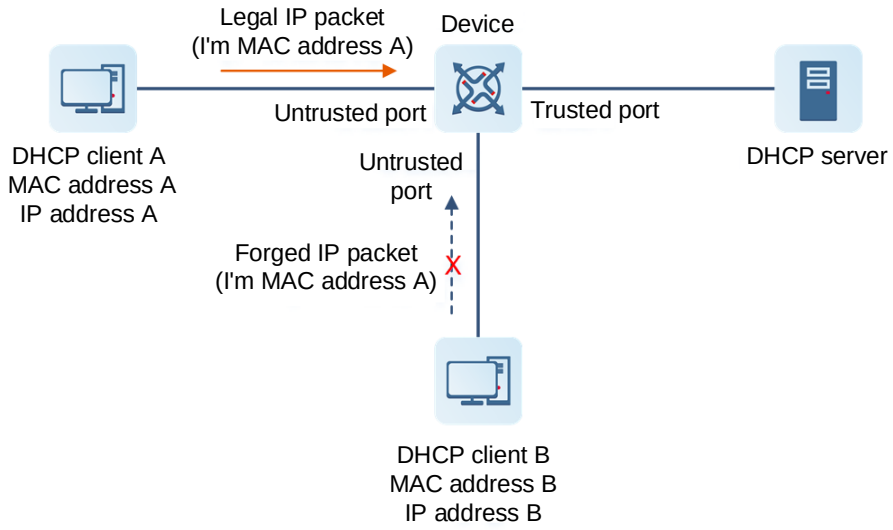
6. Detecting ARP Attacks

The device with DHCP Snooping enabled checks the ARP packets from untrusted ports and filter out the ARP packets unmatched with the assignments of the DHCP server.

As shown in the following figure, ARP packets sent from DHCP clients are checked.

The ports receiving ARP packets, the L2 MAC addresses, and the hardware addresses of ARP packet senders must be consistent with the snooped DHCP packet records.

Figure 1-4 Detecting ARP Attacks



Device is an access device, and DHCP Server is in the controlled area.

- DHCP Snooping is enabled on Device to monitor DHCP packets.
- The port on Device for connecting to DHCP Server is configured as a DHCP trusted port.
- All downlink ports on Device are configured as DHCP untrusted ports.
- IP Source Guard and ARP check are enabled on all untrusted ports on Device to filter ARP packets.

1.1.4 Protocols and Standards

- RFC 2131: Dynamic Host Configuration Protocol
- RFC 2132: DHCP Options and BOOTP Vendor Extensions

1.2 Restrictions and Guidelines

- All security control functions are effective only to DHCP untrusted ports.
- The ports on the device for connecting to trusted DHCP servers must be configured as DHCP trusted ports.
- DHCP Snooping takes effect to switching ports, L2 APs, and L2 encapsulation sub-interfaces.
- DHCP Snooping and DHCP Relay are mutually exclusive in virtual routing and forwarding (VRF) scenarios.
- A device uses the following rules to process the received DHCP packets:
  - If all interfaces on the device are in the UNTRUST state, the device will discard all the DHCP request and response packets received by the interfaces.
  - If any interface on the device is configured as a trusted port, the DHCP request packets received by the untrusted port will be forwarded through the trusted port, and the DHCP response packets received by the untrusted port will be discarded.

1.3 Configuration Task Summary

DHCP Snooping configuration includes the following tasks:

The following configuration tasks are mutually exclusive. Please configure only one task.

- Configuring DHCP Snooping
  - a Configuring Basic DHCP Snooping Functions
  - b (Optional) Disabling DHCP Snooping on a VLAN
  - c (Optional) Enabling Source MAC Address Verification
  - d (Optional) Writing Dynamic User Information in the DHCP Snooping Binding Database to the Flash Memory in Real Time
  - e (Optional) Writing Dynamic User Information in the DHCP Snooping Binding Database to a Flash Memory at a Scheduled Time
  - f (Optional) Writing Dynamic User Information in the DHCP Snooping Binding Database to a Specified Backup File at a Scheduled Time
  - g (Optional) Enabling DHCP Snooping to Support BOOTP Binding
  - h (Optional) Enabling DHCP Snooping to Support Relay Request Packet Processing
  - i (Optional) Enabling DHCP Snooping to Support Binding Entry Migration
  - j (Optional) Enabling DHCP Snooping to Fast Age Client Migration Entries
  - k (Optional) Configuring an Interface in the Suppression State
  - l (Optional) Configuring the Maximum Number of Users Bound to a VLAN
  - m (Optional) Configuring Option 82
- Enabling DHCP Snooping Monitoring

1.4 Configuring Basic DHCP Snooping Functions

1.4.1 Overview

This section describes how to enable basic DHCP Snooping functions to copy DHCP packets, generate binding entries based on interaction, check packet validity to filter out unauthorized DHCP packets and control the transmission scope of DHCP packets.

1.4.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable DHCP Snooping globally.

ip dhcp-snooping service

DHCP Snooping is disabled globally by default.

(4) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the Layer 2 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(5) Configure an interface as a DHCP Snooping trusted port.

ip dhcp-snooping service trust

All interfaces are DHCP Snooping untrusted ports by default.

Run this command to configure the interface connected to an authorized DHCP server as a trusted port. DHCP Response packets received on trusted ports are forwarded normally, while those received on untrusted ports are discarded.

The command is only configured and takes effect on Layer 2 Ethernet interfaces, Layer 2 link aggregation interfaces, Layer 2 Ethernet subinterfaces and Layer 2 link aggregation subinterfaces.

1.5 Disabling DHCP Snooping on a VLAN

1.5.1 Overview

DHCP Snooping is enabled on a per-VLAN basis. You can flexibly configure the VLANs on which DHCP Snooping is enabled.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable DHCP Snooping on a VLAN.

unset ip dhcp-snooping service vlan { *vlan-range* | { *vlan-min* [*vlan-max*] } }

After DHCP Snooping is enabled globally, it takes effect to all VLANs by default.

1.6 Enabling Source MAC Address Verification

1.6.1 Overview

After source MAC address verification is enabled, the MAC addresses in link layer headers and the **CLIENT MAC** fields in DHCP request packets from untrusted ports are checked for consistence. If the verification fails, packets are discarded.

1.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable source MAC address verification.

ip dhcp-snooping service verify mac-address

Source MAC address verification is disabled by default.

1.7 Writing Dynamic User Information in the DHCP Snooping Binding Database to the Flash Memory in Real Time

1.7.1 Overview

You can write dynamic user information in the DHCP Snooping binding database to the flash memory in real time.

1.7.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Write dynamic user information in the DHCP Snooping binding database to a flash memory in real time.

ip dhcp-snooping service database write-to-flash

1.8 Writing Dynamic User Information in the DHCP Snooping Binding Database to a Flash Memory at a Scheduled Time

1.8.1 Overview

You can write dynamic user information in the DHCP Snooping binding database to the flash memory at a scheduled time.

1.8.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Write dynamic user information in the DHCP Snooping binding database to a flash memory at a scheduled time.

ip dhcp-snooping servicedatabase write-interval *time*

The function of writing all dynamic user information in the DHCP Snooping binding database to a flash memory at a scheduled time is not configured by default.

1.9 Writing Dynamic User Information in the DHCP Snooping Binding Database to a Specified Backup File at a Scheduled Time

1.9.1 Overview

You can write dynamic user information in the DHCP Snooping binding database to a specified backup file at a scheduled time.

1.9.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Write dynamic user information in the DHCP Snooping binding database to a specified backup file at a scheduled time.

ip dhcp-snooping service database sata0 [**interval** *time*]

The function of writing all dynamic user information in the DHCP Snooping binding database to a specified backup file at a scheduled time is not configured by default.

1.10 Enabling DHCP Snooping to Support BOOTP Binding

1.10.1 Overview

By default, the DHCP snooping feature only forwards BOOTP packets. After this feature is enabled, the DHCP snooping feature snoops BOOTP packets. After a BOOTP client successfully applies for an IP address, the DHCP Snooping feature adds the BOOTP client to the static binding database.

1.10.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable DHCP Snooping to support Bootstrap Protocol (BOOTP) binding.

ip dhcp-snooping service bootp-bind

DHCP Snooping does not support BOOTP binding by default.

1.11 Enabling DHCP Snooping to Support Relay Request Packet Processing

1.11.1 Overview

You can enable DHCP Snooping to support relay request packet processing.

1.11.2 Procedure

(1) Enter the privileged EXEC mode.

peivilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable DHCP Snooping to support relay request packet processing.

ip dhcp-snooping service check-giaddr

DHCP Snooping does not support relay request packet processing by default.

After this function is enabled, services (IP Source Guard and Dot1x authentication) using DHCP Snooping binding entries generated based on relay requests cannot be deployed. Otherwise, users fail to access the Internet.

After this function is enabled, the **ip dhcp-snooping service verify mac-address** command cannot be configured. Otherwise, DHCP request packets of the relay are discarded, and users fail to obtain addresses.

1.12 Enabling DHCP Snooping to Support Binding Entry Migration

1.12.1 Overview

You can enable DHCP Snooping to support binding entry migration.

1.12.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable DHCP Snooping to support binding entry migration.

ip dhcp-snooping service station-move permit

DHCP Snooping does not support binding entry migration by default.

1.13 Enabling DHCP Snooping to Fast Age Client Migration Entries

1.13.1 Overview

You can enable DHCP Snooping to fast age client migration entries.

1.13.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable DHCP Snooping to fast age client migration entries.

ip dhcp-snooping service station-move aging

Fast aging of client migration entries is enabled for DHCP Snooping by default.

1.14 Configuring an Interface in the Suppression State

1.14.1 Overview

You can configure an interface in the suppression state to discard all DHCP packets sent to the interface.

1.14.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 2 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Configure an interface in the suppression state so as to discard all DHCP packets sent to the interface.

ip dhcp-snooping service suppression

No interface is configured in the suppression state by default.

1.15 Configuring the Maximum Number of Users Bound to a VLAN

1.15.1 Overview

You can configure the maximum number of users bound to a VLAN.

1.15.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 2 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Configure the maximum number of users bound to a VLAN.

ip dhcp-**snooping service vlan** *vlan-range* **max-user** *user-number*

The maximum number of users bound to a VLAN is not configured by default.

1.16 Configuring Option 82

1.16.1 Overview

Option 82 includes **Circuit ID** (sub-option 1) and **Remote ID** (sub-option 2). **Circuit ID** identifies the VLAN ID and interface information of a client. **Remote ID** identifies the access device of a client, which is the MAC address of the device generally.

A DHCP server assigns IP addresses based on Option 82 carried in DHCP request packets. Option 82 is transparent to clients.

1.16.2 Restrictions and Guidelines

Option 82 and DHCP Relay are mutually exclusive.

1.16.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Add Option 82 to DHCP request packets.

ip dhcp-**snooping service option format** [**standard-format** | **user-defined**]

Option 82 is not added to DHCP request packets by default.

(4) (Optional) Set **Remote ID** to a user-defined string or the host name.

ip dhcp-**snooping service option format remote-id format info** { **hostname** | **string** *ascii-string* }

Remote ID in Option 82 is not set to a user-defined string or host name by default.

(5) (Optional) Customize **Remote ID** in Option 82.

ip dhcp-**snooping service option format user-defined remote-id** *text*

Customization of **Remote ID** in DHCP Option 82 is disabled by default.

(6) (Optional) Customize **Circuit ID** in Option 82.

ip dhcp-**snooping service option format user-defined circuit-id** *text*

Customization of **Circuit ID** in DHCP Option 82 is disabled by default.

(7) (Optional) Configure the format of the MAC address string in the sub-option of DHCP Option 82.

ip dhcp-**snooping service option format user-defined mac-str-format** *type*

The user-defined MAC address string is in **H.H.H** format by default.

(8) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 2 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(9) (Optional) Configure the VLAN in **Circuit ID** as the specified VLAN.

ip dhcp-**snooping service vlan** *vlan-id* **option format change-vlan-to vlan** *vlan-id*

When Option 82 is in extended mode, the VLAN in **Circuit ID** is not configured as the specified VLAN by default.

(10) (Optional) Set **the Circuit ID** to user-defined content.

ip dhcp-**snooping service vlan** *vlan-id* **option format info circuit-id string** *ascii-string*

When Option 82 is in extended mode, **Circuit ID** is not set to user-defined content for forwarding by default.

1.17 Enabling DHCP Snooping Monitoring

1.17.1 Overview

After the DHCP Snooping monitoring function is enabled, DHCP Snooping only copies DHCP packets and generates binding entries based on the interaction status, but does not check the validity of the packets.

1.17.2 Restrictions and Guidelines

The DHCP Snooping monitoring and DHCP Snooping functions are mutually exclusive.

1.17.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable DHCP Snooping monitoring globally.

ip dhcp-snooping service monitor

DHCP Snooping monitoring is disabled globally by default.

1.18 Monitoring

This section describes the **display** commands used for checking the running status of a configured function to verify the configuration effect.

Run the **debugging** command to display the output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** command to clear the information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
display ip dhcp-server snooping	Displays DHCP Snooping configurations.
display ip dhcp-server snooping binding-database	Displays user information in the DHCP Snooping database.
display ip dhcp-server snooping packet	Displays packet statistics of DHCP Snooping.
cleanup ip dhcp-server snooping binding-database [*ipv4-address*] [*mac-address*] [**vlan** *vlan-id*] [**interface** *interface-type interface-number*]	Clears all dynamic users in the DHCP Snooping database.
debugging snooping ipv4 event	Debugs the DHCP Snooping function based on the MAC address.
debugging snooping ipv4 mac-address *H.H.H*	Debugs all DHCP Snooping functions.

1.19 Configuration Examples

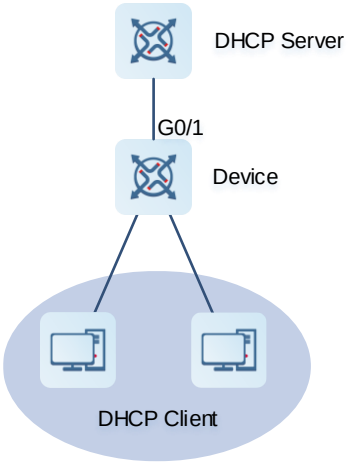
1.19.1 Configuring DHCP Snooping

1. Requirements

DHCP clients can obtain IP addresses from legitimate DHCP servers dynamically.

2. Topology

Figure 1-5 Configuring DHCP Snooping



3. Notes

- Enable DHCP Snooping on Device (an access device).
- Configure the uplink interface GigabitEthernet 0/1 as a trusted port.

4. Procedure

Configure Device (an access device).

```
Device> privilege
Device# config-system terminal
Device(cfg)# ip dhcp-snooping service
Device(cfg)# interface gigabitethernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# ip dhcp-server snooping trust
```

5. Verification

Display DHCP Snooping configurations on Device and check whether the trusted port is correct.

```
Device# display ip dhcp-snooping service
Switch DHCP snooping status                    :    ENABLE
DHCP snooping verify hardware address status   :    DISABLE
DHCP snooping database write-delay time        :    0 seconds
DHCP snooping option 82 status                 :    DISABLE
DHCP snooping Support bootp bind status        :    DISABLE
Interface                          Trusted          Rate limit (pps)
-------------------------          -------          ----------------
GigabitEthernet 0/1                YES              unlimited
```

Display entries generated by Device.

```
Device# display ip dhcp-snooping service binding-database
Total number of bindings: 1
NO.    MACADDRESS          IPADDRESS        LEASE(SEC)    TYPE            VLAN    INTERFACE
```

```
1       0013.2049.9014      172.16.1.2      86207       DHCP-Snooping 1     GigabitEthernet 0/1
```

6. Configuration Files

Device configuration file

```
hostname Device
!
ip dhcp-snooping service
!
interface GigabitEthernet 0/1
 ip dhcp-snooping service   trust
!
end
```

7. Common Misconfigurations

- The uplink interface is not configured as a DHCP Snooping trusted port.
- Another access security option is configured for the uplink interface. As a result, the uplink interface fails to be configured as a DHCP Snooping trusted port.

Contents

1 Configuring DNS 1
1.1 Overview 1
1.2 Principles 1
1.2.1 Static Domain Name Resolution 1
1.2.2 Dynamic Domain Name Resolution 1
1.2.3 DNS Proxy 2
1.2.4 Protocols and Standards 2
1.3 Configuration Task Summary 2
1.4 Notes 2
1.5 Configuring Static Domain Name Resolution 3
1.5.1 Overview 3
1.5.2 Procedure 3
1.6 Configuring Dynamic Domain Name Resolution 3
1.6.1 Overview 3
1.6.2 Restrictions and Guidelines 3
1.6.3 Prerequisites 4
1.6.4 Procedure 4
1.7 Configuring DNS Proxy 5
1.7.1 Overview 5
1.7.2 Prerequisites 5
1.7.3 Procedure 5
1.8 Monitoring 6

1.9 Configuration Examples....7
1.9.1 Configuring Static Domain Name Resolution....7
1.9.2 Configuring Dynamic Domain Name Resolution....8
1.9.3 Configuring DNS Proxy....9

1 Configuring DNS

1.1 Overview

Domain Name System (DNS) is important and commonly used in a Transmission Control Protocol/Internet Protocol (TCP/IP) network-advertise environment. The main function of DNS is to convert a domain name that is easy to remember into an IP address that is hard to remember. The process of obtaining an IP address mapped to a domain name or host name is called domain name resolution or host name resolution.

A domain name can be resolved firstly through static domain name resolution. If this fails, dynamic domain name resolution can be used instead. You can put the commonly used domain names into the static domain name resolution table, which improves the efficiency of domain name resolution.

Both the TCP and User Datagram Protocol (UDP) port IDs of DNS are 53, and generally a UDP port is used.

1.2 Principles

1.2.1 Static Domain Name Resolution

Static domain name resolution means that a user presets the mappings between domain names and IP addresses on a device. When you perform domain name operations by using application programs (such as ping and telnet), the system can find the required IP address with no need to connect to the DNS server on the network.

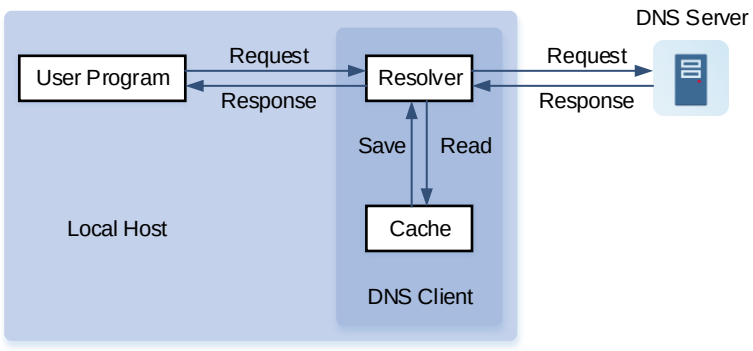
When the application programs access a small number of relatively fixed hosts, the domain name-IP mappings can be configured in advance in the form of static domain name entries. Then, no search request needs to be sent to the DNS server, improving the resolution efficiency.

1.2.2 Dynamic Domain Name Resolution

1. System Structure

Dynamic domain name resolution means that when a user performs domain name operations by using some application programs, the DNS resolver of the system queries an external DNS server for the IP addresses mapped to the domain names.

Figure 1-1 Structure of Dynamic Domain Name Resolution



As shown in Figure 1-1, a resolver is integrated with a cache to form a DNS client, which accepts DNS requests from user programs and makes responses accordingly. In general, user programs (such as ping and traceroute), a cache, and a resolver are on the same host, but a DNS server is on a different host.

2. Resolution Process

The process of dynamic domain name resolution is as follows:

(1) A user application program requests the IP address mapped to a domain name from the DNS resolver of the system.

(2) The DNS resolver queries the dynamic cache at first. If the domain name in the dynamic cache does not expire, the DNS resolver returns the domain name to the application program.

(3) If all domain names expire, the DNS resolver initiates a request for domain name-IP address conversion to the external DNS server.

(4) After receiving a response from the DNS server, the DNS resolver caches and transfers the response to the application program.

3. Cache Function

The dynamic domain name resolution supports the cache function. The cache stores the domain name-IP mappings successfully and dynamically resolved each time. When the same domain name is queried next time, the mapping can be read from the cache, with no need to send a request to the DNS server every time, thereby improving the resolution efficiency.

1.2.3 DNS Proxy

DNS proxy is generally deployed on the front-end egress gateway between a DNS server and a PC, and acts as a proxy of the DNS server to process users' DNS domain name resolution requests. When processing a request packet from the DNS client, the DNS proxy searches for the domain name-IP mapping locally first. If the required domain name-IP mapping is found, the DNS proxy directly returns it to the client. If the required domain name-IP mapping is not found, the DNS resolver searches an external DNS server for the IP address mapped to the domain name and returns a reply packet to the client.

1.2.4 Protocols and Standards

- RFC1034: DOMAIN NAMES - CONCEPTS AND FACILITIES
- RFC1035: DOMAIN NAMES - IMPLEMENTATION AND SPECIFICATION

1.3 Configuration Task Summary

All the following configuration tasks are optional and may be selected as needed.

- Configuring Static Domain Name Resolution
- Configuring Dynamic Domain Name Resolution
- Configuring DNS Proxy

1.4 Notes

This section describes the precautions for configuring the DNS function.

Whether the DNS function is supported depends on devices.

Table 1-1 Products Supporting the DNS Function

Function	Product Supporting the Function
Configuring DNS Proxy Function	S6980-64QC S6580-48CQ8QC S6580-24DC8QC

1.5 Configuring Static Domain Name Resolution

1.5.1 Overview

Static domain name resolution allows a user to preset the mappings between domain names and IP addresses on a device. The system can find the IP address mapped to a domain name from the device without obtaining the required IP address from a DNS server on the network.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the IP address mapped to a domain name.

(IPv4 network)

ip hosts-table *host-name* [*telnet-port*] *ipv4-address*

(IPv6 network)

Configure the IP**v**6 address mapped to a domain name.

ipv6 hosts-table *host-name* [*telnet-port*] *ipv6-address*

1.6 Configuring Dynamic Domain Name Resolution

1.6.1 Overview

Dynamic domain name resolution allows the system DNS resolver to query an external DNS server for the IP addresses mapped to domain names when a user performs domain name operations through applications.

If a source interface is specified in the command, when an IPv4 DNS query packet is sent, the primary IPv4 address of the source interface is used as the source address of the DNS query packet. When an IPv6 DNS query packet is sent, the first effective IPv6 address of the source interface is used as the source address of the DNS query packet.

1.6.2 Restrictions and Guidelines

If no address is configured for a source interface, a DNS query packet fails to be sent. If an IPv4 source address is specified in the command, when an IPv4 DNS query packet is sent, the configured IPv4 address serves as the source address of the IPv4 DNS query packet, and an IPv6 DNS query packet will fail to be sent.

If an IPv6 source address is specified in the command, when an IPv6 DNS query packet is sent, the configured IPv6 address serves as the source address of the IPv6 DNS query packet, and an IPv4 DNS query packet will fail to be sent.

⚠ Caution

An effective IPv6 address is a unicast address, but it cannot be a local link address or a loopback address.

1.6.3 Prerequisites

The IP address of a DNS server has been obtained.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable DNS domain name resolution.

ip dns [**oob** [**via** *mgmt-name*] | **vrf** *vrf-name*]

DNS domain name resolution is enabled by default.

(4) Configure a DNS server.

(IPv4 network)

ip nameserver-config [**oob**] { *ipv4-address&<1-6>* } [**via** *mgmt-name*]

No IPv4 address is configured for a DNS server by default.

(IPv6 network)

ip nameserver-config [**oob**] { *ipv6-address&<1-6>* } [**via** *mgmt-name*]

No IPv6 address is configured for a DNS server by default.

(5) Configure the domain name resolution using a specified source interface or address. The configuration steps below are mutually exclusive. Please configure only one task.

- Configure the domain name resolution using a specified source interface.

  ip dns [**vrf** *vrf-name*] **source** *interface-type interface-number*

- Configure the domain name resolution using a specified source IP address.

  (IPv4 network)

  ip dns [**vrf** *vrf-name*] **source ip** *ipv4-address*

  (IPv6 network)

  ip dns [**vrf** *vrf-name*] **source ipv6** *ipv6-address*

By default, no source interface is specified, and the source IP address of a DNS query packet is decided through the routing process.

1.7 Configuring DNS Proxy

Note

DNS proxy is supported by only the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1.7.1 Overview

DNS proxy is deployed between a DNS server and a PC, and acts as a proxy of a DNS server to process users' DNS domain name resolution requests.

1.7.2 Prerequisites

The IP address of a DNS server has been obtained.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the DNS proxy function.

ip dns proxy enable

DNS proxy is disabled by default.

(4) (Optional) Configure the IP address mapped to a domain name.

(IPv4 network)

ip dns proxy hosts-config *host-name ipv4-address*

No static mapping between a domain name and an IPv4 address is configured by default.

(IPv6 network)

ip dns proxy hosts-config *host-name ipv6-address*

No static mapping between a domain name and an IPv6 address is configured by default.

(5) Configure a DNS server.

(IPv4 network)

ip dns proxy nameserver *ipv4-address*

No IPv4 address is configured for a DNS server by default.

Each time this command is run, the device will add one DNS server. The system supports up to six servers. When a domain name cannot be obtained from the first server, the device tries to send a DNS request to subsequent servers until it receives a correct reply.

(IPv6 network)

ip dns proxy nameserver *ipv6-address*

No IPv6 address is configured for a DNS server by default.

Each time this command is run, the device will add one DNS server. The system supports up to six servers. When a domain name cannot be obtained from the first server, the device tries to send a DNS request to subsequent servers until it receives a correct reply.

(6) (Optional) Enable the function of caching dynamic entries on the DNS proxy.

ip dns proxy cache

The function of caching dynamic entries is enabled on the DNS proxy by default.

(7) (Optional) Configure the time to live (TTL) of a reply packet.

ip dns proxy ttl *ttl*

The default TTL of the reply packet in response to a static entry is 3,600 seconds.

(8) (Optional) Configure the range of a source port used for query in upper-level DNS servers.

ip dns proxy port-range *port-min port-max*

The default source port range is from 55,000 to 58,000.

If the port range is too narrow, the concurrent processing performance of the device is affected. If the port range is too wide, excessive flow table entry resources will be occupied and the egress device needs to adjust the flow table restriction synchronously.

1.8 Monitoring

Run the display command to check the running status of a configured function to verify the configuration effect.

Run the debugging command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the cleanup command to clear information.

Caution

Running the **cleanup** command may lose vital information and thus interrupt services.

Table 1-1 DNS Monitoring

Command	Purpose
cleanup hosts-config [*host-name*]	Clears the dynamic domain name cache table.
display hosts-config [*host-name*]	Displays DNS parameters.
debugging ip dns	Debugs the DNS function.
cleanup dns proxy hosts [*host-name*] | *****]	Clears the dynamic domain name cache table on the DNS proxy.
display dns proxy hosts-config	Displays DNS parameters on the DNS proxy.

Command	Purpose
debugging dns-proxy	Debugs the DNS proxy.

1.9 Configuration Examples

1.9.1 Configuring Static Domain Name Resolution

1. Requirements

As shown in Figure 1-2, device A needs to access the server through the domain name www.test.com. In this case, static domain name resolution can be used to implement the access to the server whose IP address is 192.168.10.2 through the domain name "www.test.com".

2. Topology

Figure 1-2 Configuring Static Domain Name Resolution



3. Notes

Manually create the mapping between www.test.com and IP address 192.168.10.2 in the static resolution table of device A.

4. Procedure

(1) Configure device A.

Configure an IP address for the Layer 3 Ethernet interface GigabitEthernet 0/1.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-GigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-GigabitEthernet 0/1)# ip address 192.168.10.1 24
DeviceA(cfg)# exit
```

Configure the mapping between www.test.com and IP address 192.168.10.2.

```
DeviceA(cfg)# ip hosts-table www.test.com 192.168.10.2
```

(2) Configure the server.

Set the IP address of the server to 192.168.10.2.

5. Verification

Run the display hosts-config command to check whether the static domain name entry is configured.

```
DeviceA# display hosts-config
```

```
Name servers are:

Host                    type      Address          TTL(sec)
www.test.com            static    192.168.10.1       ---
```

Run the ping www.test.com command on device A. The IP address mapped to www.test.com is resolved as 192.168.10.2.

```
DeviceA# ping www.test.com

Translating "www.test.com"...[OK]
Sending 5, 100-byte ICMP Echoes to 192.168.10.2, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
 interface gigabitethernet 0/1
 ip address 192.168.10.1 24
!
 ip   hosts-table www.test.com 192.168.10.2
!
```

1.9.2 Configuring Dynamic Domain Name Resolution

1. Requirements

As shown in Figure 1-3, device A is a DNS client, which can resolve domain names through the DNS server on the network.

2. Topology

Figure 1-3 Configuring Dynamic Domain Name Resolution



3. Notes

On device A, set the IP address of a DNS server to 192.168.10.2.

4. Procedure

(1) Configure device A.

Configure an IP address for port GigabitEthernet 0/1.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitEthernet 0/1
DeviceA(cfg-GigabitEthernet 0/1)# ip address 192.168.10.1 24
DeviceA(cfg)# exit
```

Configure the IP address of the DNS server.

```
DeviceA(cfg)# ip nameserver-config 192.168.10.2
DeviceA(cfg)# exit
```

(2) Configure the DNS server.

The way to configure a DNS server varies with the operating system. For details, see the configuration instructions from system vendors.

5. Verification

Run the display hosts-config command to display the configured DNS server.

```
DeviceA(cfg)# display hosts-config
Name servers are:
192.168.10.2 static

Host            type    Address                                 TTL(sec)
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
 interface gigabitEthernet 0/1
 ip address 192.168.10.1 24
!
 ip nameserver-config 192.168.10.2
!
```

1.9.3 Configuring DNS Proxy

Note

DNS proxy is supported by only the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1. Requirements

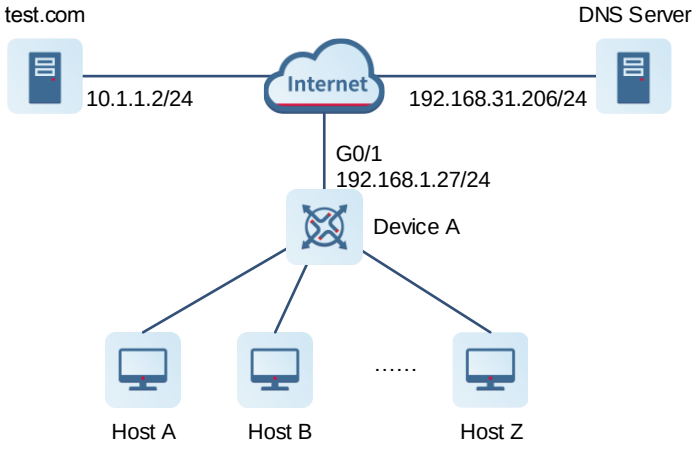
As shown in Figure 1-4, a DNS server is deployed on the network to provide the domain name service for external devices. test.com is deployed on the network to provide services for external devices over test.com.

There are several hosts A-Z in the local area network (LAN). If the IP address of the DNS server is specified on each host, the network administrator needs to change the configuration of all the hosts one by one when the IP address of the DNS server changes. The workload is very heavy. Device A in the LAN needs to be configured as a DNS proxy, the IP address of the DNS server specified on the DNS proxy is the real address of the DNS

server. In this way, when the IP address of the DNS server changes, only the configuration on the DNS proxy needs to be changed, which can greatly reduce the workload.

2. Topology

Figure 1-4 Configuring DNS Proxy



3. Notes

- On device A, enable the DNS proxy function.
- On device A, set the IP address mapped to test.com to 10.1.1.2/24.
- On device A, set the IP address of the DNS server to 192.168.31.206/24.

4. Procedure

(1) Configure device A.

Configure an IP address for port GigabitEthernet 0/1.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-GigabitEthernet 0/1)# ip address 192.168.1.27 24
DeviceA(cfg-GigabitEthernet 0/1)# exit
```

Configure the DNS proxy function.

```
DeviceA(cfg)# ip dns proxy enable
DeviceA(cfg)# ip dns proxy    hosts-config test.com 10.1.1.2
DeviceA(cfg)# ip dns proxy nameserver 192.168.31.206
DeviceA(cfg)# exit
```

(2) Configure the DNS server.

The way to configure a DNS server varies with the operating system. For details, see the configuration instructions from system vendors.

5. Verification

Run the display dns proxy hosts-config command to check whether the DNS proxy is configured.

```
DeviceA(cfg)# display dns proxy hosts-config
proxy state: enable
Name servers are:
192.168.31.206

static host        max: 1024
static host     count: 1
cache domain       max: 5120
cache domain    count: 0
Host                              type              TTL(sec)          Address
test.com                          static            ---               10.1.1.2
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
interface gigabitethernet 0/1
ip address 192.168.1.27 24
!
ip dns proxy enable
ip dns proxy hosts-config test.com 10.1.1.2
ip dns proxy nameserver 192.168.31.206
!
```

Contents

1 Configuring IPv6 Basics 1
1.1 Overview 1
1.2 Advantages 1
1.3 Principles 2
1.3.1 IPv6 Basics 2
1.3.2 ICMPv6 Protocol 12
1.3.3 IPv6 PMTUD 13
1.3.4 IPv6 Neighbor Discovery Protocol 14
1.3.5 Protocols and Standards 17
1.4 Configuring IPv6 Basic Information 17
1.4.1 Overview 17
1.4.2 Configuration Task Summary 17
1.4.3 Configuring an IPv6 Address 18
1.4.4 Configuring the IPv6 MTU for an Interface 19
1.4.5 Configuring IPv6 Source Routing 20
1.4.6 Configuring the IPv6 Hop Limit 20
1.4.7 Configuring the Default Gateway for a Management Interface 21
1.4.8 Configuring the Specified Source Address for Sending ICMPv6 Packets 21
1.4.9 Configuring the Transmission Rate of ICMPv6 Error Messages 22
1.4.10 Configuring the Function of Sending ICMPv6 Destination Unreachable Messages 23
1.5 Configuring IPv6 NDP 23
1.5.1 Overview 23

1.5.2 Configuration Task Summary....23
1.5.3 Configuring a Static Neighbor Entry....25
1.5.4 Configuring ND Entry Management....25
1.5.5 Configuring Address Resolution....27
1.5.6 Configuring NUD....28
1.5.7 Configuring DAD....29
1.5.8 Configuring Redirection....30
1.5.9 Configuring Parameters Related to RA Packets....31
1.5.10 Configuring ND Logging....33
1.5.11 Configuring the Interval for ND Packet Rate Statistics Collection....33
1.5.12 Configuring Local ND Proxy....33
1.5.13 Restraining an Interface from Sending NS Packets to Authenticated VLANs....35
1.5.14 Configuring Scheduled Automatic ND Scanning....35
1.5.15 Making Remote ND Entries Take Effect....36
1.6 Monitoring....37
1.7 Configuration Examples....38
1.7.1 Manually Configuring IPv6 Addresses....38
1.7.2 Enabling Stateless IPv6 Address Auto-configuration....40

1 Configuring IPv6 Basics

1.1 Overview

As the Internet develops rapidly and IPv4 address space is becoming exhausted, IPv4 limitations become more and more obvious. At present, many researches and practices on Internet Protocol Next Generation (IPng) have been conducted. The IPng working group of the Internet Engineering Task Force (IETF) has formulated an IPng protocol named IP Version 6 (IPv6).

1.2 Advantages

- Larger address space

Compared with 32-bit IPv4 addresses, the length of an IPv6 address is extended to 128 bits. Therefore, the address space has approximately 2^{128} addresses. IPv6 supports a hierarchical address allocation mode from the Internet core network to intranet subnet.

- Simplified packet header format

Since the design principle of the IPv6 packet header is to minimize the cost of the packet header, multiple non-key fields and optional fields are removed from the packet header to the extended packet header. Therefore, although the length of an IPv6 address is four times that of an IPv4 address, the IPv6 packet header is only twice the IPv4 packet header in length. Without the checksum field in the IPv6 packet header, IPv6 devices do not need to process fragments during forwarding. Therefore, the packet can be more efficiently forwarded.

- Efficient hierarchical addressing and routing structure

IPv6 uses a convergence mechanism and defines a flexible hierarchical addressing and routing structure. Multiple networks at the same layer are represented with a uniform network prefix on the upstream device, greatly reducing routing entries maintained by the device and routing and storage costs of the device.

- Easy management: plug and play (PnP)

IPv6 provides automatic discovery and auto-configuration functions to simplify management and maintenance of network nodes. For example, neighbor discovery (ND), maximum transmission unit (MTU) discovery, router advertisement (RA), router solicitation (RS), and auto-configuration technologies provide related services for PnP.

IPv6 supports stateful and stateless address configuration modes. In IPv4, Dynamic Host Configuration Protocol (DHCP) realizes auto-configuration of the host IP addresses and related parameters. IPv6 inherits this auto-configuration service from IPv4, which is called stateful auto-configuration. For details, see *Configuring DHCPv6*. Besides, IPv6 also offers the stateless auto-configuration service. During stateless auto-configuration, a host automatically obtains the link-local address, address prefix of the local device, and other parameter configurations.

- Security

Internet Protocol Security (IPSec) is an optional extension protocol of IPv4 but a part of IPv6 to ensure security of IPv6. At present, IPv6 provides two mechanisms: authentication header (AH) and encapsulated security payload (ESP). AH ensures data integrity and authenticates IP packet sources to ensure that the packets are

from the nodes identified by the source addresses. ESP provides data encryption to realize end-to-end encryption.

- Better QoS support

A new field in the IPv6 packet header defines how to identify and process data streams. The **Flow Label** field in the IPv6 packet header is used to authenticate a data stream, and users can propose requirements on the communication quality by using this field. A device can identify all packets belonging to a specific data stream based on this field and process these packets according to user requirements.

- New protocol for neighboring node interaction

IPv6 Neighbor Discovery Protocol (NDP) uses a series of Internet Control Message Protocol Version 6 (ICMPv6) packets to implement interactive management of adjacent nodes (nodes on the same link). IPv6 uses NDP packets and efficient multicast/unicast ND packets to replace the broadcast-based Address Resolution Protocol (ARP) and Internet Control Message Protocol Version 4 (ICMPv4) router discovery packets.

- Extensibility

IPv6 features strong extensibility, allowing users to add new features to the extension packet headers following the IPv6 packet header. The IPv4 packet header supports at most 40 bytes of options. By contrast, the length of an IPv6 extension packet header is subject to only the maximum number of bytes in the packet.

1.3 Principles

1.3.1 IPv6 Basics

1. IPv6 Address Format

An IPv6 address is represented in the X:X:X:X:X:X:X format. A 128-bit IPv6 address is separated into eight groups by colons (:), and the 16 bits in each group are represented by four hexadecimal characters (0-9, A-F). Each "X" represents a group of four hexadecimal values. The following are three valid IPv6 addresses:

2001:ABCD:1234:5678:AAAA:BBBB:1200:2100

800:0:0:0:0:0:0:1

1080:0:0:0:8:800:200C:417A

To simplify the writing of IPv6 addresses, the number "0" in an IPv6 address can be abbreviated as follows:

- Each integer in the address must be represented, except the leading zeros in each integer. For example, 2001:00CD:0034:0078:000A:000B:1200:2100 can be abbreviated as 2001:CD:34:78:A:B:1200:2100.
- If an IPv6 address contains a string of zeros, double colons (::) can be used to represent these zeros. For example, 800:0:0:0:0:0:0:1 can be represented as 800::1. The double colons indicate that this address can be extended to a complete 128-bit address. In this approach, only when the 16-bit integers are all zeros, can they be replaced with double colons. Double colons can exist only once in an IPv6 address.

In an IPv4/IPv6 hybrid environment, in an IPv6 address, the least significant 32 bits can be used to represent an IPv4 address. This IPv6 address can be represented in a hybrid manner, that is, X:X:X:X:X:X:d.d.d.d, "X" is a hexadecimal integer and "d" is an 8-bit decimal integer. For example, 0:0:0:0:0:0:192.168.20.1 is a valid IPv6 address. It can be abbreviated to ::192.168.20.1. One typical application of this hybrid representation is to map an IPv4 address to an IPv6 address, for example, IPv4 address 1.1.1.1 is mapped to an IPv6 address ::FFFF:1.1.1.1.

2. IPv6 Address Structure

An IP address consists of two parts:

- Network prefix: n bits, similar to the network ID in an IPv4 address.
- Interface ID: (128 – n) bits, similar to the host ID in an IPv4 address.

Therefore, the length of the network prefix can be expressed using an additional value according to the classless inter-domain routing (CIDR), and this value is separated from an IPv6 address by a slash (/), for example, 12AB::CD30:0:0:0:0/60, the prefix length used for routing in the address is 60 bits.

3. IPv6 Address Type

RFC 4291 defines three types of IPv6 addresses:

- Unicast address: ID of a single interface. Packets destined to a unicast address are sent to the interface identified by this address.
- Multicast address: ID of an interface group (the interfaces generally belong to different nodes). Packets destined to a multicast address are sent to all interfaces included in this address.
- Anycast address: ID of an interface group. Packets destined to an anycast address are sent to one of the interfaces included in this address (the nearest interface according to the routing protocol).

Note

IPv6 does not define broadcast addresses.

4. Unicast Addresses

Unicast addresses are classified into five types: unspecified address, loopback address, link-local addresses, site-local addresses, and global unicast addresses. Currently site-local addresses have been abolished.

- Unspecified address

  The unspecified address is 0:0:0:0:0:0:0:0, which is usually abbreviated to ::. It has two general purposes:

  - If a host has no unicast address when started, it uses the unspecified address as the source address to send an RS packet to obtain prefix information from the gateway and thereby generate a unicast address.
  - When an IPv6 address is configured for a host, the device detects whether the address conflicts with addresses of other hosts in the same network segment and uses the unspecified address as the source address to send a neighbor solicitation (NS) packet (similar to a gratuitous ARP packet).

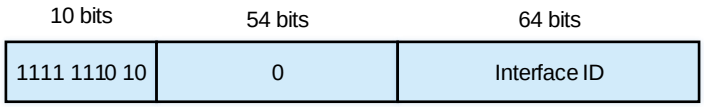
- Loopback address

  The loopback address is 0:0:0:0:0:0:0:1, which is usually abbreviated to ::1. Similar to IPv4 address 127.0.0.1, the loopback address is generally used by a node to send packets to itself.

- Link-local address

  The format of a link-local address is as follows:

Figure 1-1 Link-Local Address



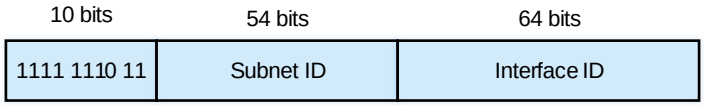
A link-local address is used to number hosts on a single network link. The address identified by the first 10 bits in the prefix is the link-local address. A device will not forward packets in which the source or destination address contains the link-local address. The middle 54 bits of this address is 0, and the last 64 bits represent the interface ID. The address part in the middle allows a single network to connect up to 2^{64} – 1 hosts.

IPv6 link-local addresses can be generated automatically by a device or specified manually.

- Site-local address

  The format of a site-local address is as follows:

Figure 1-2 Site-Local Address

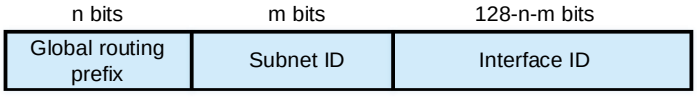


A site-local address can be used to transmit data within a site. A device never forwards packets in which the source or destination address contains a site-local address to the Internet. A site can be assumed as an enterprise's local area network (LAN). Such addresses are similar to IPv4 private addresses such as 192.168.0.0/16. RFC 3879 has abolished site-local addresses. New addresses do not support this prefix and are all regarded as global unicast addresses. Existing addresses can continue to use this prefix.

- Global unicast address

  Except unspecified, loopback, and link-local addresses, all other unicast addresses are global unicast addresses. The format of a global unicast address is as follows:

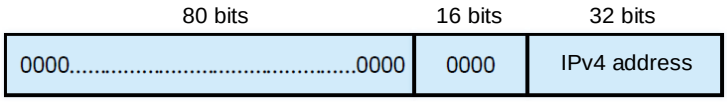
Figure 1-3 Global Unicast Address



Among global unicast addresses, there is a type of IPv4-embedded IPv6 addresses, including IPv4-compatible IPv6 addresses and IPv4-mapped IPv6 addresses. They are used for interconnection between IPv4 nodes and IPv6 nodes.

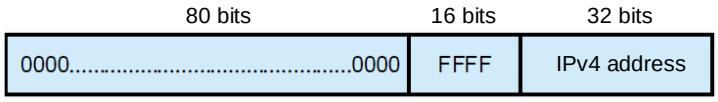
  - The format of an IPv4-compatible IPv6 address is as follows:

Figure 1-4 Format of an IPv4-compatible IPv6 Address



- The format of an IPv4-mapped IPv6 address is as follows:

Figure 1-5 Format of an IPv4-mapped IPv6 Address



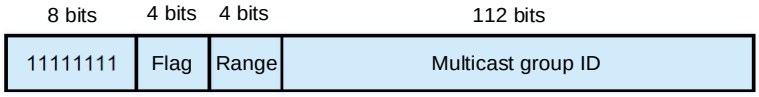
IPv4-compatible IPv6 addresses are mainly used on automatic tunnels. Nodes on automatic tunnels support both IPv4 and IPv6. Using these addresses, IPv4 devices transmit IPv6 packets over tunnels. At present, IPv4-compatible IPv6 addresses have been abolished. IPv4-mapped IPv6 addresses are used by IPv6 nodes to access IPv4-only nodes. For example, if an IPv6 application on an IPv4/IPv6 host requests to resolve the name of an IPv4-only host, the name server dynamically generates an IPv4-mapped IPv6 address and returns it to the IPv6 application.

5. Multicast Addresses

One IPv6 multicast address usually identifies interfaces on a series of different nodes. After a packet is sent to a multicast address, the packet is then forwarded to the interfaces on each node identified by this multicast address.

A multicast address is prefixed with FF00::/8. The format of an IPv6 multicast address is as follows:

Figure 1-6 Multicast Address



- Flag field

  The flag field consists of four bits. Currently only the fourth bit is specified. If this flag bit is 0, the address is a permanent known multicast address assigned by the Internet Assigned Numbers Authority (IANA). If this flag bit is 1, the address is a temporary multicast address used in a certain scenario. The remaining three flag bits are reserved for future use and set to 0.

- Scope field:

  The scope field consists of four bits to indicate the scope type of a multicast address. The following table lists common multicast address scope types and their corresponding values. Unlisted values indicate reserved or unspecified.

Table 1-1 Common Multicast Address Scope Types

Binary Value	Hexadecimal Value	Scope Type
0001	1	Local interface scope
0010	2	Local link scope
0011	3	Local subnet scope
0100	4	Local management scope
0101	5	Local site scope
1000	8	Organization scope
1110	E	Global scope

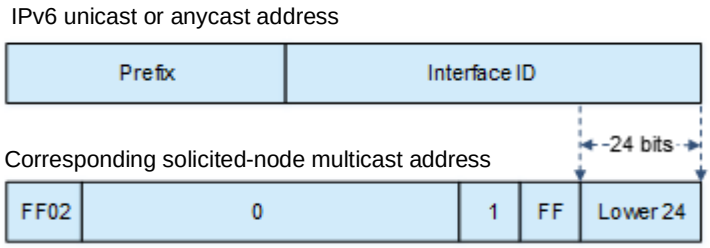
- Multicast group ID field:

  The multicast group ID field consists of 112 bits to identify a multicast group. A multicast group ID can represent different groups based on the flag and scope fields.

Multicast addresses for special purposes:

- FF01::1 represents the multicast address of all nodes in the node-local scope.
- FF02::1 represents the multicast address of all nodes in the link-local scope.
- FF01::2 represents the multicast address of all devices in the node-local scope.
- FF02::2 represents the multicast address of all devices in the link-local scope.
- Solicited-node multicast address. The address is mainly used to obtain the link layer address of a neighbor node on the same link and perform the duplicate address detection. Each unicast or anycast IPv6 address has a corresponding solicited-node address. The address is in the format of FF02:0:0:0:0:1:FFXX:XXXX, FF02:0:0:0:0:1:FF is the fixed 104 bits, and XX:XXXX is the last 24 bits of a unicast or anycast IPv6 address. Solicited-node multicast addresses are usually used in NS packets. The address format is as follows:

Figure 1-7 Format of a Solicited-Node Multicast Address



6. Anycast Addresses

Similar to a multicast address, an anycast address can also be shared by multiple nodes. The difference is that only one node using the anycast address receives data packets while all nodes using the multicast address receive data packets. Since anycast addresses are allocated from the normal IPv6 unicast address space, they use the same format as unicast addresses. Therefore, each member of an anycast address group must be configured explicitly for easier recognition.

Caution

Anycast addresses can be allocated only to devices not hosts and cannot be used as source addresses of packets.

RFC 2373 defines an anycast address called subnet-router anycast address. The following figure shows the format of a subnet-router anycast address. Such an address consists of the subnet prefix and a series of zeros (interface ID).

The subnet prefix identifies a specified link (subnet). Packets destined to the subnet-router anycast address will be forwarded to a device on this subnet. A subnet-router anycast address is usually used by the application on a node to communicate with a device on a remote subnet.

Figure 1-8 Format of a Subnet Router Anycast Address



Note

The format of an anycast address is shown in Figure 1-8, that is, n bits of subnet prefix plus (128 – n) bits of all zeros. For example, in the anycast address 1000:1::100/120, the length of the subnet prefix is 120 bits, and the host address is 8-bit all zeros. An anycast address cannot be used as the interface address of a switch.

7. IEEE EUI-64 Interface ID

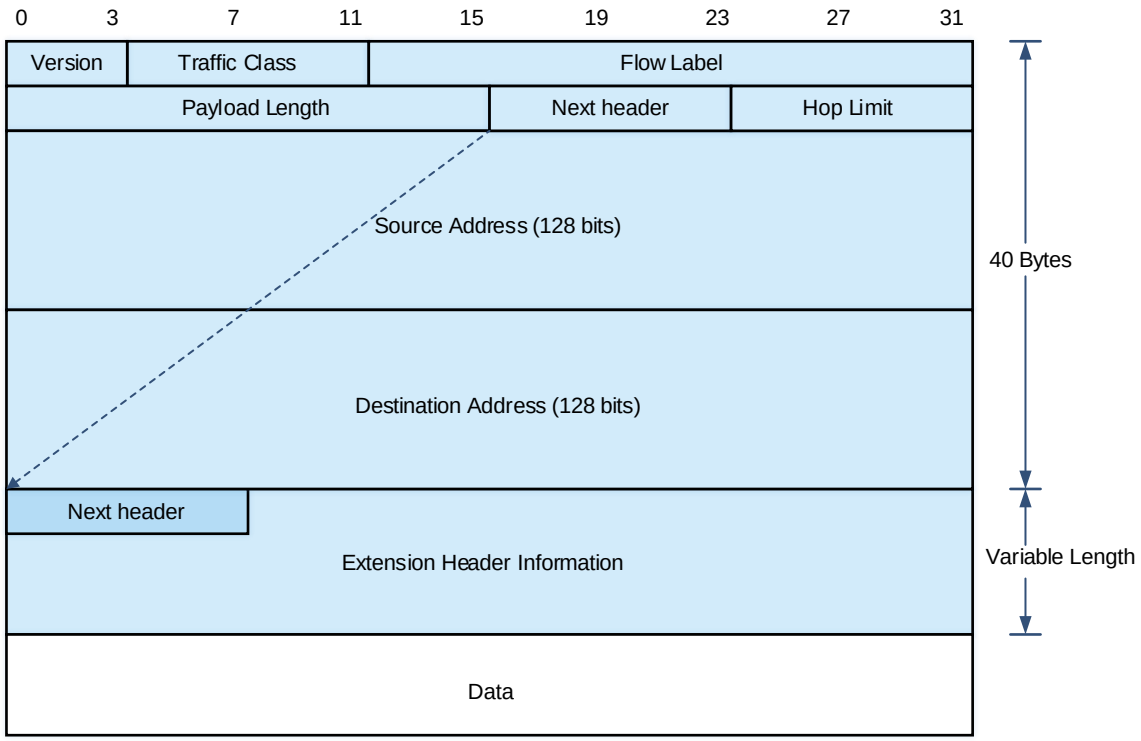
The 64-bit interface ID in an IPv6 address is used to identify a unique interface on the link. This address is a variation of a link-layer address (such as an MAC address) of an interface. This address is generated by inserting a hexadecimal number FFFE after the 24th bit of an MAC address. To make the scope of an interface ID the same as the original MAC address, the U/L bit (the 7th bit starting from the most significant bit) is negated. The resulting set of numbers is used as the interface ID in EUI-64 format.

For the loopback, VBDIF, and tunnel interfaces, the EUI-64 addresses are generated from the MAC addresses of these interfaces, with the last two bytes of the interface index filled in the middle.

8. Format of an IPv6 Packet Header

The format of an IPv6 packet header is shown in the following figure.

Figure 1-9 Format of an IPv6 Packet Header



The IPv6 packet header consists of 40 bytes, in unit of eight bytes. An IPv6 packet header has the following fields:

- Version

  This field consists of 4 bits. In an IPv6 address, this field must be 6.

- Traffic Class

  This field consists of 8 bits. This field indicates the service provided for this packet, and is similar to the type of service (ToS) field in an IPv4 address.

- Flow Label

  This field consists of 20 bits to identify packets belonging to the same service flow. One node can act as the transmission source of multiple service flows. The flow label and source node address uniquely identify one service flow.

- Payload Length

  This field consists of 16 bits, including the packet payload length and the length of IPv6 extended options (if available). That is, it includes the IPv6 packet length except the IPv6 packet header.

- Next Header

  This field indicates the protocol type in the header field following the IPv6 packet header. Similar to the protocol field in the IPv4 address header, the **Next Header** field is used to indicate whether the upper layer uses Transmission Control Protocol (TCP) or User Datagram Protocol (UDP). It can also be used to indicate existence of an IPv6 extension header.

- Hop Limit

This field contains eight bits. Every time a device forwards a packet, the hop value is reduced by 1. If the hop value reaches 0, this packet will be discarded. It is similar to the lifetime field in the IPv4 packet header.

- Source Address

  This field consists of 128 bits and indicates the sender address in an IPv6 packet.

- Destination Address

  This field consists of 128 bits and indicates the receiver address in an IPv6 packet.

At present, IPv6 defines the following extension headers.

- Hop-By-Hop Options

  This extension header must follow the IPv6 packet header. It consists of option data to be checked on each node along the path.

- Routing Options

  This extension header indicates the nodes that a packet passes through from the source address to the destination address. It consists of the address list of the passerby nodes. The initial destination address in an IPv6 packet header is the first address among the addresses in the routing header, but not the final destination address of the packet. After the node corresponding to the destination address in an IPv6 packet header receives the packet, it processes the IPv6 packet header and routing header, and sends the packet to the second address, the third address, and so on in the list in the routing header till the packet reaches the final destination address.

- Fragment

  The source node uses this extension header to fragment the packets, of which the length exceeds the path MTU (PMTU).

- Destination Options

  This extension header replaces the option fields of IPv4. At present, the **Destination Options** field can only be filled with integral multiples of 64 bits (eight bytes) if required. This extension header can be used to carry information to be checked by the destination node.

- Upper-layer Header

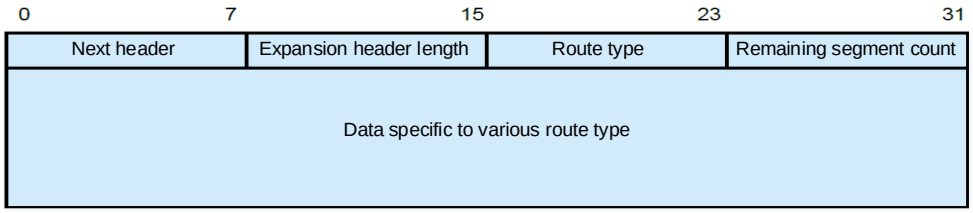
  This extension header indicates the protocol used by the upper layer for data transmission, such as TCP (6) and UDP (17).

Another two extension headers **Authentication** (AH) and **Encapsulating Security Payload** (ESP) will be described in "Configuring IPSec" in "VPN Configuration Guide".

9. IPv6 Source Routing

Similar to the IPv4 loose source routing and loose record routing options, the IPv6 routing header is used to specify the intermediate nodes that a packet passes through. It uses the following format:

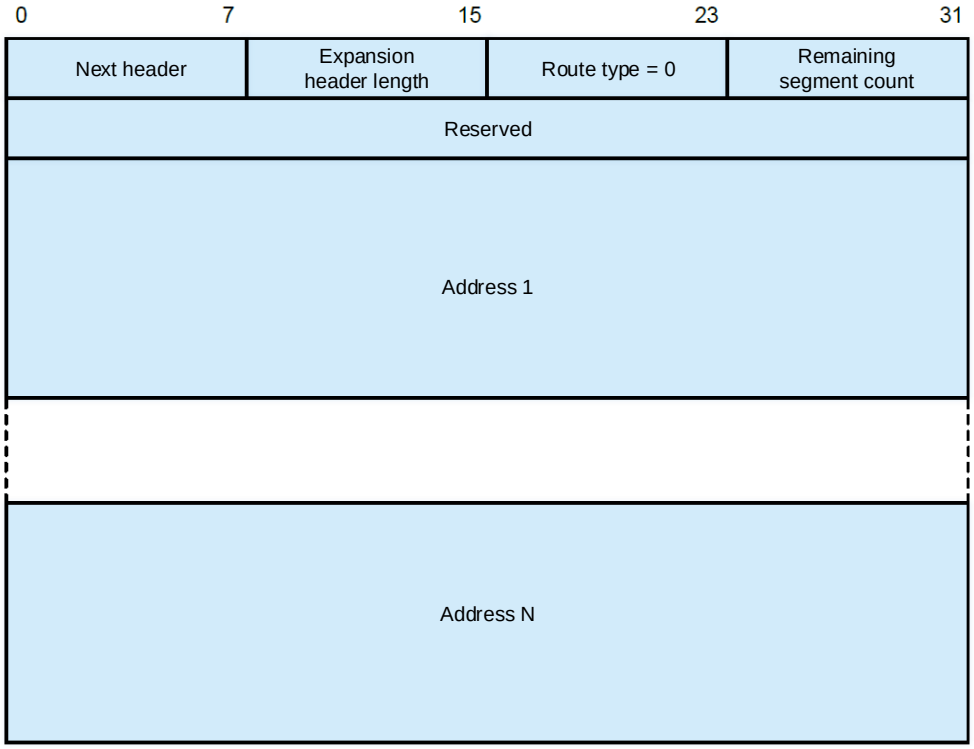
Figure 1-10 IPv6 Routing Header



The **Segments Left** field is used to indicate the number of intermediate nodes specified in the routing header that a packet passes through from the current node to the final destination address.

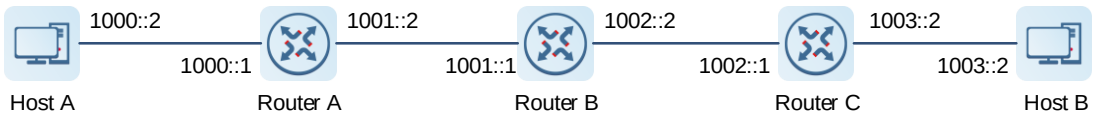
Currently, two routing types are defined: 0 and 2. The Type 2 routing header is used for mobile communication. Type 0 routing header is similar to the loose source routing option of IPv4. The format of the Type 0 routing header is as follows:

Figure 1-11 Type 0 Routing Header



The example describes the application of the Type 0 routing header.

Figure 1-12 Applications of the Type 0 Routing Header



Host A sends host B a packet, which specifies the intermediate nodes router B and router C. The following table lists the changes of fields related to the IPv6 header and routing header during the forwarding process.

Table 1-2 Changes of Fields

Transmission Node	Fields in the IPv6 Header	Fields Related to the Type 0 Routing Header
Host A	Source address = 1000::2 Destination address = 1001::1 (address of router 2)	Segments Left = 2 Address 1 = 1002::1 (address of router 3) Address 2 = 1003::2 (address of host 2)
Router A	No change	
Router B	Source address = 1000::2 Destination address = 1002::1 (address of router 3)	Segments Left = 1 Address 1 = 1001::1 (address of router 2) Address 2 = 1003::2 (address of host 2)
Router C	Source address = 1000::2 Destination address = 1003::2 (address of host 2)	Segments Left = 0 Address 1 = 1001::1 (address of router 2) Address 1 = 1002::2 (address of router 3)
Host B	No change	

The forwarding process is described as follows:

(1) Host A sends a packet in which the destination address is router B's address 1001::1, the Type 0 routing header is filled with router C's address 1002::1 and host B's address 1003::2, and the value of the **Segments Left** field is **2**.

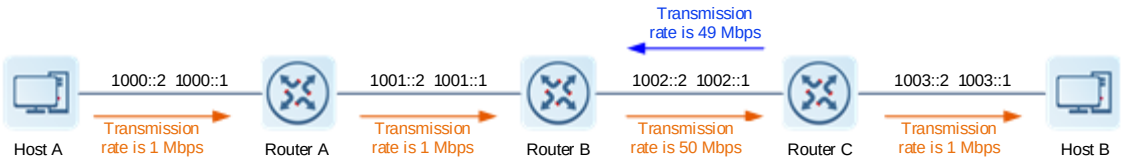
(2) Router A forwards the packet to router B.

(3) Router B interchanges the destination address in the IPv6 header with address 1 in the routing header. That is, the destination address becomes router C's address 1002::1, address 1 in the routing header becomes router B's address 1001::1, and the value of the **Segments Left** field becomes **1**. After modification, router B forwards the packet to router C.

(4) Router C interchanges the destination address in the IPv6 header with address 2 in the routing header. That is, the destination address becomes host B's address 1003::2, address 2 in the routing header becomes router C's address 1002::1, and the value of the **Segments Left** field becomes **0**. After modification, router C forwards the packet to host B.

The Type 0 routing header may be used to initiate denial of service (DoS) attacks. As shown in the following figure, host 1 sends a packet to host 2 at a rate of 1 Mbps and forges a routing header with the **Segments Left** field of 100 to cause multiple round-trips of the packet between router 2 and router 3 (50 times from router 2 to router 3 and 49 times from router 3 to router 2). At the time, the routing header generates the traffic amplification effect: 50 Mbps from router 2 to router 3 and 49 Mbps from router 3 to router 2. Due to this security problem, RFC 5095 abolished the Type 0 routing header.

Figure 1-13 Type 0 Routing Header Used to Initiate DoS Attacks



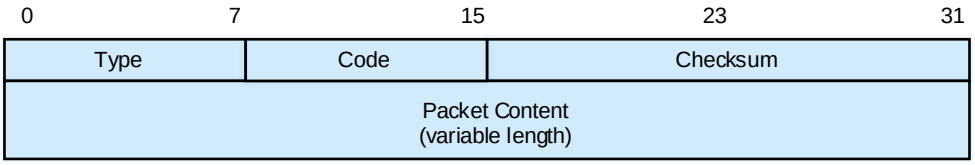
10. IPv6 Hop-limit

An IPv6 data packet passes through routers from the source address to the destination address. If a hop limit is configured, it decreases by one every time the packet passes through a router. When the hop limit decreases to 0, the router discards the packet to prevent this useless packet from being unlimitedly transmitted on the network and wasting network bandwidth. The hop limit is similar to the time to live (TTL) of IPv4.

1.3.2 ICMPv6 Protocol

1. Packet Format

Figure 1-14 ICMPv6 Packet Format



Field descriptions are as follows:

- The **Type** field indicates the type of a packet, 0 to 127 indicates error messages, and 128 to 255 indicates informational messages.
- The **Code** field indicates the subdivision type of this message type.
- The **Checksum** field indicates the checksum of ICMPv6 packets.

2. ICMPv6 Error Messages

ICMPv6 error messages are sent by a destination node or intermediate router to report the errors incurred during IPv6 data packet forwarding and transmission. There are mainly four types of error messages:

- Destination Unreachable (Type=1), indicating that packets cannot be forwarded to the destination node or the upper-layer protocol.
  - Code=0, indicating that there is no route reachable to the destination.
  - Code=1, indicating that the communication with the destination is prohibited according to the administration policy.
  - Code=2, unspecified.
  - Code=3, address unreachable.
  - Code=4, port unreachable.
- Packet Too Big (Type = 2, Code = 0), indicating that the size of a packet exceeds the link MTU of an

interface.

- Time Exceeded (Type=3)
  - Code=0, indicating that the hop limit is exceeded during transmission.
  - Code=1, fragmentation timeout.
- Parameter Problem (Type=4), indicating that an error occurs in the IPv6 header or extension header.
  - Code=0, indicating an incorrect header field.
  - Code=1, indicating that the next header cannot be identified.
  - Code=2, indicating that the IPv6 option cannot be identified.

3. ICMPv6 Informational Messages

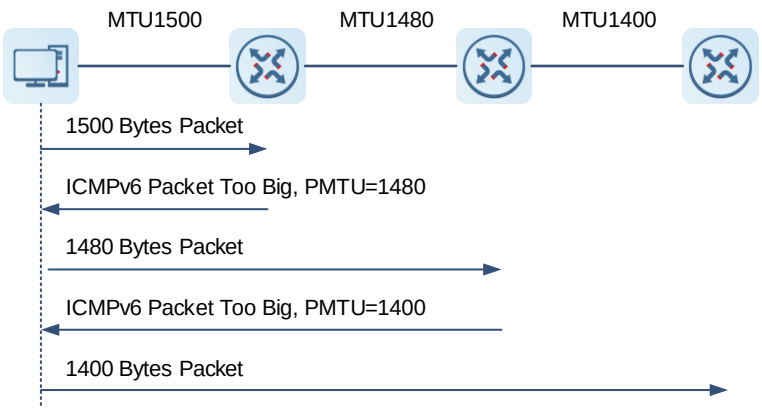
The following describes common informational messages:

- Type=128, echo request, which is sent to the destination node to trigger the destination node to immediately send back an echo reply packet.
- Type=129, echo reply, which is sent in response to an echo request.
- Type=133, router solicitation (RS) packet. After a node is started, it initiates a solicitation to the router via an RS packet to request the prefix and other configurations.
- Type=134, router advertisement (RA) packet, which is sent in response to an RS packet and advertises prefix options and certain flag bits.
- Type=135, neighbor solicitation (NS) packet, which is similar to an ARP request and used to obtain the link layer address of a neighbor, check whether the neighbor is reachable, and perform duplicate address detection.
- Type=136, neighbor advertisement (NA) packet, which is similar to an ARP reply packet.
- Type=137, Redirect packet. By sending a Redirect packet to the source host, the default gateway triggers the host to re-select a better next-hop address for delivering subsequent packets.

1.3.3 IPv6 PMTUD

Similar to IPv4 path MTU discovery (PMTUD), IPv6 PMTUD allows a host to dynamically discover the MTU size in the data path. If the length of a data packet to be sent by a host is greater than the PMTU, the host performs packet fragmentation.

Figure 1-15 IPv6 PMTUD



As shown in the preceding figure, if the length of a packet to be sent by a host is greater than the PMTU, the router discards this packet and sends an ICMPv6 Packet Too Big message containing its PMTU to the host. The host fragments the packet based on the new PMTU. In this manner, the router does not need to perform fragmentation, saving router resources and improving the IPv6 network efficiency.

1.3.4 IPv6 Neighbor Discovery Protocol

Neighbor Discovery Protocol (NDP) is a basic part of IPv6. Its main functions include router discovery, prefix discovery, parameter discovery, address auto-configuration, address resolution (like ARP), next-hop determination, neighbor unreachability detection (NUD), duplicate address detection (DAD), and redirection.

NDP defines five types of ICMP packets: RS, RA, NS, NA, and Redirect.

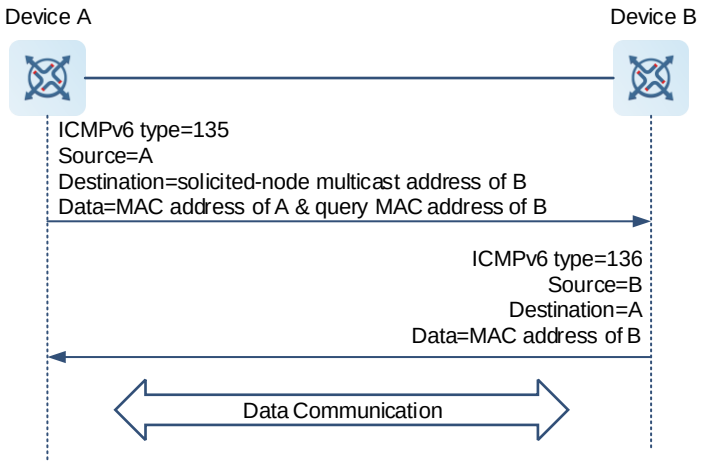
All the above ICMP packets carry one or more options. These options are optional in some cases but are significant in other cases. NDP mainly defines five options: **Source Link-Layer Address** option, Type = 1; **Target Link-Layer Address** option, Type = 2; **Prefix Information** option, Type = 3; **Redirect Header** option, Type = 4; **MTU** option, Type = 5.

Different NDP mechanisms use different ICMPv6 messages.

1. Address Resolution

When a node attempts to communicate with another, the node has to obtain the link-layer address of the peer by sending an NS packet to the peer. In this packet, the destination address is the solicited-node multicast address corresponding to the IPv6 address of the destination node. This packet also contains the link-layer address of the source node. After receiving this NS packet, the destination node replies with an NA packet, in which the destination address is the source address of the NS packet and the content is the link-layer address of the solicited node. After receiving this NA packet, the source node can communicate with the destination node.

Figure 1-16 Address Resolution



The process of address resolution is as follows:

(1) Device A sends an NS packet in multicast mode. The source address of the NS packet is the IPv6 address of the interface of device A, and the destination address is the solicited-node multicast address of device B. The packet contains the link layer address of device A, and solicits the link layer address of device B.

(2) After receiving the NS packet, device B judges whether the destination address of the packet is its own IPv6 address. If yes, device B can learn the link layer address of device A and returns an NA packet containing its link layer address in unicast mode.

(3) Device A can get the link layer address of device B from the received NA packet. The two devices can then begin to communicate with each other.

2. NUD

If the amount of time that a neighbor is considered reachable has elapsed but an IPv6 unicast packet needs to be sent to the neighbor, the device performs NUD. The NS and NA packets can be used to detect whether the neighbor node is reachable.

(1) The node sends an NS packet, in which the destination address is the IPv6 address of the neighbor node.

(2) If the acknowledgment packet is received from the neighbor node, the neighbor is considered reachable. Otherwise, the neighbor is considered unreachable.

While NUD is underway, the device can continue to forward IPv6 packets to the neighbor.

3. DAD

When obtaining an IPv6 address, a node needs to perform DAD to determine whether the IPv6 address is unique on the link. This feature is similar to the gratuitous ARP in IPv4.

The process of DAD is as follows:

(1) A node sends an NS packet. The source IPv6 address of the NS packet is the unspecified address, and the destination address is the solicited-node multicast address corresponding to the IPv6 address to be detected. The packet content is the IPv6 address to be detected.

(2) After receiving the NS packet, other neighbor nodes check whether the IPv6 address exists in their local IPv6 address collections. If yes, the neighbor node that has the IPv6 address responds with an NA packet that carries the IPv6 address to the source node.

(3) When receiving the NA packet from the neighbor, the source node considers the IPv6 address in use. On the contrary, if the source node receives no NA packet in response to the NS packet, the IPv6 address is available.

If a device detects an address conflict, this address is set in the duplicate state and the device cannot receive IPv6 packets destined to this address. Meanwhile, the device also starts a timer for this duplicate address to periodically perform DAD. If no address conflict is detected, this address can be properly used.

4. RA Packets

An RA packet usually contains the following content:

- One or more IPv6 address prefixes (used for on-link determination or stateless address auto-configuration);
- Validity period of the IPv6 address prefix;
- Host auto-configuration method (stateful or stateless);
- Default device information (whether the device acts as the default device; if yes, the duration for acting as the default device is also included);
- Other information provided for host configuration, such as hop limit, MTU and NS retransmission interval.

RA packets can also be used as replies to the RS packets sent by a host. Using RS packets, a host can obtain the auto-configured information immediately after startup rather than wait for the RA packets sent by the device.

If no unicast address is configured for a newly started host, the host uses the unspecified address (0:0:0:0:0:0:0:0) as the source address of the RS packet. Otherwise, the host uses the configured unicast address as the source address and the multicast address of all devices in the local link (FF02::2) as the destination address in the RS packet. As a reply to the RS packet, the RA packet uses the source address of the RS packet as the destination address. If the source address is the unspecified address, it uses the multicast address of all nodes in the local link (FF02::1).

In an RA packet, the following parameters can be configured in the IPv6 interface attributes:

- Ra-interval: Indicates the interval for sending the RA packets.
- Ra-lifetime: Indicates the lifetime of a router, that is, whether the device acts as the default router on the local link and the duration for acting as the default router.
- Prefix: Indicates the prefix of an IPv6 address of the local link. It is used for on-link determination or stateless address auto-configuration, including other parameter configurations related to the prefix.
- Ns-interval: Indicates the NS packet retransmission interval.
- Reachable-time: Indicates the period in which the device regards a neighbor reachable after detecting the neighbor reachability event.
- Ra-hoplimit: Indicates the hop count in an RA packet, which is used to set the hop limit for a host to send a unicast packet.
- Ra-mtu: Indicates the value of the MTU field in an RA packet.
- Managed-config-flag: Determines whether a host receiving this RA packet obtains the address through stateful auto-configuration.
- Other-config-flag: Determines whether a host receiving this RA packet uses DHCPv6 to obtain other information except the IPv6 address for auto-configuration.
- RDNSS and DNSSL options: Provide IPv6 terminals with the address of the DNS recursive query server and a search list of DNS domain names.

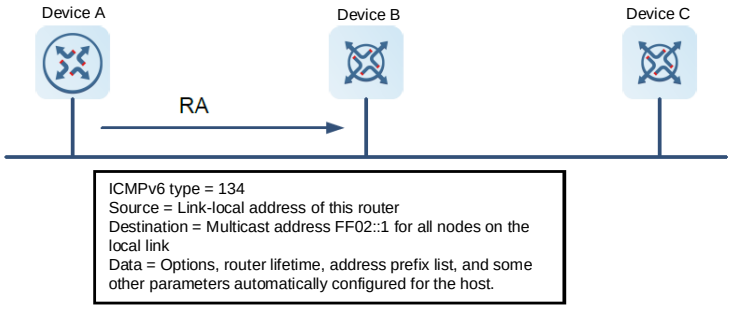
5. Router Discovery/Prefix and Parameter Discovery

RA packets are sent by the device to all link-local nodes on a regular basis. From the RA packets received, the nodes obtain neighbor router information and other configuration parameters. This is called router discovery/prefix and parameter discovery.

The nodes automatically configure the IPv6 address based on the information obtained from router discovery/prefix discovery, which is called stateless auto-configuration.

The RA packet sending process during router discovery/prefix and parameter discovery is shown in the following figure:

Figure 1-17 Router Discovery/Prefix and Parameter Discovery



6. Redirection

If a router receiving an IPv6 packet finds a better next hop, it sends an ICMP Redirect packet to inform the host of the better next hop. The host will directly send the IPv6 packet to the better next hop next time. This function is the same as ICMP Redirect packets of IPv4.

1.3.5 Protocols and Standards

- RFC4291: IPv6 Addressing Architecture.
- RFC2460: Internet Protocol, Version 6 (IPv6) Specification
- RFC4443: Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification
- RFC4861: Neighbor Discovery for IP version 6 (IPv6)
- RFC4862: IPv6 Stateless Address Autoconfiguration
- RFC5059: Deprecation of Type 0 Routing Headers in IPv6

1.4 Configuring IPv6 Basic Information

1.4.1 Overview

IPv6 is updated from IPv4 and solves many shortcomings of the IPv4.

The most significant difference between IPv6 and IPv4 is the upgrade of the IP address length from 32 bits to 128 bits. IPv6 features simplified header format, adequate address space, hierarchical address structure, flexible extension headers, and enhanced neighbor discovery mechanism and will be full of vitality in the future market competition.

1.4.2 Configuration Task Summary

The IPv6 basic configuration includes the following tasks:

(1) Configuring an IPv6 Address

(2) (Optional) Configuring the IPv6 MTU for an Interface

(3) (Optional) Configuring IPv6 Source Routing

(4) (Optional) Configuring the IPv6 Hop Limit

(5) (Optional) Configuring the Default Gateway for a Management Interface

(6) Configure the sending of ICMPv6 packets. The following configuration tasks are optional. Select either of them for configuration according to the actual condition.

- Configuring the Specified Source Address for Sending ICMPv6 Packets
- Configuring the Transmission Rate of ICMPv6 Error Messages
- Configuring the Function of Sending ICMPv6 Destination Unreachable Messages

1.4.3 Configuring an IPv6 Address

1. Overview

This section describes how to enable an interface to support the IPv6 protocol, and configure an IPv6 address for the interface to achieve IPv6 network communication.

The IPv6 link-local address can be manually configured or automatically obtained.

2. Restrictions and Guidelines

- Anycast addresses cannot be configured as interface IPv6 addresses. Exceptionally, anycast addresses with a subnet prefix of 127 or greater can be configured.
- If an interface is bound to a multiprotocol virtual routing and forwarding (VRF) instance not configured with IPv6 address family, IPv6 cannot be enabled on this interface or an IPv6 address cannot be configured for this interface. You can enable IPv6 or configure an IPv6 address for this interface only after configuring an IPv6 address family for the multiprotocol VRF instance.
- IPv6 can be enabled on an interface by two methods: (1) Run the **ipv6 enable** command in interface configuration mode by referring to Step (5); (2) Configure an IPv6 address for the interface by referring to Step (6). If an IPv6 address is configured on an interface, IPv6 is automatically enabled on this interface. In this case, IPv6 cannot be disabled even when you run the **unset ipv6 enable** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure an IPv6 general prefix.

ipv6 common-prefix *prefix-name ipv6-prefix/prefix-length*

No IPv6 general prefix is configured by default.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the switch virtual interface (SVI) configuration mode.

  interface vlan *interface-number*

- Enter the management interface configuration mode.

  interface mgmt *interface-number*

(5) (Optional) Enable IPv6 on the interface.

ipv6 enable

IPv6 is disabled on an interface by default.

(6) Configure an IPv6 address for the interface. The configuration steps below are mutually exclusive. Please configure only one task.

- Manually configure an IP address for the interface.

  - **ipv6 address** { *ipv6-address*/*prefix-length* | *ipv6-prefix*/*prefix-length* **eui-64** | *prefix-name sub-bits*/*prefix-length* [**eui-64**] }

  No IPv6 address is configured by default.

  If an IPv6 interface is created and is in up state, the system automatically generates a link-local address for this interface.

- Enable the IPv6 stateless address auto-configuration on the interface.

  ipv6 address auto-config

  The IPv6 stateless address auto-configuration is disabled on an interface by default.

1.4.4 Configuring the IPv6 MTU for an Interface

1. Overview

IPv6 PMTUD allows a host to dynamically discover the MTU size on the data path. When sending an IPv6 packet, a host compares the IPv6 MTU of the outbound interface with the PMTU and determines which MTU is smaller. If the packet length is greater than the smaller MTU, the host fragments the packet based on the smaller MTU.

2. Restrictions and Guidelines

The IPv6 MTU of an interface must be less than or equal to the interface MTU.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

(4) Configure the IPv6 MTU for the Interface.

ipv6 mtu *bytes*

The MTU value of an IPv6 packet is the same as the value on an interface by default.

1.4.5 Configuring IPv6 Source Routing

1. Overview

Similar to the IPv4 loose source routing and loose record routing options, the IPv6 routing header is used to specify the intermediate nodes that the packet passes through.

2. Restrictions and Guidelines

Since the Type 0 routing header may cause the device vulnerable DoS attacks, the device does not forward IPv6 packets carrying the routing header by default, but still processes IPv6 packets that are finally destined to itself and carry the Type 0 routing header.

An administrator can run the **ipv6 source-route** command in global configuration mode to enable IPv6 source routing.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enables the device to forward IPv6 packets carrying the routing header.

ipv6 source-routing

The device does not forward IPv6 packets carrying the routing header by default.

1.4.6 Configuring the IPv6 Hop Limit

1. Overview

This section describes how to configure the hop limit for unicast packets to prevent infinite transmission of the packets on the network.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the hop limit for unicast packets.

ipv6 max-hops *hop*

The default hop limit for unicast packets is 64.

1.4.7 Configuring the Default Gateway for a Management Interface

1. Overview

This section describes how to configure the default gateway for a management interface. After configuration, a default route is generated, with the outbound interface of the management interface and the next hop of the configured gateway.

2. Restrictions and Guidelines

This command can be configured on management interfaces only.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the management interface configuration mode.

interface mgmt *interface-number*

(4) Configure the default gateway for the management interface.

ipv6 gateway *ipv6-address*

No IPv6 default gateway is configured for a management interface by default.

1.4.8 Configuring the Specified Source Address for Sending ICMPv6 Packets

1. Overview

In a network with a large number of IPv6 addresses configured, it is complex for receivers to recognize the device, from which an ICMPv6 packet is sent. To simplify the judgment, you can configure a specified source address for ICMPv6 packets.

2. Restrictions and Guidelines

You can configure a specified address, like the address of the loopback interface, as the source address of ICMPv6 packets to simplify judgment.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a specified source address for sending ICMPv6 packets.

ipv6 icmp source [**vrf** *vrf-name*] *ipv6-address*

No address is configured as the specified source address for sending ICMPv6 packets by default.

1.4.9 Configuring the Transmission Rate of ICMPV6 Error Messages

1. Overview

When receiving an invalid IPv6 packet, a device will discard the packet and send an ICMPv6 error message to the source IPv6 address. When attacked by a large number of invalid IPv6 packets, the device continuously replies with ICMPv6 error messages till resources are exhausted and thereby fail to properly provide services. This is called DoS attacks. To prevent DoS attacks, you can restrict the transmission rate of ICMPv6 error messages.

If the length of an IPv6 packet to be forwarded exceeds the IPv6 MTU of the outbound interface, the router discards this IPv6 packet and sends an ICMPv6 Packet Too Big message to the source IPv6 address. This error message is mainly used for IPv6 PMTUD. When there are too many ICMPv6 error messages, the ICMPv6 Packet Too Big message may be filtered out, causing the failure of the IPv6 PMTUD function. To avoid this problem, you are advised to restrict the transmission rates of the ICMPv6 Packet Too Big message and other ICMPv6 error messages separately.

Note

Although ICMPv6 Redirect packets are not a type of ICMPv6 error messages, the device restricts their rates together with other ICMPv6 error messages except Packet Too Big messages.

2. Restrictions and Guidelines

Since the precision of the timer is 10 milliseconds, you are advised to set the refresh cycle of a token bucket to an integer multiple of 10 milliseconds. If the refresh cycle is set to a value greater than 0 and smaller than 10 milliseconds, the refresh cycle that actually takes effect is 10 milliseconds. For example, if the refresh rate is set to send one packet within the refresh cycle of 5 milliseconds, the refresh rate that actually takes effect is that two packets are sent within the refresh cycle of 10 milliseconds. If the refresh cycle is not an integral multiple of 10 milliseconds, the refresh cycle that actually takes effect is automatically converted into an integral multiple of 10 milliseconds. For example, if the refresh rate is set to send three packets within the refresh cycle of 15 milliseconds, the refresh rate that actually takes effect is that two packets are sent within the refresh cycle of 10 milliseconds.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the transmission rate of the ICMPv6 error message.

ipv6 icmp error-rate-limit [too-big] *interval* [*bucket-size*]

Ten messages are transmitted within 100 ms by default.

1.4.10 Configuring the Function of Sending ICMPv6 Destination Unreachable Messages

1. Overview

If a device receives an IPv6 unicast UDP packet destined for itself and finds that no upper layer protocol is capable of processing the packet, the device sends an ICMPv6 port unreachable message.

If a device cannot forward a data packet because there is no route, it sends an ICMPv6 route unreachable message to the data source.

If a device cannot forward a data packet because it has no neighbor discovery (ND) entry, it sends an ICMPv6 host unreachable message to the data source.

This command affects all ICMPv6 destination unreachable messages.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the L3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the L3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the L3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the L3 aggregate subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Enable the function of sending ICMP destination unreachable messages.

ipv6 unreachables

The function of sending ICMPv6 destination unreachable messages is enabled by default.

1.5 Configuring IPv6 NDP

1.5.1 Overview

Neighbor Discovery Protocol (NDP) is a basic part of IPv6. Its main functions include router discovery, prefix discovery, parameter discovery, address auto-configuration, address resolution (like ARP), next-hop determination, neighbor unreachability detection (NUD), duplicate address detection (DAD), and redirection.

1.5.2 Configuration Task Summary

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring a Static Neighbor Entry
- Configuring ND Entry Management

- Configuring Address Resolution
- Configuring NUD
- Configuring DAD
- Configuring Redirection
- Configuring Parameters Related to RA Packets
- Configuring ND Logging
- Configuring the Interval for ND Packet Rate Statistics Collection
- Configuring ND Proxy
- Overview

When the device receives an NS request destined to another IPv6 network segment and it knows the route, it sends an NA response containing its own MAC address.

1. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the overlay router interface configuration mode

  interface overlayrouter *interface-number*

(4) Enable ND proxy.
ipv6 proxy-nd

- ND proxy is enabled by default.Configuring Local ND Proxy
- Restraining an Interface from Sending NS Packets to Authenticated VLANs
- Configuring Scheduled Automatic ND Scanning
- Making Remote ND Entries Take Effect

1.5.3 Configuring a Static Neighbor Entry

1. Overview

A static neighbor entry is uniquely identified by the IPv6 address of a neighbor node and the ID of an L3 interface connected to this neighbor node.

2. Restrictions and Guidelines

A static neighbor entry can be configured on IPv6-enabled interfaces only. If the neighbor entry to be configured is learned through NDP and stored in the neighbor table, the dynamic neighbor entry will be automatically converted into a static one. An effective static neighbor entry will be always reachable. An invalid static neighbor entry refers to a static neighbor entry with the configured IPv6 address not matching the address configured on the interface (not within any IPv6 network segment of this interface, or in conflict with the address of this interface). In this case, packets will not be forwarded through the MAC address specified in the static neighbor entry.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static neighbor entry.

ipv6 neighbor *ipv6-address interface-type interface-number mac-address*

No static neighbor entry is configured by default.

1.5.4 Configuring ND Entry Management

1. Overview

- You can restrict the number of ND entries learned by an interface to prevent malicious neighbor learning attacks. If this number is not restricted, a large number of ND entries will be generated on the device, occupying excessive memory space and affecting forwarding performance.
- You can restrict the number of unresolved ND entries to prevent malicious scanning attacks from causing the generation of a large number of unresolved ND entries and occupying entry resources.
- You can configure the maximum number of ND options to prevent forged ND packets from carrying unlimited ND options and occupying excessive CPU space on the device.
- When an interface is enabled to learn ND entries via DAD NS packets, the interface will create ND entries in stale state when receiving the DAD NS packets.

2. Configuration Tasks

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring the Maximum Number of ND Entries That Can Be Learned
- Configuring the Maximum Number of Unresolved ND Entries
- Configuring the Maximum Number of ND Options That Can Be Processed
- Enabling an Interface to Learn ND Entries via DAD NS Packets

3. Restrictions and Guidelines

By default, a general interface is disabled to learn ND entries via DAD NS packets but interfaces in a super VLAN are allowed to do so.

4. Configuring the Maximum Number of ND Entries That Can Be Learned

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the SVI configuration mode.

interface vlan *interface-number*

(4) Configure the maximum number of ND entries that can be learned by an interface.

ipv6 nd interface cache-limit *limit*

The number of neighbor cache entries that can be learned by an interface is not limited by default.

The configured value must be equal to or greater than the number of the neighbor cache entries learned by the current interface. Otherwise, the configuration does not take effect. The configuration is subject to the ND entry capacity supported by the device.

5. Configuring the Maximum Number of Unresolved ND Entries

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of unresolved ND entries.

ipv6 nd unresolved *number*

The default maximum number of unresolved ND entries is 0, indicating no restriction. The number is only subject to the ND entry capacity supported by the device.

6. Configuring the Maximum Number of ND Options That Can Be Processed

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the maximum number of ND options that can be processed.

ipv6 nd max-opt *option*

The device supports 10 ND options by default.

7. Enabling an Interface to Learn ND Entries via DAD NS Packets

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the switch virtual interface (SVI) configuration mode.

interface vlan *interface-number*

(4) Enable the interface to learn ND entries via DAD NS packets.

ipv6 nd dad auto-learning enable

By default, a common interface is not allowed to learn neighbor entries through DAD NS packets. A super VLAN interface is allowed to learn neighbor entries through DAD NS packets.

1.5.5 Configuring Address Resolution

1. Overview

When a node attempts to communicate with another, the node has to obtain the link-layer address of the peer by sending an NS packet to the peer. In this packet, the destination address is the solicited-node multicast address corresponding to the IPv6 address of the destination node. This packet also contains the link-layer address of the source node. After receiving this NS packet, the destination node replies with an NA packet, in which the destination address is the source address of the NS packet and the content is the link-layer address of the solicited node. After receiving the NA packet, the source node can communicate with the destination node.

2. Restrictions and Guidelines

If the device is configured to actively send an NS packet to a specific VLAN in a super VLAN, then it will only send the NS packet to the VLAN specified in the super VLAN.

If an authentication-exempt VLAN is configured but the authentication-exempt VLAN is not in the VLAN list, NS packets will not be actively broadcast to the authentication-exempt VLAN.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configures a device not to forcibly use the link-local address as the source address when NS packets are sent.

unset ipv6 ns-linklocal-src

The link-local address is always used as the source address for sending NS packets by default.

(4) (Optional) Configure the device to actively send NS packets to specific sub-VLANs in the super VLAN.

ipv6 nd resolve vlan { *vlan-list* | **none** }

A device sends NS packets to all sub-VLANs in a super VLAN by default.

Note

This command is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

(5) (Optional) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

interface vlan *interface-number*

(6) Configure the NS packet retransmission interval.

ipv6 nd ns-retrans-interval *interval*

The default NS packet retransmission interval on an interface is unspecified when it is to be filled in the RA packets. The default NS packet retransmission interval is 1000 ms when it is used to control the device to send NS packets.

1.5.6 Configuring NUD

1. Overview

If the duration in which a neighbor is considered reachable or the neighbor is in stale state has elapsed but an IPv6 unicast packet needs to be sent to the neighbor, the device performs NUD.

2. Restrictions and Guidelines

A shorter duration in which a neighbor is considered reachable or in stale state indicates that the device detects unreachable neighbors more quickly but more network bandwidth and device resources will be consumed. Therefore, you are not advised to set the duration to a very small value.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

(4) Configure the duration in which a neighbor is considered reachable.

pv6 nd reachable-period *time*

By default, the value in an RA packet is 0 (indicating unspecified), and the duration in which a neighbor is considered reachable in neighbor discovery is 30,000 ms.

(5) (Optional) Configure the duration in which a neighbor keeps in stale state.

ipv6 nd stale-time *time*

The default duration of a neighbor in stale state is 1 hour.

1.5.7 Configuring DAD

1. Overview

You need to enable DAD before configuring an IPv6 address for an interface. At this moment, the address is in tentative state. If no address conflict is detected by DAD, this address can be correctly used. If an address conflict is detected and the interface ID of this address is an EUI-64 ID, duplicate link-layer addresses exist on this link. In this case, the system automatically disables this interface to prevent IPv6-related operations on this interface. At the time, you must configure a new address for the interface and disable and then enable the interface to start DAD again.

The software will retry DAD on a conflicting address according to the configured number of NS packets to be sent consecutively and interval. If no address conflict is detected, this address can be normally used.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

(4) Configure the number of NS packets to be sent consecutively during DAD.

ipv6 nd dad times *attempts*

The default number of NS packets to be sent consecutively during DAD is 1.

(5) Configure the DAD interval.

ipv6 nd dad retry *retry*

The default DAD interval is 1 second.

1.5.8 Configuring Redirection

1. Overview

If a router receiving an IPv6 packet finds a better next hop, it sends an ICMP Redirect packet to inform the host of the better next hop. The host will directly send the IPv6 packet to the better next hop next time. ICMPv6 redirection has the same function as ICMP Redirect packets of IPv4.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

(4) Enable the IPv6 redirection function on the interface.

ipv6 redirects

The IPv6 redirection function is enabled on an interface by default.

1.5.9 Configuring Parameters Related to RA Packets

1. Overview

An RA packet is a reply to an RS packet and advertises prefix options and certain flag bits.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the URL option type value in an RA packet.

ipv6 nd ra-url *type*

The URL option type value in an RA packet is not configured by default.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

(5) Enable the sending of RA packets on this interface.

unset ipv6 nd ra-send-disable

No RA packets are sent by an IPv6 interface by default.

(6) (Optional) Configure the address prefix to be advertised in an RA packet.

ipv6 nd ra-prefix { *ipv6-prefix*/*prefix-length* | **default** } [[*valid-lifetime* { **infinite** | *preferred-lifetime* } | **at** *valid-date preferred-date* | **infinite** { **infinite** | *preferred-lifetime* }] [**no-autoconfig** | **off-link** | **preference** { *high* | **low** | *medium* } | **proxy**] * | **no-advertise**]

By default, the prefix in an RA packet to be sent by an interface is the prefix configured using the **ipv6 address** command on the interface.

(7) (Optional) Configure the hop limit for RA packets.

pv6 nd ra-hoplimit *ra-hoplimit*

The default hop limit of RA packets is 64.

(8) (Optional) Configure the interval for sending RA packets.

ipv6 nd ra-rate { *interval* | **min-max** *min-interval max-interval* }

The default interval for sending packets is 200 seconds. The actual interval will fluctuate around 200 seconds by ±20%.

(9) (Optional) Configure the MTU for RA packets to be sent by the interface.

ipv6 nd ra-mtu *ra-mtu*

The default MTU of RA packets is the same as the interface MTU value. The interface MTU value is usually 1500. The MTU value may vary with different interface types and is subject the actual interface.

(10) Configure related parameters of RA packets. All the configuration steps below are optional. Select the configuration steps as required.

- Configure the **Managed address configuration** flag bit in an RA packet.

  ipv6 nd managed-config-flag

  The **Managed address configuration** flag bit in an RA packet is not configured by default.

  The settings of the flag bit determine whether a host receiving this RA packet obtains an address through stateful auto-configuration. If this flag bit is configured, an address will be obtained through stateful address auto-configuration. Otherwise, an address will not be obtained through stateful address auto-configuration.

- Configure the **Other stateful configuration** flag bit in an RA packet.

  ipv6 nd other-config-flag

  The **Other stateful configuration** flag bit in an RA packet is not configured by default.

  When this flag bit is set, the flag bit in an RA packet sent by the device is set to 1. A host receiving this flag bit uses DHCPv6 to obtain other information except the IPv6 address for auto-configuration.

  When **Managed address configuration** is set, **Other stateful configuration** is also set by default.

- Configure the address of the DNS recursive resolution server contained in an RA packet. Run the following commands in turn.

  unset ipv6 nd ra dns server send-disable

  The RDNSS option is not carried in an RA packet by default.

  ipv6 nd ra dns server *ipv6-address* { *valid-lifetime* | **infinite** } **sequence** *number*

  The address of the DNS recursive resolution server in an RA packet is not configured by default.

- Configure the DNS suffix in an RA packet. Run the following commands in turn.

  unset ipv6 nd ra dns search-list send-disable

The DNSSL option is not carried in an RA packet by default.

ipv6 nd ra dns search-list *ipv6-domain-name* { *valid-lifetime* | **infinite** } **sequence** *number*

The DNS suffix contained in an RA packet is not configured by default.

- Configure the router lifetime in an RA packet.

  ipv6 nd ra-lifetime *lifetime*

  The default router lifetime in an RA packet is 1,800 seconds.

- Configure the URL address of an RA packet.

  ipv6 nd ra-url *ra-url*

  The URL address of an RA packet is not configured by default.

1.5.10 Configuring ND Logging

1. Overview

After ND logging is configured, system logs about ND packets received and sent by the device will be printed.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable ND logging.

ipv6 nd log enable

ND logging is disabled by default.

(4) Configure ND logging rate.

ipv6 nd log rate *rate*

At most 20 ND logs are printed per minute by default.

1.5.11 Configuring the Interval for ND Packet Rate Statistics Collection

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the interval for ND packet rate statistics collection.

ipv6 nd packet rate-statistics interval *interval*

The ND packet rate statistics collection is disabled by default.

1.5.12 Configuring ND Proxy

2. Overview

When the device receives an NS request destined to another IPv6 network segment and it knows the route, it sends an NA response containing its own MAC address.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the overlay router interface configuration mode

  interface overlayrouter *interface-number*

(4) Enable ND proxy.
ipv6 proxy-nd

1.5.13 ND proxy is enabled by default.Configuring Local ND Proxy

1. Overview

After local ND proxy is enabled on an interface, when receiving NS messages that request addresses of other hosts, the device replies with NA messages containing its MAC address.

If L2 access isolation or inter-subnet isolation (such as sub VLANs) is configured, after local ND proxy is enabled on the gateway, the gateway serves as a proxy to process NS packets from the downlink users and replies with NA packets containing the gateway's MAC address. Thus, the traffic of communication among these users is forwarded by the gateway at L3.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI interface configuration mode.

interface vlan *vlan-id*

(4) Enable local ND proxy.

local-proxy-nd enable [**strictly**]

Local ND proxy is disabled on an interface by default.

1.5.14 Restraining an Interface from Sending NS Packets to Authenticated VLANs

1. Overview

In gateway authentication mode, all sub VLANs in a super VLAN are authenticated VLANs by default. Users in an authenticated VLAN have to pass authentication to access the network. After authentication, a static ND entry is generated on the device. Therefore, when accessing an authenticated user, the device does not need to send NS packets to the authenticated VLAN. If the device attempts to access users in an authentication-exempt VLAN, it only needs to send ARP requests to the authentication-exempt VLAN.

In gateway authentication mode, the function of restraining from sending NS packets to authenticated VLANs is enabled on the device by default. If the device needs to access non-authenticated users in an authenticated VLAN, disable this function.

2. Restrictions and Guidelines

The configuration is supported only on SVIs and takes effect only in gateway authentication mode.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI interface configuration mode.

interface vlan *vlan-id*

(4) Restrain the interface from sending NS packets to authentication-exempt VLANs.

ipv6 nd suppress-auth-vlan-ns

NS packets are not sent to authenticated VLANs by default.

1.5.15 Configuring Scheduled Automatic ND Scanning

1. Overview

After scheduled automatic ND scanning is enabled on an interface, the device scans neighbors in the specified range to learn ND entries of these neighbors. If ND entries of a neighbor are not learned, after the current scan ends, the device waits for the scanning interval (5 minutes by default) and then starts the next scan until the entries are learned.

2. Restrictions and Guidelines

- The value range of the scanning interval is from 1 to 30, in minutes.
- The value range of the scanning rate is from 1 to 100, in IP addresses per second.
- Up to 1000 instances can be configured.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the L3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the L3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the L3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the L3 aggregate subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN routing interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable scheduled automatic ND scanning.

ipv6 nd scan auto { *start-ipv6-address end-ipv6-address* }

The scheduled automatic ND scanning function is disabled by default.

(5) (Optional) Configure the interval of scheduled automatic ND scanning.

ipv6 nd scan interval *minutes*

The default interval of scheduled automatic ND scanning is **5** minutes.

(6) (Optional) Configure the rate of scheduled automatic ND scanning.

ipv6 nd scan rate *rate-value*

The default rate of scheduled automatic ND scanning is **20** IP addresses per second.

1.5.16 Making Remote ND Entries Take Effect

Note

This function is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1. Overview

Remote neighbor entries synchronized by the Ethernet virtual private network (EVPN) configured on OverlayRouter interfaces do not take effect by default, that is, remote entries are not used for forwarding L3 routing packets. After this function is configured, remote entries can be used for forwarding L3 routing packets.

2. Restrictions and Guidelines

For details about the VXLAN, see "Configuring VXLAN" in the *Data Center Configuration Guide*.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VXLAN routing interface configuration mode.

interface overlayrouter *interface-number*

(4) Make remote neighbor entries synchronized by EVPN take effect.

unset ipv6 nd remote inactive

Remote neighbor entries synchronized by EVPN do not take effect by default.

1.6 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** command to clear information.

Caution

Running the **cleanup** command may lose vital information and thus interrupt services.

Table 1-3 IPv6 Monitoring

Command	Purpose
cleanup ipv6 dynamic-neighbor [**vrf** *vrf-name*] [*interface-type interface-number* | **vxlan [force]**]	Clears the dynamically learned neighbor entries.
cleanup ipv6 dynamic-neighbor oob	Clears the dynamically learned neighbor entries on a management interface.
display ipv6 address [*interface-type interface-number*]	Displays IPv6 addresses.
display ipv6 common-prefix	Displays the prefix information in a general prefix.
display ipv6 nd-packet [**interface** *interface-type interface-number*] **statistics**	Displays the statistics on IPv6 ND packets.
display ipv6 nd ra dns search-list	Displays the DNSSL option in RA packets and the suppression status and interface for configuring the DNSSL option.
display ipv6 nd ra dns server	Displays the recursive DNS server address provided by the RDNSS option in RA packets and the suppression status and interface for configuring the RDNSS option.
display ipv6 neighbor-table [**vrf** *vrf-name*] **statistics** [**all**]	Displays the statistics on IPv6 neighbor tables.
display ipv6 packet statistics [*interface-type interface-number* | **total**]	Displays the statistics on IPv6 packets.

Command	Purpose
display ipv6 neighbor-table [[**vrf** *vrf-name*] [*interface-type interface-number*] [*ipv6-address*] | [**detail**] [**oob**] [**verbose**] [*ipv6-address*] | **static**]	Displays neighbor information.
display ipv6 raw-socket [*protocol*]	Displays all IPv6 raw sockets
display ipv6 neighbor-routers [*interface-type interface-number*]	Displays the neighbor router information and RA packets.
display ipv6 sockets	Displays all IPv6 sockets.
display ipv6 udp [**local-port** *port-number*] [**peer-port** *port-number*]	Displays all IPv6 UDP sockets.
display ipv6 udp statistics	Displays the statistics on IPv6 UDP sockets.
display ipv6 interface-status [[*interface*-type interface-number] [**ra-info**] | **brief** [*interface-type interface-numbe*r | **oob**]]	Displays the information about IPv6 interfaces.

1.7 Configuration Examples

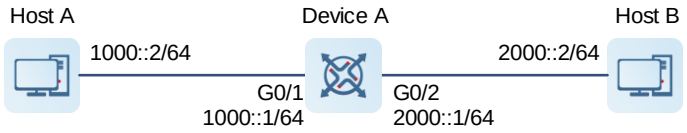
1.7.1 Manually Configuring IPv6 Addresses

1. Requirements

Host A and host B communicate with each other through IPv6 addresses.

2. Topology

Figure 1-18 Manually Configuring IPV6 Addresses



3. Notes

Enable IPv6 on an interface and configure an IPv6 address.

4. Procedure

(1) Configure device A.

Configure an IPv6 address for port GigabitEthernet 0/1 and enable the sending of RA packets on this interface.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
```

```
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 enable
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 address 1000::1/64
DeviceA(cfg-if-GigabitEthernet 0/1)# unset ipv6 nd ra-send-disable
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Configure an IPv6 address for port GigabitEthernet 0/2 and enable the sending of RA packets on this interface.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# ipv6 enable
DeviceA(cfg-if-GigabitEthernet 0/2)# ipv6 address 2000::1/64
DeviceA(cfg-if-GigabitEthernet 0/1)# unset ipv6 nd -ra- send-disable
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

(2) Configure hosts.

Set the IPv6 address of host A to 1000::2/24.

Set the IPv6 address of host B to 2000::2/24.

5. Verification

Run the **display ipv6 interface** command to check that IPv6 addresses are successfully added to the interfaces.

```
DeviceA# display ipv6 interface-status

interface GigabitEthernet 0/1 is Up, ifindex: 2, vrf_id 0
  address(es):
    Mac Address: 00:50:56:b0:05:99
    INET6: FE80::250:56FF:FEB0:599 , subnet is FE80::/64
    INET6: 1000::1 , subnet is 1000::/64
  Joined group address(es):
    FF01::1
    FF02::1
    FF02::2
    FF02::1:FF00:0
    FF02::1:FF00:1
    FF02::1:FFB0:599
  MTU is 1500 bytes
  ICMP error messages limited to one every 100 milliseconds
  ICMP redirects are enabled
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND stale time is 3600 seconds
  ND advertised reachable time is 0 milliseconds
  ND retransmit interval is 1000 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 600 seconds<480--720>
  ND router advertisements live for 1800 seconds

interface GigabitEthernet 0/2 is Down, ifindex: 3, vrf_id 0
```

```
  address(es):
    Mac Address: 00:50:56:b0:05:9a
    INET6: FE80::250:56FF:FEB0:59A [ TENTATIVE ], subnet is FE80::/64
    INET6: 2000::1 [ TENTATIVE ], subnet is 2000::/64
  Joined group address(es):
  MTU is 1500 bytes
  ICMP error messages limited to one every 100 milliseconds
  ICMP redirects are enabled
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND stale time is 3600 seconds
  ND advertised reachable time is 0 milliseconds
  ND retransmit interval is 1000 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 600 seconds<480--720>
  ND router advertisements live for 1800 seconds
```

On host A, run the **ping** command to test the interconnectivity with host B.

On host B, run the **ping** command to test the interconnectivity with host A

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
interface gigabitethernet 0/1
ipv6 enable
ipv6 address 1000::1/64
unset ipv6 nd ra-send-disable
!
interface gigabitethernet 0/2
ipv6 enable
ipv6 address 2000::1/64
unset ipv6 nd ra-send-disable
!
```

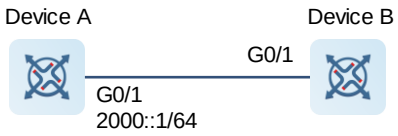
1.7.2 Enabling Stateless IPv6 Address Auto-configuration

1. Requirements

Device A is connected to device B through the GigabitEthernet 0/1 interface which is configured with an IPv6 address. Stateless address auto-configuration needs to be configured on device B to obtain an IPv6 address so that the two devices can communicate with each other through the IPv6 addresses.

2. Topology

Figure 1-19 Enabling Stateless IPv6 Address Auto-configuration



3. Notes

- Configure an IPv6 address on device A.
- Enable stateless address auto-configuration on device B to obtain an IPv6 address.

4. Procedure

(1) Configure device A.

Configure an IPv6 address for port GigabitEthernet 0/1 and enable the sending of RA packets on this interface.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 address 2000::1/64
DeviceA(cfg-if-GigabitEthernet 0/1)# unset ipv6 nd ra-send-disable
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

(2) Configure device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# ipv6 address auto-config
```

5. Verification

On device B, run the **display ipv6 interface** command to check that the interface automatically obtains an IPv6 address.

```
DeviceB# display ipv6 interface

interface GigabitEthernet 0/1 is Up, ifindex: 2, vrf_id 0
  address(es):
    Mac Address: 00:50:56:b0:2f:50
    INET6: FE80::250:56FF:FEB0:2F50 , subnet is FE80::/64
    INET6: 2000::250:56FF:FEB0:2F50 [ PRE ], subnet is 2000::/64
      valid lifetime 2591906 sec, preferred lifetime 604706 sec
  Joined group address(es):
    FF01::1
    FF02::1
    FF02::2
```

```
    FF02::1:FF00:0
    FF02::1:FFB0:2F50
  MTU is 1500 bytes
  ICMP error messages limited to one every 100 milliseconds
  ICMP redirects are enabled
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND stale time is 3600 seconds
  ND advertised reachable time is 0 milliseconds
  ND retransmit interval is 1000 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 600 seconds<480--720>
  ND router advertisements live for 1800 seconds
  Hosts use stateless autoconfig for addresses.
```

On device A, run the **ping** command to test the interconnectivity with device B.

```
DeviceA# ping ipv6 2000::250:56FF:FEB0:2F50
Sending 5, 100-byte ICMP Echoes to 2000::250:56ff:feb0:2f50, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/13/63 ms.
```

On device B, run the **ping** command to test the interconnectivity with device A.

```
DeviceAB# ping ipv6 2000::1
Sending 5, 100-byte ICMP Echoes to 2000::1, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface gigabitEthernet 0/1
ipv6 enable
ipv6 address 2000::1/64
unset ipv6 nd ra-send-disable
!
```

- Device B configuration file

```
hostname DeviceB
!
interface gigabitEthernet 0/1
ipv6 address auto-config
!
```

Contents

1 Configuring DHCPv6 1

1.1 Introduction 1

1.2 Principles 1

1.2.1 Basic Concepts 1

1.2.2 Packet Format 4

1.2.3 Requesting/Allocating Addresses 4

1.2.4 Requesting/Allocating Prefixes 8

1.2.5 Stateless Service 9

1.2.6 DHCPv6 Relay Agent 9

1.2.7 Address Advertisement 11

1.2.8 Protocols and Standards 12

1.3 Restrictions and Guidelines 12

1.4 Configuration Task Summary 12

1.5 Configuring the DHCPv6 Server to Assign Addresses 12

1.5.1 Overview 12

1.5.2 Configuration Tasks 12

1.5.3 Configuring the IA_NA Address Prefix to Be Assigned from a DHCPv6 Server to a DHCPv6 Client 13

1.5.4 Configuring the Prefix of a Statically Bound Address to Be Assigned from a DHCPv6 Server to a DHCPv6 Client 13

1.5.5 Configuring a DCHPv6 Server to Assign Prefixes from a Local Prefix Pool 13

1.5.6 Configuring Excluded Addresses on a DHCPv6 Server 14

1.5.7 Enabling the DHCPv6 Server Service 14

1.6 Configuring the DHCPv6 Server to Assign Other Network Parameters 15
1.6.1 Overview 15
1.6.2 Configuration Tasks 15
1.6.3 Configuring the DNS Address to Be Assigned from a DHCPv6 Server to a DHCPv6 Client 15
1.6.4 Configuring the Domain Name to Be Assigned from a DHCPv6 Server to a DHCPv6 Client 15
1.6.5 Configuring the Boot File URL to Be Assigned from a DHCPv6 Server to a DHCPv6 Client 16
1.6.6 Configuring Option 52 on a DHCPv6 Server 16
1.7 Configuring the DHCPv6 Relay 16
1.7.1 Overview 16
1.7.2 Restrictions and Guidelines 16
1.7.3 Procedure 17
1.8 Configuring Source Interface Designation Function on a DHCPv6 Relay Agent 18
1.8.1 Overview 18
1.8.2 Restrictions and Guidelines 18
1.8.3 Procedure (Global Configuration Mode) 18
1.8.4 Procedure (Interface Configuration Mode) 18
1.9 Monitoring 19
1.10 Configuration Examples 20
1.10.1 Dynamically Assigning IPv6 Addresses 20
1.10.2 Dynamically Assigning IPv6 Address Prefixes 22
1.10.3 Configuring DHCPv6 Relay 25

1 Configuring DHCPv6

1.1 Introduction

The Dynamic Host Configuration Protocol for IPv6 (DHCPv6) is a protocol that allows a DHCPv6 server to transfer configurations to IPv6 nodes, such as IPv6 address, domain name server (DNS) address, network information service (NIS) server address, and Simple Network Time Protocol (SNTP) server address.

Compared with other IPv6 address assignment methods, such as manual configuration and Stateless Address Autoconfiguration (SLAAC), DHCPv6 has the following advantages:

- DHCPv6 provides the address assignment, prefix delegation (PD), and configuration parameter allocation functions.
- DHCPv6 is a Stateful Address Autoconfiguration (SAAC) protocol for flexibly adding and reusing network addresses. It can record assigned addresses, which enhances network manageability.
- By using DHCPv6 PD, uplink network devices can assign address prefixes to downlink network devices, ensuring flexible station-level automatic configuration and flexible control of station address space.
- DHCPv6 configuration parameter allocation solves the problem that parameters cannot be obtained through the SLAAC protocol and allows a DHCPv6 server to assign DNS addresses and domain names to hosts.
- DHCPv6 is a protocol based on the client/server model. A DHCPv6 client is used to obtain various configurations, whereas a DHCPv6 server is used to provide various configurations. If the DHCPv6 client and DHCPv6 server are not on the same network link, they can interact with each other by using a DHCPv6 relay.

1.2 Principles

1.2.1 Basic Concepts

1. IPv6 address assignment methods

- Manual configuration: IPv6 addresses/prefixes and other network configuration parameters are manually configured.
- SLAAC: A host uses the prefix carried in a received Router Advertisement (RA) message and the local interface ID to automatically generate an IPv6 address.
- SAAC (DHCPv6): DHCPv6 includes the following three types:
  - DHCPv6 SAAC: A DHCPv6 server automatically configures IPv6 addresses/prefixes and other network configuration parameters.
  - DHCPv6 SLAAC: A host automatically generates an IPv6 address by using an RA message. The DHCPv6 server assigns configuration parameters other than IPv6 addresses.
  - DHCPv6 PD: A lower-layer network device applies for a prefix from an upper-layer network device, and the upper-layer network device assigns a proper address prefix to the lower-layer network device. The lower-layer network device automatically divides the delegated prefix (less than 64 bits) into subnet segments with a 64-bit prefix and assigns these prefixes to the links directly connected to IPv6 hosts

through RA messages, ensuring automatic address configuration for the hosts.

2. DUID

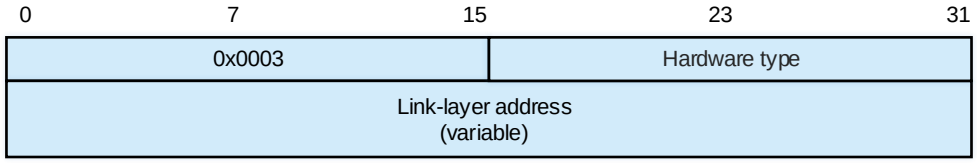
A DHCP Unique Identifier (DUID) uniquely identifies a DHCPv6 device (DHCPv6 client, relay, or server). It is used for mutual authentication during DHCPv6 message exchanges.

RFC 3315 defines three types of DUIDs:

- DUID Based on Link-Layer address plus Time (DUID-LLT)
- DUID Assigned by Vendor Based on Enterprise Number (DUID-EN)
- Link-Layer address (DUID-LL)

DHCPv6 devices use DUID-LLs.

Figure 1-1 DUID-LL Structure



DUID type is set to **0x0003**, **Hardware type** is set to **0x0001** (indicating the supported hardware type of Ethernet), and **Link-layer address** is set to the Media Access Control (MAC) address of the interface with the minimum serial number on the device.

3. Multicast address

In DHCP, clients broadcast DHCP packets to servers. To prevent broadcast storms, DHCPv6 uses multicast packets instead of broadcast packets. DHCPv6 uses the following multicast addresses:

- FF02::1:2: a link-scoped multicast address used by a client to communicate with neighboring relay agents and servers. All DHCPv6 servers and relay agents on the link are members of this multicast group.
- FF05::1:3: a site-scoped multicast address used by a DHCPv6 relay agent to communicate with servers. All DHCPv6 servers within the site are members of this multicast group.

4. Identity association (IA)

A DHCPv6 server uses IAs to assign addresses to DHCPv6 clients. Each IA is uniquely identified by an identity association identifier (IAID). IAIDs are generated by DHCPv6 clients. One-to-one mappings are established between IAs and clients. An IA may contain several addresses, which can be allocated by the client to other interfaces. An IA may contain one of the following types of addresses:

- Non-temporary addresses (NAs), namely, globally unique addresses
- Temporary addresses (TAs), which are hardly used
- PD, prefix delegation

Based on the address type, IAs are classified into IA_NA, IA_TA, and IA_PD types. DHCPv6 servers support only IA_NA and IA_PD.

5. Binding

A DHCPv6 binding is a manageable address information structure. An address binding on a DHCPv6 server records the IA and other configurations of a client. A client can request multiple bindings. The address binding data on a server is presented in the form of an address binding table. Bindings that contain IAs use DUID, IA-Type, or IAID as the index, and bindings that contain configurations use DUID as the index.

6. DHCPv6 address pool

A DHCPv6 server uses address pools to save IP addresses and network parameters to be assigned to clients.

7. Sequence for selecting addresses/prefixes

A DHCPv6 server selects to-be-assigned IPv6 addresses/prefixes based on the following sequence:

(1) Addresses/prefixes previously assigned to the client

(2) Idle address/prefixes in the address pool/prefix pool that meet the client's requirements

(3) Other idle addresses/prefixes in the address pool/prefix pool

(4) Expired or previously conflicted addresses/prefixes

(5) If no address is available, the DHCPv6 server does not assign an address to the client.

8. Address conflict-record

When an address assigned to a DHCPv6 client is conflicted, the client sends a DECLINE packet to notify the DHCPv6 server that the address is rebound. Then, the server adds the address to the address conflict queue. The server does not assign addresses in the address conflict queue to clients, and can display and clear address information in the address conflict queue.

9. Packet type

DHCPv6 uses User Datagram Protocol (UDP) ports 546 and 547 for packet exchanges. Specifically, a DHCPv6 client uses port 546 for receiving packets, while a DHCPv6 server and a DHCPv6 relay agent use port 547 for receiving packets. The following describes types of packets that can be exchanged among a DHCPv6 server, client, and relay agent:

- Packets that may be sent from a DHCPv6 client to a DHCPv6 server include SOLICIT, REQUEST, CONFIRM, RENEW, REBIND, RELEASE, DECLINE, and INFORMATION-REQUEST packets.
- Packets that may be sent from a DHCPv6 server to a DHCPv6 client include ADVERTISE, REPLY, and RECONFIGURE packets.
- Packets that may be sent from a DHCPv6 relay agent to another DHCPv6 relay agent or a DHCPv6 server include RELAY-FORWARD packets.
- Packets that may be sent from a DHCPv6 server or DHCPv6 relay agent to another DHCPv6 relay agent include RELAY-REPLY packets.

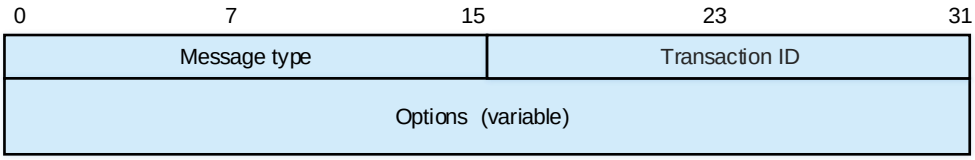
Note

- Devices serving as DHCPv6 servers do not support RECONFIGURE packets.
- Devices serving as DHCPv6 servers do not support CONFIRM and RECONFIGURE packets.

1.2.2 Packet Format

1. Format of Packets Exchanged Between a DHCPv6 Client and a DHCPv6 Server

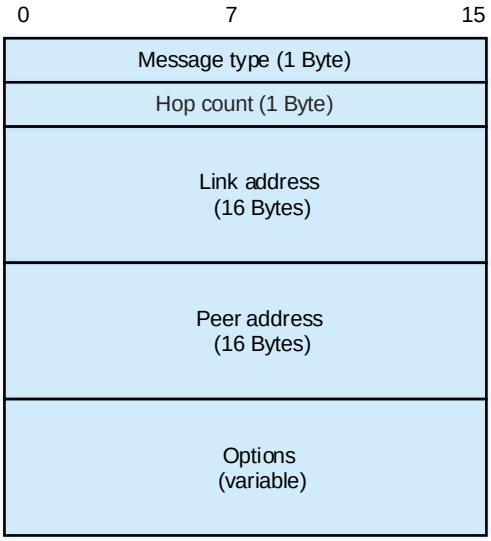
Figure 1-2 Format of Packets Exchanged Between a DHCPv6 Client and a DHCPv6 Server



- **Message type**: Identifies the message type of a DHCPv6 packet.
- **Transaction ID**: Identifies a DHCPv6 packet exchange.
- **Options**: variable-length options carried in a packet.

2. Format of Packets Exchanged Between a DHCPv6 Relay Agent and a DHCPv6 Server

Figure 1-3 Format of Packets Exchanged Between a DHCPv6 Relay Agent and a DHCPv6 Server



- **Message type**: Identifies the message type of a DHCPv6 packet. Only RELAY-FORWARD and RELAY-REPLY packets can be exchanged between a DHCPv6 relay agent and a DHCPv6 server.
- **Hop count**: Number of relays that a packet passes through.
- **Link address**: Used by a DHCPv6 server to identify the access link of a client. It is a global IPv6 unicast address or link-local address.
- **Peer address**: Source address of a relayed packet, that is, the address of a client or a relay agent.
- **Options**: Variable-length options. The **Relay message** option is mandatory, and a relay agent can add other options.

1.2.3 Requesting/Allocating Addresses

A DHCPv6 client can request an IPv6 address from a DHCPv6 server.

After being configured with available addresses, a DHCPv6 server can assign IPv6 addresses to hosts in the network and record the assigned addresses to improve the network manageability.

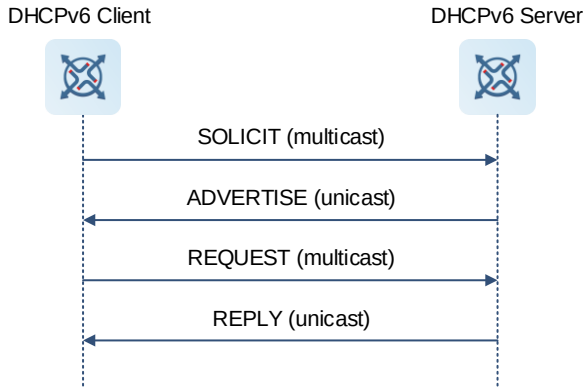
1. Principles

Network hosts serve as DHCPv6 clients and DHCPv6 servers to implement address assignment, update, confirmation, and release through message exchanges.

2. Four-Way Message Exchange

Generally, a DHCPv6 client and a DHCPv6 server implement address, prefix, and parameter configuration through a four-way message exchange.

Figure 1-4 Process of Four-Way Message Exchange



(1) A DHCPv6 client sends a SOLICIT message with the destination address FF02::1:2 and destination port 547 on the local link in multicast mode to request an address, a prefix, and configuration parameters. All DHCPv6 servers or DHCPv6 relay agents on the link receive this SOLICIT message.

(2) After receiving the SOLICIT message, the DHCPv6 server sends an ADVERTISE message in unicast mode if it can provide the information requested in the SOLICIT message. The ADVERTISE message includes the requested address, prefix, and configuration parameters.

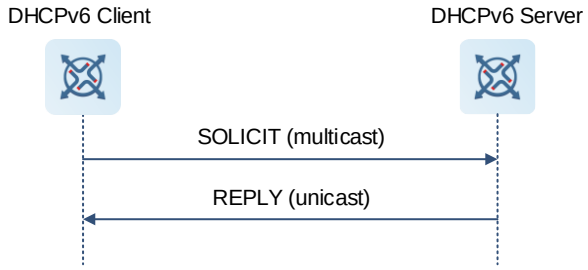
(3) The DHCPv6 client may receive the ADVERTISE messages from multiple DHCPv6 servers. Generally, the DHCPv6 client selects the server from which it first receives the ADVERTISE message as the DHCPv6 server. Then, the DHCPv6 client sends a REQUEST message with the destination address FF02::1:2 and destination port 547 to request an address, a prefix, and configuration parameters.

(4) After receiving the REQUEST message, the DHCPv6 server creates a binding locally and sends a REPLY message in unicast mode. The REPLY message includes the address, prefix and configuration parameters that the DHCPv6 server assigns to the DHCPv6 client. The DHCPv6 client completes address, prefix, or parameter configuration based on the information in the REPLY message.

3. Two-Way Message Exchange

The two-message exchange can be used to complete address, prefix, and parameter configuration for DHCPv6 clients more quickly.

Figure 1-5 Two-Way Message Exchange



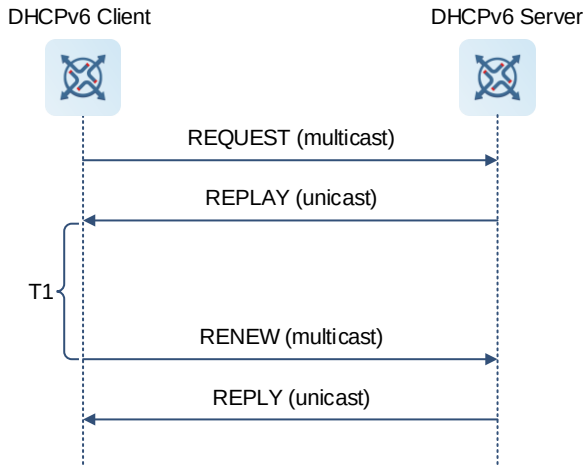
(1) A DHCPv6 client sends a SOLICIT message with the destination address FF02::1:2 and destination port 547 on the local link to request an address, a prefix, and configuration parameters. The SOLICIT message contains the **Rapid Commit** option.

(2) If a DHCPv6 server supports the **Rapid Commit** option, the DHCPv6 server creates a binding locally and sends a REPLY message in unicast mode. The REPLY message includes the address, prefix, and configuration parameters to be assigned to the DHCPv6 client. The DHCPv6 client completes configuration based on the information in the REPLY message. If the DHCPv6 server does not support the **Rapid Commit** option, that is, it does not support rapid assignment, the process of four-way message exchange is used to assign an IPv6 address/a prefix and other network configuration parameters to the client.

4. Update and Rebinding

The DHCPv6 server provides the control address and the updated T1 and T2 in the IA of the REPLY message sent to the DHCPv6 client.

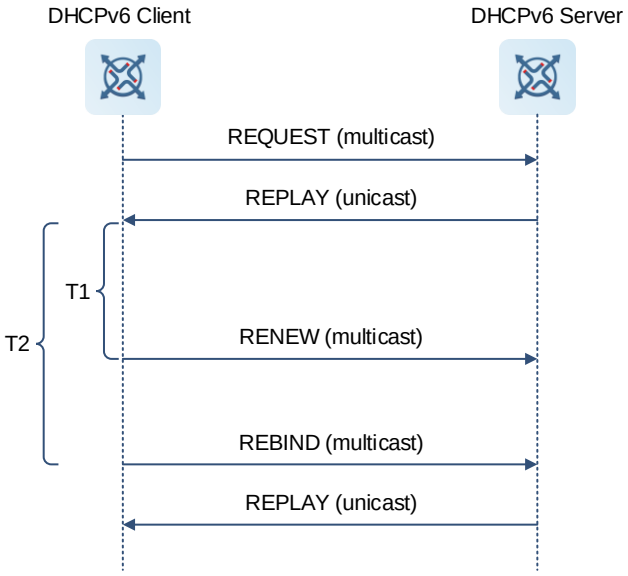
Figure 1-6 Update



(1) The DHCPv6 client sends a RENEW multicast message to the DHCPv6 server for updating the address and prefix after T1 seconds. The RENEW message contains the DUID of the DHCPv6 server and the IA information to be updated.

(2) After receiving the RENEW message, the DHCPv6 server checks whether the DUID value in the RENEW message is the same as its DUID. If yes, the DHCPv6 server updates the local binding and sends a REPLY message in unicast mode. The REPLY message contains the new T1 and other parameters.

Figure 1-7 Rebinding



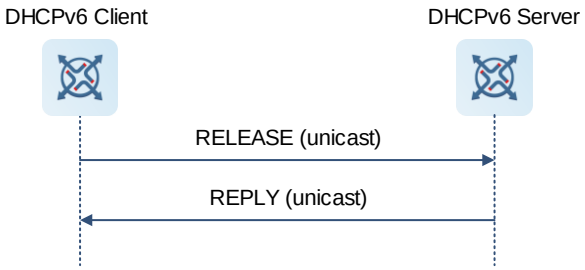
(1) If no response is received within T2 after the DHCPv6 client sends a RENEW message to the DHCPv6 server, the DHCPv6 client sends a REBIND multicast message to the DHCPv6 server for rebinding the address and prefix.

(3) After receiving the REBIND message, the DHCPv6 server (perhaps a new DHCPv6 server) sends a REPLY message based on the content of the REBIND message.

5. Release

If a DHCPv6 client needs to release an address or a prefix, the DHCPv6 client needs to notify the DHCPv6 server of the to-be-released address or prefix. In this way, the DHCPv6 server can assign the address or prefix to other DHCPv6 clients.

Figure 1-8 Release



(1) A DHCPv6 client sends a RELEASE message to a DHCPv6 server to notify the DHCPv6 server that it no longer needs to use an address or prefix, that is, to release the address or prefix.

(2) After receiving RELEASE message, the DHCPv6 server removes the corresponding binding based on the address or prefix in the RELEASE message and sends a REPLY message carrying the state option to the DHCPv6 client.

6. Confirmation

After moving to a new link or encountering a restart, a DHCPv6 client needs to confirm whether the original address is still available.

Figure 1-9 Confirmation



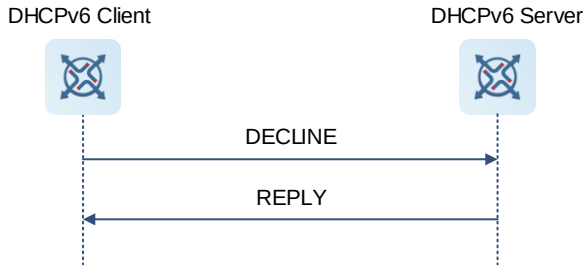
(1) The DHCPv6 client sends a CONFIRM message to the DHCPv6 server on the new link to check whether the original address is still available.

(2) After receiving the CONFIRM message, the DHCPv6 server performs confirmation based on the address information contained in the CONFIRM message and sends a REPLY message carrying the state option to the DHCPv6 client. If the confirmation indicates unavailability, the DHCPv6 client may initiate a new address assignment request.

7. Address conflict-record

If a DHCPv6 client finds that the assigned address has been occupied on the link, the DHCPv6 client sends a DECLINE message to notify the DHCPv6 server of the address conflict.

Figure 1-10 Address conflict-record



(1) The DHCPv6 client sends a DECLINE message to the DHCPv6 server. The DECLINE message includes the IA information of the conflicted address.

(2) After receiving the DECLINE message, the DHCPv6 server marks the address in the DECLINE message as **Declined** and does not assign the address to other clients. Then, the DHCPv6 server sends a REPLY message carrying the state option to the DHCPv6 client. You can manually clear addresses marked as **Declined** to facilitate re-assignment.

1.2.4 Requesting/Allocating Prefixes

After available prefixes are configured on a DHCPv6 server, uplink network devices can assign the address prefixes to downlink network devices by serving as the PD of the DHCPv6 server. This ensures flexible station-level automatic address configuration and flexible control of station address space.

Downlink network devices serve as DHCPv6 clients to exchange messages with the DHCPv6 server to implement prefix assignment, update, and release. Downlink network devices obtain, update, rebind, and release prefixes by using the four-way or two-way message exchange mechanism similar to that for assigning addresses. However, prefix assignment is different from address assignment in the following aspects:

- In message exchanges via PD, the CONFIRM and DECLINE messages are not used.
- If a DHCPv6 client connects to a new link and needs to check whether the prefix is available, it performs confirmation by exchanging the REBIND and REPLY messages.
- These messages use IA_PD.

Note

For message exchanges via PD, see Requesting/Allocating Addresses

1.2.5 Stateless Service

When a DHCPv6 client only needs configuration parameters, the DHCPv6 stateless service can be used to obtain related configuration parameters. This solves the problem that configuration parameters such as the DNS address cannot be obtained through the SLAAC protocol.

Figure 1-11 Message Exchange Using the Stateless Service



(1) A DHCPv6 client sends an INFORMATION-REQUEST message to a DHCPv6 server to request stateless information. Usually, this message does not contain the DUID of a specific DHCPv6 server.

(2) The DHCPv6 server sends a REPLY message containing the configuration parameters to the DHCPv6 client.

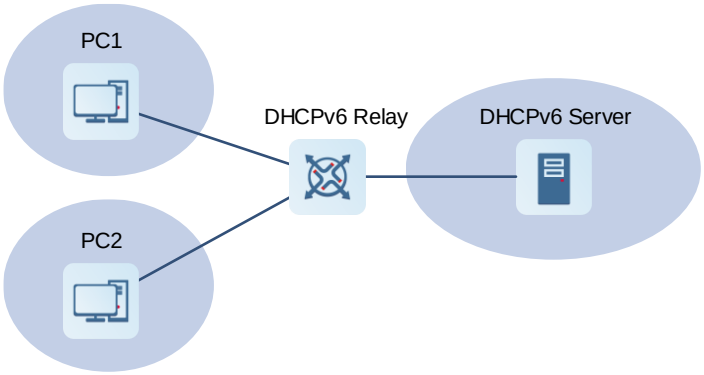
(3) The DHCPv6 client checks whether information provided in the REPLY message is consistent with configuration parameters requested in the INFORMATION-REQUEST message. If yes, the DHCPv6 client performs network configuration based on parameters provided in the REPLY message. Otherwise, it ignores the message. If the DHCPv6 client receives multiple REPLY messages that match the request, the DHCPv6 client completes stateless configuration based on parameters provided in the first received REPLY message.

1.2.6 DHCPv6 Relay Agent

A DHCPv6 client usually discovers a DHCPv6 server by sending Solicit packets with a multicast address as the destination address. Therefore, the DHCPv6 client and DHCPv6 server must be able to directly communicate with each other, that is, they must be deployed on the same link. This may cause inconvenience to management and upgrade and resource waste. The DHCPv6 relay agent function can solve these problems by

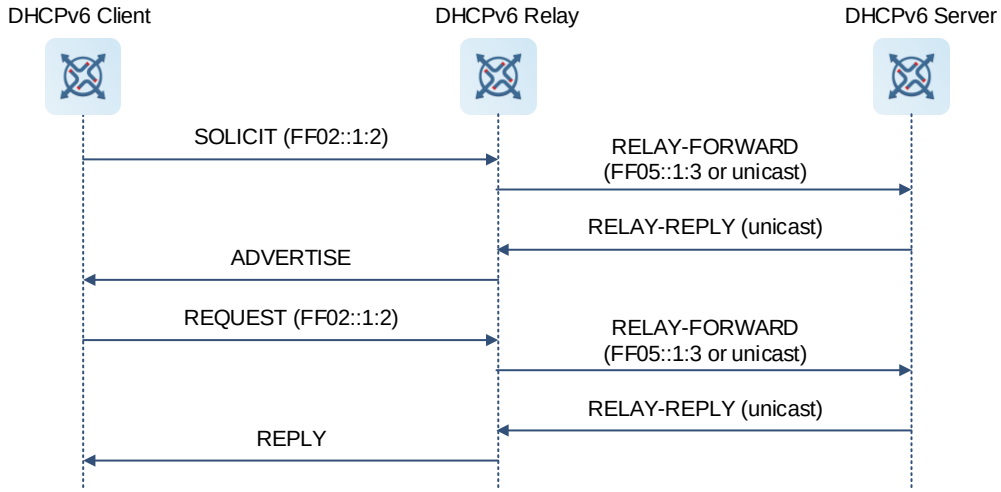
enabling a DHCPv6 client to send packets to a DHCPv6 server on a different link. A DHCP relay agent is often deployed on the link where a DHCPv6 client resides. It is used to forward interaction packets between the DHCPv6 client and a DHCPv6 server. The DHCP relay agent is transparent to the DHCPv6 client.

Figure 1-12 DHCPv6 Relay Agent



2. Principles

Figure 1-13 Packet Exchange Among a DHCPv6 Client, DHCPv6 Relay Agent, and DHCPv6 Server



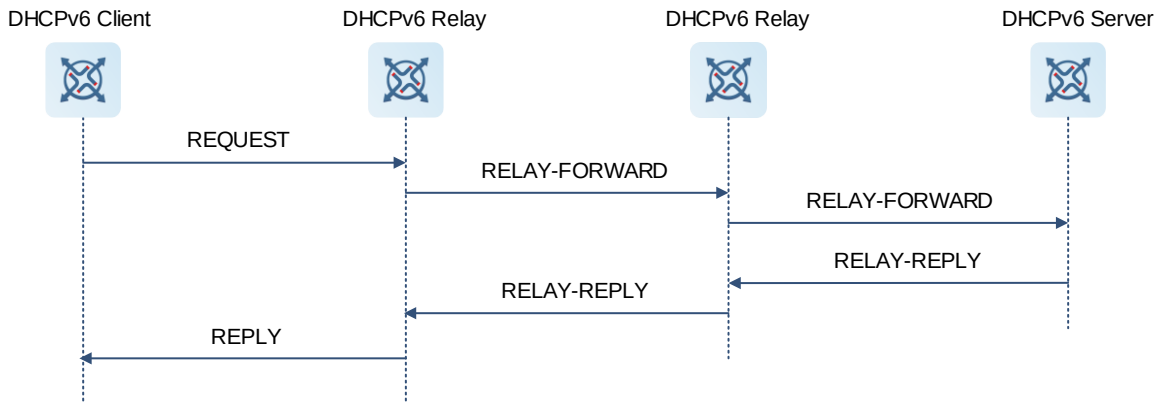
(1) When receiving a SOLICIT or REQUEST message from the DHCPv6 client, a DHCPv6 relay agent creates a RELAY-FORWARD message. This message contains the original message from the DHCPv6 client and multiple options added by the relay agent. Then, the DHCP relay agent sends the RELAY-FORWARD message to a specified DHCPv6 server or a specific multicast address FF05::1:3.

(2) After receiving the RELAY-FORWARD message, the DHCPv6 server extracts the original message for processing. Then, the DHCPv6 server constructs a response to the original message, encapsulates the response in a RELAY-REPLY message, and sends the RELAY-REPLY message to the DHCPv6 relay agent.

(3) After receiving the RELAY-REPLY message, the DHCPv6 relay agent extracts the original message from the DHCPv6 server and forwards it to the DHCPv6 client.

3. Multi-Level DHCPv6 Relay Agents

Figure 1-14 Multi-Level DHCPv6 Relay Agents



A DHCPv6 relay agent encapsulates and decapsulates messages between a DHCPv6 client and a DHCPv6 server on different links to enable communication between them. Multi-level relay agents are allowed between a DHCPv6 client and a DHCPv6 server.

4. Options Supported by a DHCPv6 Relay Agent

- Interface-ID (Option 18)

  A DHCPv6 relay agent can use **Interface-ID** to identify the interface that receives messages from the client. A DHCPv6 server configures parameters based on the value of **Interface-ID** and copies **Interface-ID** to the RELAY-REPLY message to be sent to the DHCPv6 relay agent. After receiving the RELAY-REPLY message, the DHCPv6 relay agent forwards the message through the interface identified by **Interface-ID**. **Interface-ID** exists only in RELAY-FORWARD and RELAY-REPLY packets.

- Remote-ID (Option 37)

  A DHCPv6 relay agent can use **Remote-ID** to carry the DUID, port, and virtual local area network (VLAN) information. Based on the value of **Remote-ID**, a DHCPv6 server performs address assignment, parameter configuration, and prefix delegation.

- Subscriber-ID (Option 38)

  A DHCPv6 relay agent can use **Subscriber-ID** to carry a client's MAC address and other physical information. Based on the value of **Subscriber-ID**, a DHCPv6 server performs address assignment, parameter configuration, and prefix delegation.

1.2.7 Address Advertisement

A DHCPv6 server advertises a client's IPv6 address and MAC address to the authentication module for auditing.

In a subnet without a DHCPv6 relay, a DHCPv6 server can be configured with the IA_NA address assignment and advertisement functions. When receiving a DHCPv6 packet from a DHCPv6 client, the DHCPv6 server can identify the client's MAC address based on the source MAC address in the packet. When assigning an IPv6 address to the client or renewing the lease of this address, the DHCPv6 server advertises the MAC address and IPv6 address to the authentication module. The authentication module integrates the address information with

IPv4 authentication information of the client and synchronizes the information to the authentication server. The authentication server authenticates or records the client based on the information.

1.2.8 Protocols and Standards

- RFC 3315: Dynamic Host Configuration Protocol for IPv6
- RFC 3633: IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) Version 6
- RFC 3646: DNS Configuration Options for Dynamic Host Configuration Protocol for IPv6 (DHCPv6)
- RFC 3736: Stateless DHCP Service for IPv6
- RFC 5417: Control And Provisioning of Wireless Access Points (CAPWAP) Access Controller DHCP Option

1.3 Restrictions and Guidelines

- To provide the DHCPv6 service, specify a DHCPv6 address pool.
- The DHCPv6 service can be configured only on a Switch Virtual Interface (SVI), L3 Ethernet interface, or L3 aggregate port (AP).

1.4 Configuration Task Summary

DHCPv6 configuration includes the following tasks:

- Configuring the DHCPv6 server
  a Configuring the DHCPv6 Server to Assign Addresses
  b (Optional) Configuring the DHCPv6 Server to Assign Other Network Parameters
- Configuring the DHCPv6 relay
  a Configuring the DHCPv6 Relay
  b (Optional) Configuring Source Interface Designation Function on a DHCPv6 Relay Agent

1.5 Configuring the DHCPv6 Server to Assign Addresses

1.5.1 Overview

Configure the device as a DHCPv6 server to assign IPv6 addresses or prefixes to DHCPv6 clients.

1.5.2 Configuration Tasks

(1) Configuring the DHCPv6 Server to Assign Addresses Perform either of the following configuration tasks.
  - Configuring the IA_NA Address Prefix to Be Assigned from a DHCPv6 Server to a DHCPv6 Client
  - Configuring the Prefix of a Statically Bound Address to Be Assigned from a DHCPv6 Server to a DHCPv6 Client
  - Configuring a DCHPv6 Server to Assign Prefixes from a Local Prefix Pool

(2) (Optional) Configuring Excluded Addresses on a DHCPv6 Server

(3) Enabling the DHCPv6 Server Service

1.5.3 Configuring the IA_NA Address Prefix to Be Assigned from a DHCPv6 Server to a DHCPv6 Client

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Create a DHCPv6 address pool and enter the DHCPv6 address pool configuration mode.

ipv6 dhcp-server pool *pool-name*

No DHCPv6 address pool is configured by default.

(4) Configure the IA_NA address prefix to be assigned from a DHCPv6 server to a DHCPv6 client.

iana-address prefix *ipv6-address/prefix-length* [**forever** { *valid-lifetime* | **infinite** } { *preferred-lifetime* | **infinite** } }]

No IA_NA address prefix is configured by default.

1.5.4 Configuring the Prefix of a Statically Bound Address to Be Assigned from a DHCPv6 Server to a DHCPv6 Client

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Create a DHCPv6 address pool and enter the DHCPv6 address pool configuration mode.

ipv6 dhcp-server pool *pool-name*

No DHCPv6 address pool is configured by default.

(4) Configure the prefix of a statically bound address to be assigned from a DHCPv6 server to a DHCPv6 client.

prefix-delegation *ipv6-address*/*prefix-length client*-DUID [**forever** { *valid-lifetime* | **infinite** } { *preferred-lifetime* | **infinite** }]

No static address prefix is configured by default.

1.5.5 Configuring a DCHPv6 Server to Assign Prefixes from a Local Prefix Pool

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a local prefix pool for the PD service of a DHCPv6 server.

ipv6 local-prefix pool *pool-name ipv6-prefix*/*prefix-length assigned-length*

No local prefix pool is configured for the PD service of a DHCPv6 server by default.

(4) Create a DHCPv6 address pool and enter the DHCPv6 address pool configuration mode.

ipv6 dhcp-server pool *pool-name*

(5) Configure the local prefix pool associated with the DHCPv6 server.

prefix-delegation pool *pool-name* [**forever** { *valid-lifetime* | **infinite** } { *preferred-lifetime* | **infinite** }]

No local prefix pool associated with the DHCPv6 server is configured by default.

1.5.6 Configuring Excluded Addresses on a DHCPv6 Server

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Create a DHCPv6 address pool and enter the DHCPv6 address pool configuration mode.

ipv6 dhcp-server pool *pool-name*

No DHCPv6 address pool is configured by default.

(4) Configure excluded addresses on a DHCPv6 server.

excluded-address *start-ipv6-address* [*end-ipv6-address*]

No excluded address is configured on a DHCPv6 server by default.

1.5.7 Enabling the DHCPv6 Server Service

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

interface vlan *interface-number*

(4) Enable the DHCPv6 Server service.

ipv6 dhcp-server server *pool-name* [**rapid-commit**] [**preference** *preference-value*]

The DHCPv6 Server service is disabled by default.

1.6 Configuring the DHCPv6 Server to Assign Other Network Parameters

1.6.1 Overview

A DHCPv6 server can automatically assign configuration parameters, such as the DNS address, domain name, boot file, and Option 52 to DHCPv6 clients.

The DHCPv6 server uses Option 52 to specify the IPv6 address of a Control and Provisioning of Wireless Access Points (CAPWAP) access controller (AC).

1.6.2 Configuration Tasks

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring the DNS Address to Be Assigned from a DHCPv6 Server to a DHCPv6 Client
- Configuring the Domain Name to Be Assigned from a DHCPv6 Server to a DHCPv6 Client
- Configuring the Boot File URL to Be Assigned from a DHCPv6 Server to a DHCPv6 Client
- Configuring Option 52 on a DHCPv6 Server

1.6.3 Configuring the DNS Address to Be Assigned from a DHCPv6 Server to a DHCPv6 Client

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the DHCPv6 address pool configuration mode.

ipv6 dhcp-server pool *pool-name*

No DHCPv6 address pool is configured by default.

(4) Configure the DNS address to be assigned from a DHCPv6 server a DHCPv6 client.

dns-server *ipv6-address*

No DNS address is configured by default.

1.6.4 Configuring the Domain Name to Be Assigned from a DHCPv6 Server to a DHCPv6 Client

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the DHCPv6 address pool configuration mode.

ipv6 dhcp-server pool *pool-name*

No DHCPv6 address pool is configured by default.

(4) Configure the domain name to be assigned from a DHCPv6 to a DHCPv6 client.

domain-name *domain-name*

No domain name is configured by default.

1.6.5 Configuring the Boot File URL to Be Assigned from a DHCPv6 Server to a DHCPv6 Client

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the DHCPv6 address pool configuration mode.

ipv6 dhcp-server pool *pool-name*

No DHCPv6 address pool is configured by default.

(4) Configure the boot file Uniform Resource Locator (URL) to be assigned from a DHCPv6 server to a DHCPv6 client.

bootfile-url *url-string*

No boot file URL is configured by default.

1.6.6 Configuring Option 52 on a DHCPv6 Server

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the DHCPv6 address pool configuration mode.

ipv6 dhcp-server pool *pool-name*

No DHCPv6 address pool is configured by default.

(4) Configure Option 52 on a DHCPv6 server.

option52 *ipv6-address*

Option 52 is not configured on a DHCPv6 server by default.

1.7 Configuring the DHCPv6 Relay

1.7.1 Overview

A DHCPv6 relay agent can be configured for address assignment, prefix delegation, and parameter allocation between a DHCPv6 client and a DHCPv6 server on different links.

1.7.2 Restrictions and Guidelines

A destination address must be configured for the device with the DHCPv6 relay function enabled. If the destination address is a multicast address (such as FF05::1:3), an egress interface needs to be configured.

- When a Layer 3 interface is added to or removed from VRF, the existing destination address or interface of DHCPv6 relay will be deleted. After the Layer 3 interface is added to or removed from VRF, enable the DHCPv6 relay function and specify the destination address or interface according to specific requirements.
- When VRF is deleted, the destination address or interface of DHCPv6 relay on the Layer 3 interface added to the VRF will be deleted.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the format of **Interface-ID** on a DHCPv6 relay agent.

ipv6 dhcp-server relay-agent option interface-id format user-defined *text*

The interface name is specified in **Interface-ID** on a DHCPv6 relay agent by default.

(4) (Optional) Configure the format of the MAC address in user-defined options on a DHCPv6 relay agent.

ipv6 dhcp-server relay-agent option mac-str-format *type*

The default MAC address format is H.H.H.

(5) (Optional) Add **the Remote-ID** to DHCPv6 relay packets.

ipv6 dhcp-server relay-agent option remote-id enable

DHCPv6 relay packets do not carry **Remote-ID** by default.

(6) (Optional) Configure the format of **Remote-ID** in DHCPv6 relay packets.

ipv6 dhcp-server relay-agent option remote-id format user-defined *text*

The device DUID is specified for **Remote-ID** in DHCPv6 relay packets by default.

(7) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

interface vlan *interface-number*

(8) Enable the DHCPv6 Relay function.

ipv6 dhcp-server relay-agent destination [**vrf** *vrf-name*] *ipv6-address* [*interface-type interface-number*]

The DHCPv6 Relay function is disabled by default. No destination address or interface over which packets are routed to the destination address is configured.

1.8 Configuring Source Interface Designation Function on a DHCPv6 Relay Agent

1.8.1 Overview

In some device redundancy scenarios, a DHCPv6 server regards uplink interface IP addresses of multiple DHCPv6 relay agents as one IP address. As a result, the DHCPv6 server cannot identify unique DHCPv6 relay agents by their uplink interface IP addresses. The source interface can be designated to differentiate between different DHCPv6 relay agents to ensure normal packet forwarding.

The source interface designation function allows designating an IP address or interface. When a DHCPv6 relay agent forwards a RELAY-FORWARD packet to a DHCPv6 server, specified information is filled in the source IP address and link address fields of the RELAY-FORWARD packet but the destination IP address is still the IP address of the DHCPv6 server. After receiving the RELAY-FORWARD packet, the DHCPv6 server records the source IP address of the packet and uses it as the destination address of the RELAY-REPLY packet. That is, the response packet is destined for the IP address or interface specified by using the source interface designation function of the relay agent, so as to bypass the uplink interface of the DHCPv6 relay agent.

1.8.2 Restrictions and Guidelines

The source interface designation function can be configured in global configuration mode and interface configuration mode. The source interface designation type in interface configuration mode is prior to that in global configuration mode. In one configuration mode, the last configured source interface designation type prevails.

1.8.3 Procedure (Global Configuration Mode)

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the source interface designation function.

ipv6 dhcp-server relay-agent source { **gateway-address** | **source-ip-address** } { *interface-type interface-number* | *ipv6-address* }

The source interface designation function is not configured by default.

1.8.4 Procedure (Interface Configuration Mode)

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

(4) Configure the source interface designation function.

ipv6 dhcp-server relay-agent source { **gateway-address** | **source-ip-address** } { *interface-type interface-number* | *ipv6-address* }

The source interface designation function is not configured by default.

1.9 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to outputting debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** command to clear information

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup ipv6 dhcp-server binding-database [*ipv6-address*]	Clears bindings on a DHCPv6 server.
cleanup ipv6 dhcp-server server statistics	Clears statistics of different types of packets on a DHCPv6 server.
cleanup ipv6 dhcp-server conflict-record { *ipv6-address* | * }	Clears conflicted addresses on a DHCPv6 server.

Command	Purpose
cleanup ipv6 dhcp-server relay-agent statistics	Clears statistics of different types of packets on a DHCPv6 relay agent.
display ipv6 dhcp-server	Displays the DUID of a DHCPv6 device.
display ipv6 dhcp-server binding-database [*ipv6-address*]	Displays address bindings on a DHCPv6 server.
display ipv6 dhcp-server interface [*interface-type interface-number*]	Displays interfaces on a DHCPv6 server.
display ipv6 dhcp-server pool [*pool-name*]	Displays address pools on a DHCPv6 server.
display ipv6 dhcp-server conflict-record	Displays conflicted addresses on a DHCPv6 server.
display ipv6 dhcp-server server statistics	Displays DHCPv6 server statistics.
display ipv6 dhcp-server relay-agent agent { *ipv6-address* | * }	Displays source interfaces on a DHCPv6 relay agent.
display ipv6 dhcp-server relay-agent destination { **all** | *interface-type interface-number* }	Displays destination addresses on a DHCPv6 relay agent.
display ipv6 dhcp-server relay-agent source	Displays the source interface designation configuration on a DHCPv6 relay agent.
display ipv6 dhcp-server relay-agent statistics	Displays statistics of different types of packets on a DHCPv6 relay agent.
display ipv6 local-prefix pool [*pool-name*]	Displays local prefix pool configurations and usage on a device.
debugging ipv6 dhcp-server [**detail**]	Debugs DHCPv6.
debugging ipv6 dhcp-server relay-agent	Debugs a DHCPv6 relay agent.
debugging ipv6 dhcp-server server	Debugs a DHCPv6 server.

1.10 Configuration Examples

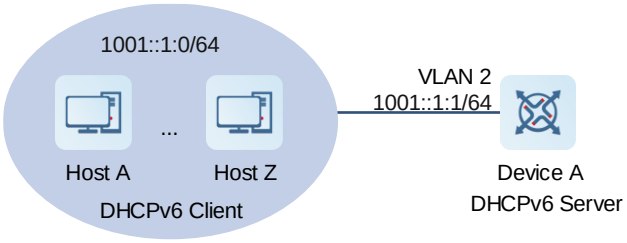
1.10.1 Dynamically Assigning IPv6 Addresses

1. Requirements

As shown in Figure 1-15, a DHCPv6 client in network segment 1001::1:0/64 requests address information from a DHCPv6 server (Device A) in the same subnet. Assignable IPv6 addresses, DNS address, domain name, and other configuration parameters are configured on the DHCPv6 server.

2. Topology

Figure 1-15 Configuring Dynamic IPv6 Address Assignment



3. Notes

Run the DHCPv6 server on Device A and configure addresses and other parameters for automatic address and parameter assignment.

4. Procedure

(1) Configure Device A:

Configure DHCPv6 address pool parameters.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6    dhcp-server pool v6
DeviceA(cfg-dhcp-pool)# iana-address prefix 1001::1:0/64
DeviceA(cfg-dhcp-pool)# excluded-address 1001::1:1 1001::1:2
DeviceA(cfg-dhcp-pool)# dns-server 1001::1:2
DeviceA(cfg-dhcp-pool)# domain-name example.com
```

Configure an interface address and enable the DHCPv6 Server service.

```
DeviceA(cfg)# interface vlan 2
DeviceA(cfg-if-VLAN 2)# ipv6 enable
DeviceA(cfg-if-VLAN 2)# ipv6 address 1001::1:1/64
DeviceA(cfg-if-VLAN 2)# ipv6    dhcp-server server v6
```

Cancel suppression of RA messages released on the device.

```
DeviceA(cfg-if-VLAN 2)# unset ipv6 nd suppress-ra
```

Set the configuration flag of managed addresses to **1**, that is, enable hosts to obtain IPv6 addresses from the DHCPv6 server.

```
DeviceA(cfg-if-VLAN 2)# ipv6 nd managed-config-flag
```

Set the configuration flags of other information to **1**, that is, enable hosts to obtain other information except IPv6 addresses from the DHCPv6 server.

```
DeviceA(cfg-if-VLAN 2)# ipv6 nd other-config-flag
```

(2) Configure the hosts:

Enable the dynamic address acquisition function. (Omitted)

5. Verification

Check whether the clients have applied for IPv6 addresses in network segment 1001::1:0/64.

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
 ipv6    dhcp-server pool v6
 iana-address prefix 1001::1:0/64
 excluded-address 1001::1:1 1001::1:2
 dns-server 1001::1:2
 domain-name example.com
!
 interface vlan 2
 ipv6 address 1001::1:1/64
 ipv6    dhcp-server server v6
!
```

7. Common Errors

- The address pool name exceeds 256 characters.
- The number of configured address pools exceeds 256.
- The number of interfaces configured with the DHCPv6 Server service exceeds 256.
- The number of configured DNS addresses exceeds 10.
- The number of configured domain names exceeds 10.

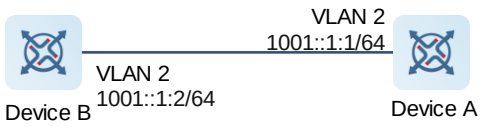
1.10.2 Dynamically Assigning IPv6 Address Prefixes

1. Requirements

As shown in Figure 1-16, Device B is a DHCPv6 client and requests an IPv6 address prefix (2001::1/64), DNS address, domain name, and other network parameters from Device A (a DHCPv6 server).

2. Topology

Figure 1-16 Configuring Dynamic IPv6 Address Prefix Assignment



3. Notes

- Run the DHCPv6 Server service and the PD service on Device A.
- Enable the DHCPv6 Client service on Device B.
- Deploy IPv6 neighbor discovery (ND) between the server and the client to configure host addresses in the subnet through RA messages.

4. Procedure

(1) Configure Device A:

Configure a prefix pool.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)#ipv6 local-prefix pool myprefix 2001::1/64 64
```

Configure DHCPv6 address pool parameters and associate it with the prefix pool.

```
DeviceA(cfg)# ipv6  dhcp-server pool v6_pd
DeviceA(cfg-dhcp-pool)# dns-server 1001::1:2
DeviceA(cfg-dhcp-pool)# domain-name example.com
DeviceA(cfg-dhcp-pool)# prefix-delegation pool myprefix
DeviceA(cfg-dhcp-pool)# exit
```

Configure an interface address and enable the DHCPv6 Server service.

```
DeviceA(cfg)# interface vlan 2
DeviceA(cfg-if-VLAN 2)# ipv6 enable
DeviceA(cfg-if-VLAN 2)# ipv6 address 1001::1:1/64
DeviceA(cfg-if-VLAN 2)# ipv6  dhcp-server server v6_pd
```

Cancel suppression of RA messages released on the device.

```
DeviceA(cfg-if-VLAN 2)# unset ipv6 nd suppress-ra
```

Set the configuration flag of managed addresses to **1**, that is, enable hosts to obtain IPv6 addresses from the DHCPv6 server.

```
DeviceA(cfg-if-VLAN 2)# ipv6 nd managed-config-flag
```

Set the configuration flags of other information to **1**, that is, enable hosts to obtain other information except IPv6 addresses from the DHCPv6 server.

```
DeviceA(cfg-if-VLAN 2)# ipv6 nd other-config-flag
```

(2) Configure Device B:

Configure an interface address.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface vlan 2
DeviceB(cfg-if-VLAN 2)# ipv6 enable
DeviceB(cfg-if-VLAN 2)# ipv6 address 1001::1:2/64
```

Enable the DHCPv6 Client service and request address/prefix information from the DHCPv6 server.

```
DeviceB(cfg-if-VLAN 2)# ipv6  dhcp-server client pd mypd
```

5. Verification

Check whether the DHCPv6 Server function is enabled on VLAN 2 of Device A.

```
DeviceA# display ipv6  dhcp-server interface
VLAN 2 is in count mode
  Server pool: v6_pd
  Rapid-Commit: disable
```

Check whether the DHCPv6 Client function is enabled on VLAN 2 of Device B.

```
DeviceB# display ipv6  dhcp-server interface
```

```
VLAN 2 is in client mode
  State is IDLE
  next packet will be send in : 43049 seconds
  List of known servers:
    DUID: 00:03:00:01:00:50:56:b0:05:98
    Reachable via address: ::
    Preference: 0
    Configuration parameters:
      IA PD: IA ID 0x2, T1 43200, T2 69120
        Prefix: 2001::/64
          preferred forever 86400, valid forever 86400
          expires at Jan 15 2020 16:0 (86249 seconds)
  Prefix name: mypd
  DNS server: 1001::1:2
  Domain name: example.com
  Rapid-Commit: disable
```

Check the common prefix of Device B.

```
DeviceB# display ipv6 general-prefix
There is 1 general prefix.
IPv6 general prefix mypd, acquired via DHCP Prefix Discovery
  2001::/64 valid forever 86368, preferred forever 86368
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
 ipv6 local-prefix pool myprefix 2001::1/64 64
!
 ipv6   dhcp-server pool v6_pd
 dns-server 1001::1:2
 domain-name example.com
 prefix-delegation pool myprefix
!
 interface vlan 2
 ipv6 address 1001::1:1/64
 ipv6   dhcp-server server v6_pd
 unset ipv6 nd suppress-ra
 ipv6 nd managed-config-flag
 ipv6 nd other-config-flag
!
```

- Device B configuration file

```
hostname DeviceB
!
 interface vlan 2
 ipv6 address 1001::1:2/64
```

```
 ipv6    dhcp-server client pd mypd
!
```

7. Common Errors

- The specified address pool name is too long.
- The number of configured address pools exceeds 256.
- The configuration is performed on other interfaces than a SVI, routing port, or L3 AP.
- The number of interfaces configured with the DHCPv6 Server service exceeds 256.
- The number of configured DNS addresses exceeds 10.
- The number of configured domain names exceeds 10.

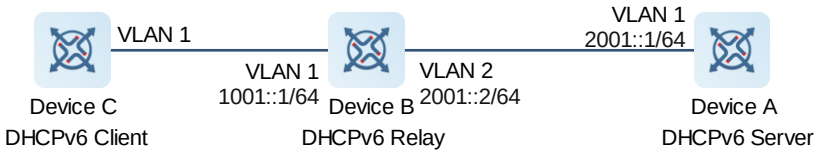
1.10.3 Configuring DHCPv6 Relay

1. Requirements

As shown in Figure 1-17, Device C (a DHCPv6 client) is in network segment 1001::/64, and the address of Device A (a DHCPv6 server) is 2001::1/64. Device B (a DHCPv6 relay agent) provides the relay service for the DHCPv6 client and DHCPv6 server on different links to enable communication between them.

2. Topology

Figure 1-17 Configuring DHCPv6 Relay



3. Notes

- Enable the DHCPv6 Server function on Device A and configure addresses and other parameters.
- Enable the DHCPv6 Relay function on Device B with the destination address pointed to VLAN 1 on Device A.
- Enable the DHCPv6 Client function on Device C.

4. Procedure

(1) Configure Device A:

Configure DHCPv6 address pool parameters.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6    dhcp-server pool v6
DeviceA(cfg-dhcp-pool)# iana-address prefix 1001::/64
DeviceA(cfg-dhcp-pool)# excluded-address 1001::1 1001::2
DeviceA(cfg-dhcp-pool)# dns-server 1001::2
DeviceA(cfg-dhcp-pool)# domain-name example.com
```

Configure an interface address and enable the DHCPv6 Server service.

```
DeviceA(cfg)# interface vlan 1
DeviceA(cfg-if-VLAN 1)# ipv6 enable
DeviceA(cfg-if-VLAN 1)# ipv6 address 2001::1/64
DeviceA(cfg-if-VLAN 1)# ipv6    dhcp-server server v6
```

(2) Configure Device B:

Configure an uplink interface address.

```
DeviceB> privilege
DeviceB#config-system terminal
DeviceB(cfg)# interface vlan 2
DeviceB(cfg-if-VLAN 2)# ipv6 enable
DeviceB(cfg-if-VLAN 2)# ipv6 address 2001::2/64
DeviceB(cfg-if-VLAN 2)# exit
```

Configure a downlink interface address and enable the DHCPv6 Relay function.

```
DeviceB(cfg)#interface vlan 1
DeviceB(cfg-if-VLAN 1)# ipv6 enable
DeviceB(cfg-if-VLAN 1)# ipv6 address 1001::1/64
DeviceB(cfg-if-VLAN 1)# ipv6    dhcp-server relay-agent destination 2001::1
```

Configure the device to disable RA suppression.

```
DeviceB(cfg-if-VLAN 1)# unset ipv6 nd suppress-ra
```

Set the configuration flag bit of the managed address to 1, namely, the device obtains the IPv6 address through the DHCPv6 server.

```
DeviceB(cfg-if-VLAN 1)# ipv6 nd managed-config-flag
```

Set the configuration flag bit of the other information to 1, namely, the device obtains other information than the IPv6 address through the DHCPv6 server.

```
DeviceB(cfg-if-VLAN 1)# ipv6 nd other-config-flag
```

(3) Configure Device C:

Enable the DHCPv6 Client function to apply for an IP address.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)#interface vlan 1
DeviceC(cfg-if-VLAN 1)#ipv6    dhcp-server client ia
```

5. Verification

Check whether Device C has applied for an IPv6 address in network segment 1001::/64.

```
DeviceC# display ipv6 interface

interface VLAN 1 is Up, ifindex: 2, vrf_id 0
  address(es):
    Mac Address: 00:50:56:b0:2f:50
    INET6: 1001::3 [ TENTATIVE ], subnet is 1001::/64
      valid forever 86400 sec, preferred forever 86400 sec
  Joined group address(es):
```

```
    FF01::1
    FF02::1
    FF02::2
    FF02::1:FF00:0
    FF02::1:FF00:3
    FF02::1:FFB0:2F50
  MTU is 1500 bytes
  ICMP error messages limited to one every 100 milliseconds
  ICMP redirects are enabled
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND stale time is 3600 seconds
  ND advertised reachable time is 0 milliseconds
  ND retransmit interval is 1000 milliseconds
  ND advertised retransmit interval is 0 milliseconds
  ND router advertisements are sent every 600 seconds<480--720>
  ND router advertisements live for 1800 seconds
  Hosts use stateless autoconfig for addresses.
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
 ipv6    dhcp-server pool v6
 iana-address prefix 1001::/64
 excluded-address 1001::1 1001::2
 dns-server 1001::2
 domain-name example.com
!
 interface vlan 1
 ipv6 address 2001::1/64
 ipv6    dhcp-server server v6
!
```

- Device B configuration file

```
hostname DeviceB
!
 interface vlan 2
 ipv6 enable
 ipv6 address 2001::2/64
!
 interface vlan 1
 ipv6 address 1001::1/64
 ipv6    dhcp-server relay-agent destination 2001::1
 unset ipv6 nd suppress-ra
 ipv6 nd managed-config-flag
```

```
 ipv6 nd other-config-flag
!
```

- Device C configuration file

```
hostname DeviceC
!
 interface vlan 1
 ipv6    dhcp-server client ia
!
```

Contents

1 Configuring DHCPv6 Client 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.1.3 Protocols and Standards 1

1.2 Restrictions and Guidelines 1

1.3 Configuring DHCPv6 Client to Request an IPv6 Address 1

1.3.1 Overview 1

1.3.2 Restrictions and Guidelines 2

1.3.3 Procedure 2

1.4 Configuring DHCPv6 Client to Request an IPv6 Prefix 2

1.4.1 Overview 2

1.4.2 Restrictions and Guidelines 2

1.4.3 Procedure 2

1.5 Monitoring 3

1 Configuring DHCPv6 Client

1.1 Introduction

1.1.1 Overview

The Dynamic Host Configuration Protocol for IPv6 (DHCPv6) is used to assign IPv6 addresses, IPv6 prefixes, and other network configuration parameters to client hosts.

When a device serves as a DHCPv6 client, the device can:

- Obtain an IPv6 address and network configuration parameters by using DHCPv6. The IPv6 address is used as the address of an interface on which the DHCPv6 Client function is enabled.
- Obtain an IPv6 prefix and network configuration parameters by using DHCPv6. The IPv6 prefix is used by the device to generate an IPv6 address.
- Obtain an IPv6 address, IPv6 prefix, and network configuration parameters by using DHCPv6. The IPv6 address is used as the address of an interface on which the DHCPv6 Client function is enabled, and the IPv6 prefix is used by the device to generate an IPv6 address.
- Obtain network configuration parameters except the IPv6 address and prefix through DHCPv6 stateless configuration.

1.1.2 Principles

DHCPv6 is a protocol based on the client/server model. For details about interaction between a DHCPv6 client and a DHCPv6 server, see *Configuring DHCPv6*.

1.1.3 Protocols and Standards

- RFC 3315: Dynamic Host Configuration Protocol for IPv6
- RFC 3633: IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) Version 6
- RFC 3646: DNS Configuration Options for Dynamic Host Configuration Protocol for IPv6 (DHCPv6)
- RFC 3736: Stateless DHCP Service for IPv6
- RFC 5417: Control And Provisioning of Wireless Access Points (CAPWAP) Access Controller DHCP Option

1.2 Restrictions and Guidelines

You are advised not to enable both the DHCPv6 Client and DHCPv6 Server functions or both the DHCPv6 Client and DHCPv6 Relay functions on the same interface. Otherwise, the functions may be abnormal.

1.3 Configuring DHCPv6 Client to Request an IPv6 Address

1.3.1 Overview

Configure a device as a DHCPv6 client to automatically request an IPv6 address or related parameters from a DHCPv6 server.

1.3.2 Restrictions and Guidelines

The DHCPv6 Client function can be configured only on L3 interfaces.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the switch virtual interface (SVI) configuration mode.

interface vlan *interface-number*

(4) Enable the DHCPv6 Client function and request an identity association non-temporary address (IA_NA) address.

ipv6 dhcp-client enable client ia [**rapid-commit**]

Both the DHCPv6 Client function and IA_NA address requesting are disabled by default.

(5) (Optional) Configure the stateless service.

ipv6 nd other-autoconfig-flag

The stateless service is not configured by default.

1.4 Configuring DHCPv6 Client to Request an IPv6 Prefix

1.4.1 Overview

Configure a device as a DHCPv6 client to automatically request an IPv6 prefix or related parameters from a DHCPv6 server.

1.4.2 Restrictions and Guidelines

The DHCPv6 Client function can be configured only on L3 interfaces.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

interface vlan *interface-number*

(4) Enable the DHCPv6 Client function to request an IPv6 address prefix.

ipv6 dhcp-client enable client pd *prefix-name* [**rapid-commit**]

IPv6 address prefix requesting is not configured on a DHCPv6 client by default.

1.5 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to outputting debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** command to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup ipv6 dhcp-client client reset*interface-type interface-number*	Restarts the DHCPv6 Client function.
display ipv6 dhcp-client	Displays the DHCP Unique Identifier (DUID) of a device.

display ipv6 dhcp-client interface [*interface-type interface-number*]	Displays DHCPv6 interfaces.
debugging ipv6 dhcp-client relay[**detail**]	Debugs DHCPv6.
debugging ipv6 dhcp-client client	Debugs a DHCPv6 client.

Contents

1 Configuring TCP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 6
1.2 Configuration Task Summary 6
1.3 Configuring TCP SYN Packet Timeout Period 6
1.3.1 Overview 6
1.3.2 Restrictions and Guidelines 6
1.3.3 Procedure 7
1.4 Configuring a TCP Window Size 7
1.4.1 Overview 7
1.4.2 Restrictions and Guidelines 7
1.4.3 Procedure 7
1.5 Configuring the Sending of TCP Reset Packets Upon the Receiving of Port Unreachable Messages 7
1.5.1 Overview 7
1.5.2 Restrictions and Guidelines 8
1.5.3 Procedure 8
1.6 Configuring MSS for a TCP Connection 8
1.6.1 Overview 8
1.6.2 Restrictions and Guidelines 8

1.6.3 Procedure....8
1.7 Configuring the TCP Path MTU Discovery Function....9
1.7.1 Overview....9
1.7.2 Restrictions and Guidelines....9
1.7.3 Procedure....9
1.8 Configuring TCP Connection Exception Detection....10
1.8.1 Overview....10
1.8.2 Procedure....10
1.9 Monitoring....10

1 Configuring TCP

1.1 Introduction

1.1.1 Overview

Transmission Control Protocol (TCP) is a connection-oriented, reliable, and byte stream-based transport layer communication protocol. It is defined in RFC 793 of the Internet Engineering Task Force (IETF).

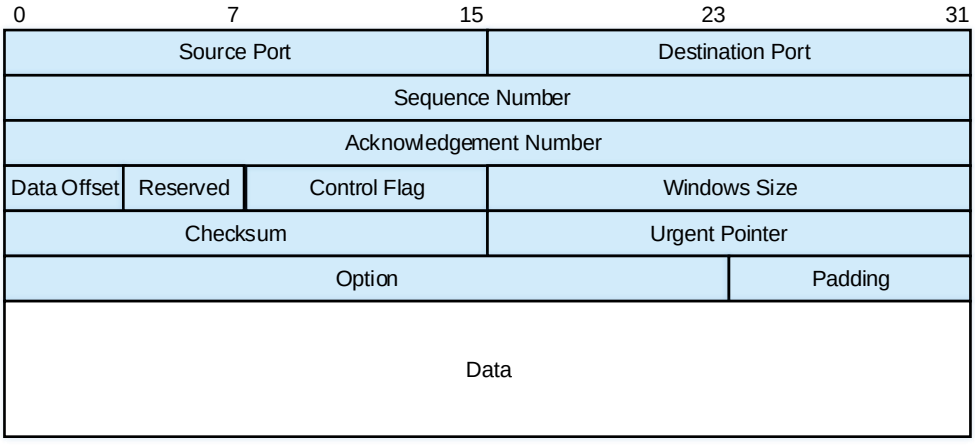
An interconnected network is very different from a single network in that topologies, bandwidth, latency, packet sizes, and other parameters may be entirely different in different parts of the interconnected network. TCP is designed to adapt dynamically to these features of interconnected networks and show robustness in the case of various failures.

- TCP uses a checksum function to check data correctness and validity. The checksum needs to be calculated before transmission and after receiving. Besides, TCP uses the message digest 5 (MD5) authentication to verify data.
- TCP uses sequential transmission, timeout retransmission, and piggyback mechanism to ensure reliability.
- TCP uses the sliding window protocol to control flows. As documented in the protocol, unacknowledged packets in a window need to be retransmitted.
- TCP uses the widely praised TCP congestion control algorithm (also known as the additive increase and multiplicative decrease (AIMD) algorithm) to control congestion.

1.1.2 Principles

1. TCP Header Format

Figure 1-1 TCP Header Format

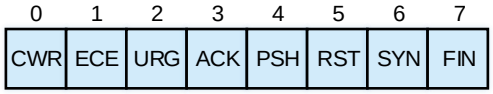


Field descriptions are as follows:

- **Source Port**: Indicates a 16-bit source port number.
- **Destination Port**: Indicates a 16-bit destination port number.

- **Sequence Number**: Indicates a 32-bit sequence number.
- **Acknowledgment Number**: Indicates a 32-bit acknowledgment sequence number.
- **Data Offset**: Indicates a 4-bit data offset. The value is the length of the TCP header (including the option) divided by 4.
- **Control Flag**: Indicates an 8-bit control flag.

Figure 1-2 Control Flag



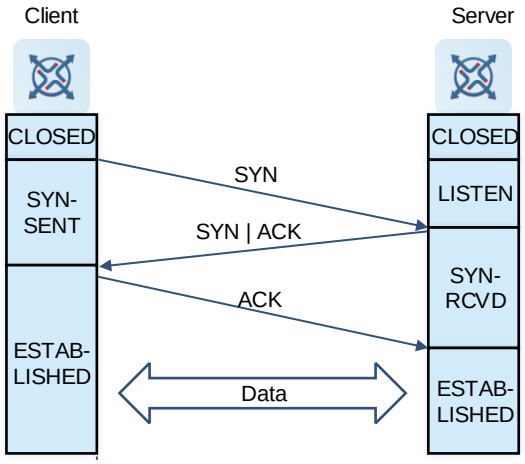
  - Congestion window reduced (**CWR**): The **CWR** flag and the **ECE** flag behind are used for the Explicit Congestion Notification (**ECN**) field in the IP header. When the **ECE** flag is **1**, the peer is notified to make the congestion window smaller.
  - ECN-Echo (**ECE**): The **ECE** flag indicates ECN-Echo. When it is set to **1**, the peer is notified of the congestion from the peer to the local network. When the **ECN** field in the IP header of a received packet is set to **1**, the **ECE** flag in the TCP header is set to **1**.
  - Urgent flag (**URG**): When this bit is set to **1**, the packet contains data that needs urgent processing. Data that needs urgent processing will be further explained in the description of the **Urgent Pointer** field.
  - Acknowledge flag (**ACK**): When this bit is set to **1**, the acknowledgment number is valid. According to the TCP protocol, this bit must be set to **1** except for the SYN packet in the initial establishment of a TCP connection.
  - Push Flag (**PSH**): When this bit is set to **1**, the received data must be immediately transferred to the upper-layer application. When **PSH** is set to **0**, data is buffered instead of being transferred to the upper-layer application immediately.
  - Reset Flag (**RST**): When this bit is set to **1**, the TCP connection must be forcibly closed due to an exception. For example, if a client attempts to establish a TCP connection with a server port that has not provided service externally, the server sends a Reset packet with **RST** value of **1** to the client. In addition, if a device is restarted due to a program crash or power cut, established TCP connections on the device will be closed because all connections will be initialized. In this case, if the peer sends a Reset packet with **RST** set to **1**, the connection is closed.
  - Synchronize Flag (**SYN**): This bit is used for establishing a connection. If **SYN** is set to **1**, a connection is needed and the initial value of the sequence number is set in the sequence number field.
  - Finish Flag (**FIN**): When this bit is set to **1**, no more data will be sent and disconnection is needed. When the communication ends and disconnection is required, both communication hosts can exchange TCP segments with **FIN** set to **1**. The connection can be closed after each host acknowledges the FIN packet from the peer. However, a host does not have to reply with a FIN packet immediately after receiving a TCP packet with **FIN** set to **1**. Instead, it can wait until all data in the buffer has been successfully sent and is automatically deleted.
- **Window Size**: Indicates 16-bit idle space in the receive buffer. It is used to notify the peer end of a TCP connection of the maximum data length supported by the local end.

- **Checksum**: Indicates a 16-bit TCP checksum.
- **Urgent Pointer**: Indicates a 16-bit field showing the offset of urgent data relative to the value of **Sequence Number**. It is meaningful only when the **URG** flag is set to **1**.
- **Option**: Indicates the option field used to improve the TCP transmission performance. The maximum length is 40 bytes because it is controlled by the data offset (header length). In addition, you are advised to set the **Option** field to an integer multiple of 32 bits.

2. Establishing a TCP Connection

TCP uses the "three-way handshake" mechanism to establish a virtual circuit connection.

Figure 1-3 Establishing a TCP Connection



(1) The client initiates a connection and ends the CLOSED phase. In this case, the server enters the LISTEN phase.

(2) The client sends a TCP SYN packet to the server, and then enters the SYN-SENT phase.

(3) The server receives the SYN packet, ends the LISTEN phase, and responds with an SYN ACK packet. Then, the server enters the SYN-RCVD phase.

(4) The client receives the SYN ACK packet from the server, confirming that the data transmission from the client to the server is normal. The client ends the SYN-SENT phase, and responds with an ACK packet. Then, the client enters the ESTABLISHED phase.

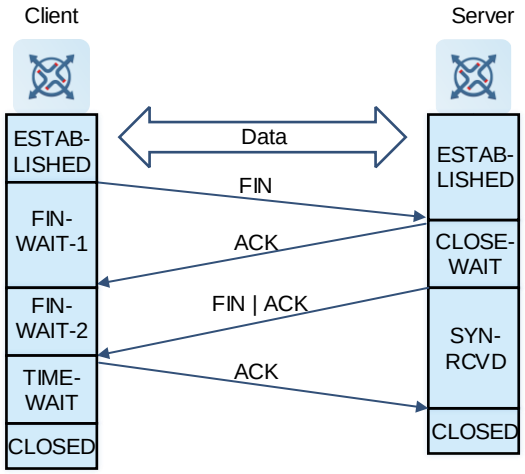
(5) The server receives the ACK packet from the client, confirming that the data transmission from the server to the client is normal. The server ends the SYN-SENT phase, and enters the ESTABLISHED phase.

(6) So far, the three-way handshake is completed, the TCP client and server establish a connection successfully and are ready for data transmission.

3. Releasing a TCP Connection

TCP uses the "four-way wavehand" mechanism to release a virtual circuit connection.

Figure 1-4 Releasing a TCP Connection



(1) To release a connection, the client sends a TCP FIN packet to the server. The client enters the FIN-WAIT-1 phase and stops sending data to the server. At this moment, the client can still receive data from the server.

(2) The server receives the FIN packet, ends the ESTABLISHED phase, enters the CLOSE-WAIT phase, and responds with an ACK packet.

(3) The client receives the ACK packet from the server, confirming that the server has received the connection release request from the client. Then, the client ends the FIN-WAIT-1 phase and enters the FIN-WAIT-2 phase.

(4) After sending the ACK packet, the server goes through the Closed-Wait phase, and sends a FIN ACK packet to the client. At this moment, the server has stopped sending data to the client, but can still receive data from the client.

(5) The client receives the TCP packet from the server, ends the FIN-WAIT-2 phase, enters the TIME-WAIT phase, and sends the ACK packet to the server.

(6) After waiting for a certain period of time, the client ends the TIME-WAIT phase and enters the CLOSED phase. The "four-way wavehand" is completed.

4. SYN Packet Retransmission Upon Timeout

If the server does not reply with an SYN ACK packet after the client sends an SYN packet, the client keeps retransmitting the SYN packet for certain times or until the timeout period expires.

If the server replies with an SYN ACK packet after the client sends an SYN packet but the client does not reply with an ACK packet, the server keeps retransmitting the SYN ACK packet for certain times or until the timeout period expires. This may be SYN flooding.

5. TCP Sliding Window

The TCP sliding window is the receive buffer used to buffer data from the peer. The data will be subsequently read by applications. The TCP window size reflects the size of idle space in the receive buffer. For bulk-data connections, enlarging the window size dramatically promotes TCP transmission performance.

6. MSS

The maximum segment size (MSS) refers to the maximum length of data payload in a TCP segment, excluding the TCP option.

Three-way handshake used for establishing a TCP connection needs MSS negotiation. Both parties of the connection add the MSS option to SYN packets. The option indicates the maximum size of a segment that can be received by the local end, that is, the maximum size of a segment that can be sent by the peer end. Both parties compare the local MSS value with the MSS value received from the peer end and take the smaller MSS value as the MSS of the connection.

The MSS option value in SYN packets to be sent is calculated as follows:

- IPv4 TCP: MSS = IP MTU of the outbound interface corresponding to the peer IPv4 address – IP header size (20 bytes) – TCP header size (20 bytes).
- IPv6 TCP: MSS = Path MTU corresponding to the peer IPv6 address – IPv6 header size (40 bytes) – TCP header size (20 bytes).

Note

The MSS that actually takes effect is the MSS calculated based on the maximum transmission unit (MTU) or the configured MSS, whichever is smaller.

If a connection supports certain options, the option length after 4-byte alignment should be deducted from the MSS value. For example, 20 bytes need to be deducted if the MD5 option is used because the length of the MD5 option is 18 bytes and the length after alignment is 20 bytes.

7. Path MTU Discovery

The TCP path MTU discovery function stipulated in RFC 1191 is used to discover the smallest MTU in a TCP path to prevent fragmentation and reassembly and enhance the network bandwidth utilization. The process of IPv4 TCP path MTU discovery is described as follows:

(1) The TCP source sets the Don't Fragment (DF) bit in the outer IP header in a TCP packet to be sent.

(2) If the outbound interface MTU value of a device in the TCP path is smaller than the IP packet length, the packet will be discarded and an ICMP error packet carrying this outbound interface MTU will be sent to the TCP source.

(3) By parsing the ICMP error packet, the TCP source knows the smallest MTU in the TCP path (that is, path MTU).

(4) The size of subsequent data segments sent by the TCP source will not surpass the MSS, which is calculated as follows: TCP MSS = Path MTU – IP header size – TCP header size.

If the MSS has reached the minimum 32 bytes specified by the system and an ICMP error packet asking the system to reduce the MSS is received again, the system will allow the fragmentation of the packets sent from this TCP connection.

The aging mechanism of the path MTU is as follows:

(1) After receiving an ICMP error packet, the TCP source not only reduces the path MTU value, but also starts an aging timer for the path MTU value.

(5) The system gradually increases the MSS value of TCP according to the MTU table specified in RFC 1191 when the timer expires.

(6) If no ICMP error packet is received within 2 minutes after one increment of the MSS, the system continues the increment until the MSS increases to the MSS value advertised by the peer during the three-way handshake.

8. TCP Keepalive Function

You may enable the keepalive function to check whether the peer end of a TCP connection works normally. If the TCP peer end does not send a packet to the local end within a period of time (namely, idle period), the local end starts sending keepalive packets to the peer end for multiple consecutive times. If no response packet is received, the local end considers the peer end abnormal and closes the TCP connection.

1.1.3 Protocols and Standards

- RFC 793: Transmission Control Protocol
- RFC 1122: Requirements for Internet Hosts -- Communication Layers
- RFC 1191: Path MTU Discovery
- RFC 1213: Management Information Base for Network Management of TCP/IP-based internets: MIB-II
- RFC 2385: Protection of BGP Sessions via the TCP MD5 Signature Option
- RFC 4022: Management Information Base for the Transmission Control Protocol (TCP)

1.2 Configuration Task Summary

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Optimizing TCP performance
  - Configuring TCP SYN Packet Timeout Period
  - Configuring a TCP Window Size
  - Configuring the Sending of TCP Reset Packets Upon the Receiving of Port Unreachable Messages
  - Configuring MSS for a TCP Connection
  - Configuring the TCP Path MTU Discovery Function
- Configuring TCP Connection Exception Detection

1.3 Configuring TCP SYN Packet Timeout Period

1.3.1 Overview

In case of SYN flooding, shortening SYN timeout period can reduce resource consumption. However, it does not work on continuous SYN flooding.

When a device actively requests to establish a connection with an external device, shortening SYN timeout period can reduces users' waiting time, for example, waiting time in the telnet connection.

You may prolong SYN timeout period properly on an unstable network.

1.3.2 Restrictions and Guidelines

Perform the configuration on both ends of a TCP connection.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure SYN timeout.

ip tcp syn-timeout *time*

The default timeout period of SYN packets used for connection establishment is 20 seconds.

1.4 Configuring a TCP Window Size

1.4.1 Overview

The TCP receive buffer is used to buffer data from the peer. The data will be subsequently read by applications. The TCP window size reflects the size of idle space in the receive buffer. For bulk-data connections, enlarging the window size dramatically promotes TCP transmission performance.

1.4.2 Restrictions and Guidelines

Perform the configuration on both ends of a TCP connection.

- The IPv4 TCP and IPv6 TCP configuration commands are the same.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configures a TCP window size.

ip tcp window-scale *size*

The default TCP window size is 65,535 bytes.

If the window size is greater than 65535 bytes, window enlarging will be enabled automatically. The window size advertised to the peer is the configured window size or the idle space in the receive buffer, whichever is smaller.

1.5 Configuring the Sending of TCP Reset Packets Upon the Receiving of Port Unreachable Messages

1.5.1 Overview

In general, when the device distributes a TCP packet, if the TCP connection, to which the packet belongs cannot be identified, the device sends a reset packet to the peer to terminate the TCP connection. This, however, can also become a target for attackers. A large number of TCP port unreachable messages can

impose attacks on the device. You can use this command to prevent the sending of TCP reset packets upon the receiving of port unreachable messages.

1.5.2 Restrictions and Guidelines

Perform the configuration on both ends of a TCP connection.

- The IPv4 TCP and IPv6 TCP configuration commands are the same.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the sending of TCP reset packets after port unreachable messages are received.

ip tcp send-reset

TCP reset packets are sent upon the receiving of port unreachable messages by default.

1.6 Configuring MSS for a TCP Connection

1.6.1 Overview

TCP negotiates about the Maximum Segment Size (MSS) value when establishing a connection. The MSS refers to the maximum length of data payload in a TCP segment. If the path MTU discovery function at one end of a TCP connection is not available, the device at this end cannot calculate the size of TCP packets based on the MTU value, and the TCP packets may be discarded due to too long size. To prevent this situation, you can manually set the MSS for TCP packets so that the MSS value negotiated by both ends will not exceed this value and the packets can successfully go through the intermediate networks.

1.6.2 Restrictions and Guidelines

- Perform the configuration on both ends of a TCP connection.
- The IPv4 TCP and IPv6 TCP configuration commands are the same.
- An excessively small MSS reduces transmission performance. You can promote TCP transmission by increasing the MSS. Configure the MSS by referring to the interface MTU. If the former is bigger than the interface MTU, TCP packets need to be fragmented and reassembled, which reduces the transmission performance.
- If an upper MSS limit is configured, the upper MSS limit that actually takes effect is the MSS calculated based on the MTU or configured MSS, whichever is smaller.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an upper MSS limit.

ip tcp max-segment-size *max-segment-size*

Default MSS Upper Limit = MTU – Header Length – TCP Header Length

Note

Upper IPv4 TCP MSS Limit = IPv4 Path MTU – 20-Byte IPv4 Header –20-Byte TCP Header

Upper IPv6 TCP MSS Limit = IPv6 Path MTU – 20-Byte IPv4 Header –20-Byte TCP Header

1.7 Configuring the TCP Path MTU Discovery Function

1.7.1 Overview

The path MTU discovery function of TCP is implemented according to RFC 1191 to improve the network bandwidth utilization. When TCP is applied to bulk transmit chunk data, this function can improve transmission performance greatly.

After discovering the path MTU, TCP can use a larger MSS to detect a new path MTU at intervals. When the device discovers a path MTU smaller than the MSS negotiated by both ends of a TCP connection, the device tries to discover a larger path MTU at intervals. The detection process is stopped when the path MTU reaches the MSS or the user turns off the timer.

1.7.2 Restrictions and Guidelines

Perform the configuration on both ends of a TCP connection.

- Run the **ip tcp track-mtu-discovery** command to enable the IPv4 Path MTU discovery function of TCP. Run the **unset** form of this command to disable the function.
  - Run the **age** *time* command to configure the interval for a new detection after TCP discovers the path MTU in minutes. The value range is from 10 to 30. The default value is 10.
  - Run the **age infinite** command to stop detection after TCP discovers the path MTU.
- The IPv6 Path MTU discovery function of TCP is enabled permanently and cannot be disabled.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IPv4 Path MTU discovery function.

ip tcp track-mtu-discovery [**age** *time* | **age infinite**]

The IPv4 Path MTU discovery function of TCP is disabled by default.

1.8 Configuring TCP Connection Exception Detection

1.8.1 Overview

You may enable the TCP keepalive function to check whether the peer works normally.

Suppose that the TCP keepalive function is enabled on a device and default interval, transmission count, and idle period settings are used. If no packet is received from the peer within 15 minutes, the device starts sending keepalive packets every 75 seconds for 6 consecutive times. If the device receives no TCP packet from the peer, it considers the TCP connection invalid and automatically releases the TCP connection.

The IPv4 TCP and IPv6 TCP configuration commands are the same.

1.8.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the TCP keepalive function.

ip tcp keep-alive [**interval** *interval*] [**times** *times*] [**idle-period** *time*]

The TCP keepalive function is disabled by default.

This command is no different to the server and client and applies to all TCP connections.

1.9 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 TCP Monitoring

Command	Purpose
display tcp connect [**local-ip** *ipv4-address*] [**local-port** *port-number*] [**peer-ip** *ipv4-address*] [**peer-port** *port-number*] [**vrf-name** *vrf-name*]	Displays the basic information about an IPv4 TCP connection.
display tcp connect statistics [**local-ip** *ipv4-address*] [**local-port** *port-number*] [**peer-ip** *ipv4-address*] [**peer-port** *port-number*] [**vrf-name** *vrf-name*]	Displays the statistics on an IPv4 TCP connection.

Command	Purpose
display tcp pmtu [**local-ip** *ipv4-address*] [**local-port** *port-number*] [**peer-ip** *ipv4-address*] [**peer-port** *port-number*]	Displays the path MTU information of IPv4 TCP.
display tcp port [*port-number*]	Displays the usage of IPv4 TCP port.
display tcp parameter	Displays IPv4 TCP parameters.
display tcp statistics	Displays IPv4 TCP statistics.
debugging ip tcp packet [**in** | **out**] [**local-ip** *ipv4-address*] [**peer-ip** *ipv4-address*] [**global** | **vrf** *vrf-name*] [**local-port** *port-number*] [**peer-port** *port-number*] [**deeply**]	Debugs IPv4 TCP packets.
debugging ip tcp transactions [**local-ip** *ipv4-address*] [**peer-ip** *ipv4-address*] [**local-port** *port-number*] [**peer-port** *port-number*]	Debugs IPv4 TCP connections.
display ipv6 tcp connect [**local-ipv6** *ipv6-address*] [**local-port** *port-number*] [**peer-ipv6** *ipv6-address*] [**peer-port** *port-number*] [**vrf-name** *vrf-name*]	Displays the basic information about an IPv6 TCP connection.
display ipv6 tcp connect statistics [**local-ip** *ipv4-address*] [**local-port** *port-number*] [**peer-ip** *ipv4-address*] [**peer-port** *port-number*] [**vrf-name** *vrf-name*]	Displays the statistics on an IPv6 TCP connection.
display ipv6 tcp pmtu [**local-ipv6** *ipv6-address*] [**local-port** *port-number*] [**peer-ipv6** *ipv6-address*] [**peer-port** *port-number*] [**vrf-name** *vrf-name*]	Displays the path MTU information of IPv6 TCP.
display ipv6 tcp port [*port-number*]	Displays the usage of IPv6 TCP port.
debugging ipv6 tcp packet [**in** | **out**] [**local-ipv6** *ipv6-address*] [**peer-ipv6** *ipv6-address*] [**global** | **vrf** *vrf-name*] [**local-port** *port-number*] [**peer-port** *port-number*] [**deeply**]	Debugs IPv6 TCP packets.
debugging ipv6 tcp transactions [**local-ipv6** *ipv6-address*] [**peer-ipv6** *ipv6-address*] [**local-port** *port-number*] [**peer-port** *port-number*]	Debugs IPv6 TCP connections.

Contents

1 Configuring IP REF 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 3
1.2 Configuration Task Summary 3
1.3 Notes 3
1.4 Configuring REF Load Balancing 4
1.4.1 Overview 4
1.4.2 Procedure 4
1.5 Configuring ECMP to Conduct Load Balancing According to a Fixed Profile 4
1.5.1 Overview 5
1.5.2 Restrictions and Guidelines 5
1.5.3 Procedure 5
1.6 Configuring ECMP to Conduct Load Balancing According to an Enhanced Profile 6
1.6.1 Overview 6
1.6.2 Restrictions and Guidelines 6
1.6.3 Procedure 6
1.7 Configuring Illegitimate Address Check 7
1.7.1 Overview 7
1.7.2 Restrictions and Guidelines 7
1.7.3 Procedure 7

1.8 Configuring a Warning Threshold for ECMP Group Resources or Next Hop Resources.....7
1.8.1 Overview.....9
1.8.2 Restrictions and Guidelines.....9
1.8.3 Procedure.....9
1.9 Configuring an L3 MAC Address.....9
1.9.1 Overview.....9
1.9.2 Restrictions and Guidelines.....9
1.9.3 Procedure.....10
1.10 Monitoring.....10
1.11 Configuration Examples.....12
1.11.1 Configuring Load Balancing.....12

1 Configuring IP REF

1.1 Introduction

1.1.1 Overview

On products that do not support hardware-based forwarding, IPv4/IPv6 packets are forwarded by software. To optimize the forwarding performance of software, the device introduces the IPv4/IPv6 technology to the software. REF can actively resolve next hops and implement load balancing.

REF maintains two tables: forwarding table and adjacency table. The forwarding table is used to store routing information. The adjacency table, equivalent to the Address Resolution Protocol (ARP) table and IPv6 neighbor table, is used to store link layer information of next hops.

1.1.2 Principles

1. Basic Concepts

- Forwarding table

  A forwarding table, also known as forwarding information base (FIB), is the lookup table where the device determines target devices for data exchange, and there are one-to-one mappings between entries in the FIB and entries in the IP routing table. Since the FIB contains all the necessary routing information, there is no need to maintain the route cache. When the network topology or route changes, the IP routing table is updated, and entries in the FIB change accordingly. During packet forwarding, IP REF selects packet transmission paths according to the forwarding table.

- Adjacency table

  An adjacency table contains output interface information of routed packets, for example, the next hop list, the next component used for packet processing, or the link layer output encapsulation. When a packet matches a certain adjacent node, the packet is directly encapsulated and then the send function of the node is called to forward the packet. For the sake of search and update, an adjacency table is often organized in the form of a hash table. To support routing load balancing, the next hop list of an adjacent node is organized in the form of a load balancing table. The next hop list of an adjacent node may not contain next hop information but may contain indexes of next components used for packet processing (such as other line cards and multi-service cards).

- Active resolution

  REF supports next hop resolution. If the MAC address of the next hop of an Ethernet interface is unknown, REF will actively resolve the next hop.

  IPv4 REF requests the ARP module to resolve the next hop. IPv6 REF requests the neighbor discovery (ND) module to resolve the next hops.

- Packet forwarding path

  Packets are routed and forwarded based on their IPv4/IPv6 addresses. If the source and destination IPv4/IPv6 addresses of a packet are specified, the forwarding path of this packet is definite.

2. REF Load Balancing

REF Load balancing is used to distribute traffic load among channels of multiple network devices. In the REF model, a route is a multi-path route when the route prefix is associated with multiple next hops. In this case, the route will be associated with a load balancing table and load balancing will be implemented based on the route weights. When an IP packet matches the load balancing table based on the longest prefix match, REF performs hash calculation based on the IP address of the packet and selects a path to forward the packet.

3. ECMP Load Balancing

Equal cost multiple path (ECMP) load balancing works in such a way that the same packet flow is transmitted by one path and different packet flows are transmitted by multiple equal-cost paths according to the load balancing algorithms, to achieve load balancing.

ECMP differentiates packet flows by using one or several load balancing algorithms based on packet characteristics, such as the source IP address, destination IP address, L4 source port number, and L4 destination port number. It also distributes the same packet flow to the same path for transmission, and different packet flows to equal-cost paths evenly.

For example, in source IP address-based load balancing mode, packets are distributed to different paths based on the source IP addresses of the packets. Packets with different source IP addresses are evenly distributed to equal-cost paths, and packets with the same source IP address are forwarded by the same path.

4. Hash Load Balancing Control

Hash load balancing allows users to control load balancing flexibly in different scenarios. Hash load balancing functions are described as below:

- Elastic hash

  In normal cases, a load balancing cluster interworks with multiple top of rack (ToR) devices via the ECMP, and the ToR devices are required to distribute all packets of a session to the same scheduling server. However, when a server is faulty, all session streams will be reorganized and balanced according to the traditional load balancing mechanism of ToR devices. In this case, packets of the same session may be distributed to two servers. The elastic hash function ensures that, when one link is faulty, traffic on other links is not affected and only the traffic of the faulty link is distributed to other active links, without affecting the traffic of other links. In this way, traffic of a session will only be distributed to the same server.

- Hash algorithm mode setting

  When traffic models are different, an appropriate hash algorithm mode can be set to adapt to different traffic to achieve the optimal load balancing effect.

- Hash disturbance

  Device traffic is balanced using the hash algorithm. For the same stream from two devices of the same type, the same path will be calculated by the load balancing algorithm. When the ECMP is deployed, the same stream from the two devices may be distributed to the same destination device. This is called hash polarization. The hash disturbance factor is used to affect the load balancing algorithm of the device. Different disturbance factors can be configured on different devices so that different paths are calculated for the same stream.

- Hash synchronization

To ensure network security, a firewall cluster is deployed between the internal and external networks for traffic cleaning. This requires that both the uplink and downlink traffic of a session be transmitted to the same device in the firewall cluster for processing. The source and destination IP addresses contained in the uplink and downlink streams of a session are reversed. If a traditional hash algorithm is used, the uplink and downlink streams will be directed to different firewalls, while hash synchronization can ensure that the same path is calculated for the uplink and downlink streams.

- Hash factor acquisition

  For packets of the virtual extensible LAN (VXLAN) and Generic Routing Encapsulation (GRE) tunnels, there are inner and outer headers. The hash factor can be obtained from the inner or outer header to achieve better load balancing. For example, when tunnel packets have the same IP address in the outer header but different IP addresses in the inner header, the IP address in the inner header can be specified as a hash factor in order to achieve better load balancing.

1.1.3 Protocols and Standards

RFC 1812: Requirements for IP Version 4 Routers

1.2 Configuration Task Summary

All the configuration tasks below are optional. Perform the configuration tasks as required.

- Configuring REF Load Balancing
- Configuring ECMP to Conduct Load Balancing According to a Fixed Profile
- Configuring ECMP to Conduct Load Balancing According to an Enhanced Profile
- Configuring Illegitimate Address Check

1.3 Notes

This section describes the precautions for configuring the IP REF function.

Whether the IP REF function is supported depends on devices.

Table 1-1 Products Supporting the IP REF Function

Function	Product Supporting the Function
Configuring ECMP to Conduct Load Balancing According to a Fixed Profile Function	S6980-64QC S6580-48CQ8QC S6580-24DC8QC
Configuring ECMP to Conduct Load Balancing According to an Enhanced Profile Function	S6980-64QC S6580-48CQ8QC S6580-24DC8QC

1.4 Configuring REF Load Balancing

1.4.1 Overview

REF Load balancing is used to distribute traffic load among channels of multiple network devices. REF supports load balancing of packets. In the REF model, a route is a multi-path route when the route prefix is associated with multiple next hops. In this case, the route will be associated with a load balancing table and load balancing will be implemented based on the route weights. When an IP packet matches the load balancing table based on the longest prefix match, REF performs hash calculation based on the IP address of the packet and selects a path to forward the packet.

IP REF supports three load balancing modes, that is, three routing policies:

- The load is balanced based on the destination IP addresses of IP packets and hash calculation is performed on the destination IP addresses of the packets. A path with a greater weight is more likely to be selected. This policy is used by default.
- The load is balanced based on the source and destination IP addresses of IP packets and hash calculation is performed on the source and destination IP addresses of the packets. A path with a greater weight is more likely to be selected.
- The load is balanced based on the source IP addresses of IP packets and hash calculation is performed on the source IP addresses of the packets. A path with a greater weight is more likely to be selected.

1.4.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an IP REF load balancing algorithm.

(IPv4 network)

ip ref load-sharing { **original | original-only** }

The default load balancing algorithm of IPv4 REF is load balancing based on the destination IP addresses.

(IPv6 network)

ipv6 ref load-sharing { **original | original-only** }

The default load balancing algorithm of IPv6 REF is load balancing based on the destination IPv6 addresses.

1.5 Configuring ECMP to Conduct Load Balancing According to a Fixed Profile

Note

Only S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC support this function.

1.5.1 Overview

ECMP differentiates packet flows by using one or several load balancing algorithms based on packet characteristics, such as the source IP address, destination IP address, L4 source port number, and L4 destination port number. It also distributes the same packet flow to the same path for transmission, and different packet flows to equal-cost paths evenly.

1.5.2 Restrictions and Guidelines

- The command for configuring a load balancing policy in basic load balancing mode is not different to IPv4 and IPv6.
- When the **enhanced profile** *profile-name* parameter is configured in the **ip ref load-balance** command, load balancing can be performed according to the enhanced profile indicated by *profile-name*. For the enhanced profile configurations, see Configuring ECMP to Conduct Load Balancing According to an Enhanced Profile.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ECMP load balancing policy.

ip ref load-balance enhanced profile *profile-name*

The load is balanced according to an enhanced profile by default.

(4) (Optional) Enable the elastic hash of ECMP load balancing.

ip ref hash-elasticity enable

or

ecmp cluster enable

The **ecmp cluster enable** command is only used for compatibility with earlier versions. The **ecmp cluster enable** and **ip ref hash-elasticity enable** commands implement the same functions, and the configurations of the command run later are displayed. You are advised to use the **ip ref hash-elasticity enable** command.

Elastic hash of ECMP load balancing is disabled by default.

(5) (Optional) Configure the balancing factor acquisition mode for GRE and IP-IN-IP packets in an enhanced load balancing profile.

ip ref hash-header { inner | outer }

The outer header of GRE and IP-IN-IP packets are used as the source for acquiring the balancing factor by default.

(6) (Optional) Configure the ECMP hash algorithm mode.

ip ref algorithm mode *hash-number*

The default ECMP hash algorithm mode is 9.

1.6 Configuring ECMP to Conduct Load Balancing According to an Enhanced Profile

Note

This function is supported by only the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1.6.1 Overview

The enhanced load balancing mode allows users to define load balancing profiles and combine multiple fields of different types of packets for load balancing.

In enhanced load balancing mode, a device first determines the type of packets to be transmitted and performs load balancing based on the specified packet fields. For example, if IPv4 packets containing changing source IP addresses need to be output through an aggregate port, configure the device to perform load balancing based on the specified source IP address field in IPv4 packets.

Configure this function on aggregation or core switches that support enhanced load balancing, and specify parameters to achieve load balancing.

1.6.2 Restrictions and Guidelines

- **dst-ip**: Performs load balancing based on the destination IP addresses of incoming packets.
- **l4-dst-port**: Performs load balancing based on the L4 destination port numbers of incoming packets.
- **l4-src-port**: Performs load balancing based on the L4 source port numbers of incoming packets.
- **protocol**: Performs load balancing based on the protocol types of incoming packets.
- **src-ip**: Performs load balancing based on the source IP addresses of incoming packets.
- **src-port**: Performs load balancing based on the source port numbers of incoming packets.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Rename the enhanced profile and enter the enhanced profile configuration mode of REF load balancing.

ip-ref-load-balance-profile [*profile-name*]

The default name of the enhanced profile is **default**.

(4) Configure the enhanced profile for load balancing. Perform at least one of the following configuration steps:

- Configure the load balancing mode for IP packets.

  (IPv4 network)

  ipv4 field [**dst-ip** | **l4-dst-port** | **l4-src-port** | **protocol** | **src-ip** | **src-port**]

  The load is balanced based on the source IP addresses, destination IP addresses, L4 source port numbers, and L4 destination port numbers of packets by default.

  (IPv6 network)

ipv6 field [**dst-ip** | **l4-dst-port** | **l4-src-port** | **protocol** | **src-ip** | **src-port**]

The load is balanced based on the source IP addresses, destination IP addresses, L4 source port numbers, and L4 destination port numbers of packets by default.

- Configure the hash disturbance factor.
  - **hash-disturb** { string | { [**seed** *seed_value*] [**offset** *offset*] } }

  The hash disturbance function of load balancing is disabled by default.
- Enable hash synchronization.
  - **hash-symmetrical** { **ipv4 | ipv6** }

  Hash synchronization is disabled by default.

1.7 Configuring Illegitimate Address Check

1.7.1 Overview

REF is capable of checking the destination IP addresses of IPv4 packets in a network, filtering out packets with illegitimate addresses, and recording these illegitimate addresses.

After the illegitimate address check function of REF is enabled, the device will check the destination IP addresses of IPv4 packets. The following types of addresses are considered illegitimate:

- Host addresses 0 (including 0.0.0.0).
- Loopback addresses (127.0.0.0–127.255.255.255).
- Class E addresses (240.0.0.0–255.255.255.254).

Invalid packets containing destination IP addresses that meet the above rules will be discarded by the device, and the illegitimate addresses will be logged.

1.7.2 Restrictions and Guidelines

REF can filter out class E addresses and loopback addresses and find out illegitimate addresses with the host address of 0 based on the specified mask length.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure illegitimate address check.

ip ref check [**masklen** *mask-length*]

IPv4 REF does not check illegitimate destination IP addresses by default.

1.8 Configuring Elastic Hashing

1.8.1 Overview

Elastic hashing is a scheme for load balance optimization upon Equal Cost Multipath (ECMP) next hop change.

A load balancing cluster is interconnected with the ToR (Top-of-Rack) switch through ECMP to balance traffic among each member node of the cluster.

An ECMP route is generated between the ToR switch and the cluster through a dynamic routing protocol. The dynamic routing protocol converges after an ECMP path fails. The traffic between the ToR switch and the cluster will be balanced again, changing the original session state on the member nodes of the cluster. The cluster needs to set up new sessions, causing some sessions to be disconnected.

Elastic hashing eliminates traffic re-balancing caused by the change in the number of ECMP paths. After elastic hashing is enabled, the traffic carried over the failed path is balanced among the active paths while the traffic on the active links remains unchanged.

If the next hop is session-sensitive, that is, the next hop or exit interface is an endpoint service device such as a server, you are advised to use elastic hashing. If it is a transient network node, elastic hashing does not stand out.

1.8.2 Restrictions and Guidelines

- Enabling elastic hashing can mitigate the effect on the original forwarding traffic when the number of ECMP next hops is increased or reduced.
- Enabling cross-device elastic hashing can mitigate the effect on the original forwarding traffic when the ECMP next hop is Up or Down.
- After cross-device elastic hashing is enabled, the next hop is sorted by the IP address again when a new next hop is added.
- After cross-device elastic hashing is enabled, the hashing result for multiple standalone devices with the same ECMP next hop is consistent.
- Elastic hashing and cross-device hashing are mutually exclusive and cannot be both configured.

To ensure hash consistency when switching traffic to another device upon device restart or switchover, make sure the following conditions are met:

- Both devices are equipped with the chip of the same model.
- Both devices are configured with same hash algorithm, factor, disturbance, and symmetrical hash. Do not configure matching domains for load balancing based on the source interface. Otherwise, hash inconsistency may occur.
- Both devices have the same number of next hops. The next hop is sorted by the same IP address order.
- If both devices have a next-hop fault before restart or switchover, the fault must be rectified to ensure hash consistency. If not, the restarted device is sorted by the new IP address while the device that is not restarted is sorted by the old IP address, leading to hash inconsistency.

1.8.3 Procedure

(1) Enter the privileged EXCE mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Configure elastic hashing for ECMP load balancing.
[**unset**] **ecmp cluster enable**

Elastic hashing is disabled by default.

(4) Configure cross-device elastic hashing for ECMP load balancing.
[**unset**] **ecmp cluster across-dev enable**

Cross-device elastic hashing is disabled by default.

1.9 Configuring a Warning Threshold for ECMP Group Resources or Next Hop Resources

1.9.1 Overview

After a warning threshold is configured for ECMP group resources or next hop resources, a system log is generated if the quantity of occupied resources reaches the warning threshold.

1.9.2 Restrictions and Guidelines

- The threshold range is related to the resource capacity supported by devices.
- After the quantity of actually occupied resources reaches the threshold, a system log is generated.
- If the configured threshold is smaller than or equal to the quantity of actually occupied resources, a system log is generated immediately.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a warning threshold for ECMP group resources or next hop resources.

ecmp group warningvalue *value*

nexthop warningvalue *value*

1.10 Configuring an L3 MAC Address

Note

This command is not supported by the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1.10.1 Overview

You can define an L3 MAC address for a device so that packets whose destination MAC addresses match the L3 MAC address are forwarded at L3. In data center scenarios, an L3 MAC address may be specified in static ARP entries of neighbors so that packets whose destination addresses match the L3 MAC address can be routed by any neighbor.

1.10.2 Restrictions and Guidelines

- A global L3 MAC address can be added to trigger the routing of packets whose destination addresses match the L3 MAC address.
- A maximum of two L3 MAC addresses can be added.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a global L3 MAC address.

l3intf-mac-add *mac*

1.11 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** command to clear information.

⚠ Caution

Running the **cleanup** command may lose vital information and thus interrupt services.

Table 1-2 Monitoring

Command	Purpose
display ip logging [**cleanup**]	Displays the recorded illegitimate addresses.
display ip ref packet statistics	Displays current packet statistics of IPv4 REF.
cleanup ip ref packet statistics	Clears current packet statistics of IPv4 REF.
display ip ref adjacency [**glean** | **local** | *ipv4-address* | **interface** *interface-type interface-number* | **discard** | **statistics**]	Displays information about the specified glean adjacencies, local adjacencies, adjacency of a specified IP address, adjacencies associated with a specified interface, and all adjacent nodes of IPv4 REF.
display ip ref resolve-list	Displays the next hop resolved by IPv4 REF.
display ip ref precise [**oob** | **vrf** *vrf-name*] *source-ipv4-address destination-ipv4-address*	Displays the actual forwarding path of a specified packet.
display ip ref route [**oob** | **vrf** *vrf-name*] [**default** | *ipv4-address mask* | **statistics**]	Displays routing information in the IPv4 REF table. The parameter **default** indicates that information about the default route will be displayed.
display ip ref precise [**vrf** *vrf-name*] *source-ipv4-address destination-ipv4-address*	Displays the actual forwarding path of a specified packet.
display ip ref route [**vrf** *vrf-name*] [**default** | *ipv4-address mask* | **statistics**]	Displays routing information in the IPv4 REF table. The parameter **default** indicates that information about the default route will be displayed.

Command	Purpose
display ip ref route detail [*ipv4-address*] [**vrf** *vrf-name*] [**slot** *slot-id*]	Displays details about the current IPv4 route. Details about a single IPv4 route, IPv4 routes of a specified line card, or IPv4 routes of all cards can be displayed.
display route-res usage [**all** | **slot** *slot_id*]	Displays the usage of routing resources in a specified slot or all slots.
display ip ref load-balance	Displays the configuration of the ECMP load balancing using a fixed profile.
display ip ref-load-balance-profile [*profile-name*]	Displays the configuration of the ECMP load balancing using an enhanced profile.
ip ref synchronize all	Synchronizes IPv4 routes between the hardware and software REF.
display ipv6 ref packet statistics	Displays current packet statistics of IPv6 REF.
cleanup ipv6 ref packet statistics	Clears current packet statistics of IPv6 REF.
display ipv6 ref adjacency [**glean** | **local** | *ipv6-address* | **interface** *interface-type interface-number* | **discard** | **statistics**]	Displays information about the specified glean adjacencies, local adjacencies, adjacency of a specified IPv6 address, adjacencies associated with a specified interface, and all adjacent nodes of IPv6 REF.
display ipv6 ref resolve-list	Displays the next hop resolved by IPv6 REF.
display ipv6 ref precise [**oob** | **vrf** *vrf-name*] *source-ipv6-address destination-ipv6-address*	Displays the actual forwarding path of a specified IPv6 packet.
display ipv6 ref route [**oob** | **vrf** *vrf-name*] [**default** | **statistics** | *ipv6-address/prefix-length*]	Displays route information in the IPv6 REF table. The parameter **default** indicates that information about the default route will be displayed.
display ipv6 ref route detail [*ipv6-address*] [**vrf** *vrf-name*] [**slot** *slot-id*]	Displays details about the current IPv6 route. Details about a single IPv6 route, IPv6 routes of a specified line card, or IPv6 routes of all cards can be displayed.
ipv6 ref synchronize all	Synchronizes IPv6 routes between the hardware and software REF.
display ecmp group count	Displays the usage of ECMP group resources.
display nexthop count	Displays the usage of next hop resources.
display l3intf-mac-add	Displays the configured L3 MAC address.

1.12 Configuration Examples

1.12.1 Configuring Load Balancing Policies

Note

Only S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC support the configuration example of the ECMP load balancing policy.

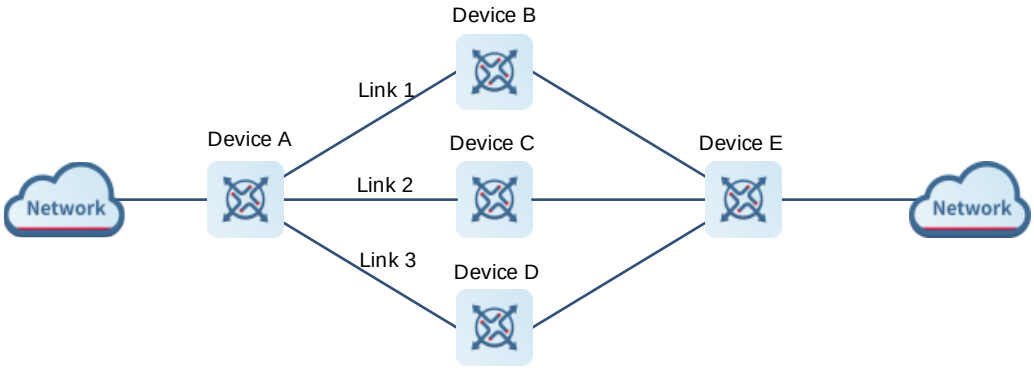
1. Requirements

As shown in Figure 1-1, device A and device E run REF, and device B, device C and device D are forwarding devices.

Device A communicates with device E. Link 1, link 2, and link 3 are three equal-cost links. A specific load balancing algorithm is configured on device A so that traffic is balanced by these three equal-cost links.

2. Topology

Figure 1-1 Configuring Load Balancing Policies



3. Notes

- Configure a policy of load balancing using a fixed profile on device A. Configure the load balancing mode, enable the elastic hash of ECMP load balancing, and configure the ECMP hash algorithm mode.
- Configure a policy of load balancing using an enhanced profile on device A: Rename the ECMP load balancing profile, configure the load balancing mode for IPv4 packets, configure the load balancing mode for IPv6 packets, configure the hash disturbance factor of load balancing, and disable the load balancing hash disturbance function.

4. Procedure

Configure a policy of load balancing using a fixed profile on device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip ref load-balance enhanced profile test
DeviceA(cfg)# ip ref hash-elasticity enable
DeviceA(cfg)# ip ref algorithm mode 8
```

Configure a policy of load balancing using an enhanced profile on device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip-ref-load-balance-profile rpn
DeviceA(ref-ip-config-load-balance-profile)# ipv4 field src-ip dst-ip
DeviceA(ref-ip-config-load-balance-profile)# ipv6 field src-ip dst-ip protocol
DeviceA(ref-ip-config-load-balance-profile)# hash-disturb A
DeviceA(ref-ip-config-load-balance-profile)# unset hash-symmetrical ipv6
```

5. Verification

Run the **display ip ref load-balance** command to display the configuration of load balancing using a fixed profile.

```
DeviceA# display ip ref load-balance
load-balance                  : src-ip.
  hash-elasticity             : enable.
Algorithm mode
  current: 8,   default: 9
```

Run the **display ip ref load-balance-profile** command to display the configuration of load balancing using an enhanced profile.

```
DeviceA# display ip ref load-balance-profile
Load-balance-profile: rpn
Packet      Hash Field:
      IPV4: src-ip dst-ip
      IPV6: protocol src-ip dst-ip
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
ip-ref-load-balance-profile rpn
 ipv4 field src-ip dst-ip
 ipv6 field protocol src-ip dst-ip
 hash-disturb A
 unset hash-symmetrical ipv6
!
ip ref algorithm mode 8
ip ref load-balance enhanced profile test
ip ref hash-elasticity enable
```

IP Routing Configuration

1. Configuring IP Routing Basics
2. Configuring Static Routing
3. Configuring RIP
4. Configuring RIPng
5. Configuring OSPFv2
6. Configuring OSPFv3
7. Configuring IS-IS
8. Configuring BGP
9. Configuring VRF
10. Configuring Routing Policies
11. Configuring PBR
12. Configuring Keys

Contents

1 Configuring IP Routing Basics 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 3
1.3 Configuring Route Limitations 4
1.3.1 Overview 4
1.3.2 Restrictions and Guidelines 4
1.3.3 Configuration Tasks 4
1.3.4 Configuring the Maximum Number of Equal-Cost Routes 4
1.3.5 Disabling IPv4 Routing 5
1.3.6 Disabling IPv6 Routing 5
1.4 Configuring Fast Convergence of Recursive Routing 6
1.4.1 Overview 6
1.4.2 Configuration Tasks 6
1.4.3 Enabling IPv4 Recursive Routing Fast Convergence 6
1.4.4 Enabling IPv6 Recursive Routing Fast Convergence 6
1.4.5 Preventing a Route from Being Iterated to a Black-Hole Route 7
1.1.2 Preventing a Route from Being Iterated to a Default Route 8
1.5 Configuring ARP-to-host Conversion 9
1.5.1 Overview 9
1.5.2 Restrictions and Guidelines 9

1.5.3 Configuration Tasks 9
1.5.4 Configuring the Interface to Enable ARP-to-host Conversion 9
1.5.5 Enabling Proxy ARP 9
1.5.6 Configuring BGP Redistributed ARP-host Route 10
1.6 Configuring ND-to-route Conversion 10
1.6.1 Overview 10
1.6.2 Restrictions and Guidelines 10
1.6.3 Configuration Tasks 11
1.6.4 Configuring the Interface to Enable ND Conversion to Route 11
1.6.5 Configuring Delayed Redistribution of ND-to-route 12
1.6.6 Configuring Tag Attribute of ND-to-route Conversion 13
1.6.7 Configuring to Ignore Warning Log of ND Conversion to Route 13
1.6.8 Configuring ND Proxy 14
1.6.9 Configuring BGP Redistributed ND-Route Route 14
1.7 Configuring BGP Route Deletion During GR 15
1.7.1 Overview 15
1.7.2 Procedure 15
1.8 Enabling Real-Time Active Route Statistics Collection 15
1.8.1 Overview 15
1.8.2 Restrictions and Guidelines 15
1.8.3 Procedure 15
1.9 Monitoring 16

1 Configuring IP Routing Basics

1.1 Introduction

1.1.1 Overview

A router receives a data packet, checks its destination address, and forwards the packet to the next router according to the routing table. The packet arrives at the destination host after being forwarded multiple times by intermediate routers. The forwarding path of the intermediate routers constitutes the routing path information.

Note

In this chapter, the term "router" refers to any network device that supports the routing function. These network devices can be Layer-3 switches, routers, or firewalls.

1.1.2 Principles

1. Composition of Routing Table

A router manages the routing table, consolidates routes sent by various routing protocols, and selects and sends preferred routes to the routing table. Routes discovered by various routing protocols are stored in the routing table. The routing table contains the following key items:

Destination: Destination address, used to identify the destination network or destination address of the IP packet.

Mask: Network mask, which identifies the network segment of the destination address together with the destination address.

Protocol: Routing protocol.

Distance: Management distance, identifying the priority of the route for joining the routing table. For the same destination address, routing management will take the route with the smallest management distance as the optimal route. The table below lists the default management distances of different routing protocols.

Table 1-1 Routing Protocols and Default Management Distances

Routing Protocol	Management Distance
Direct route	0
Static route	1
EBGP	20
OSPF	110
IS-IS	115
RIP	120

Routing Protocol	Management Distance
IBGP	200

Metric: Routing cost. For the same management distance, the route with smallest routing cost will be the optimal route.

NextHop: The next-hop IP address, the interface IP address of the packet in the next forwarding route.

Interface: Forwarding interface, indicating from which interface of the local router the packet is forwarded to the next hop.

2. Route Classification

By the source mode, routes are generally divided into the following three categories:

- Direct route: It is the route discovered by a link-layer protocol and is also called interface route.
- Static route: It is manually configured by the network administrator. A static route is easy to configure and less demanding on the system, and therefore applicable to a small-sized network that is stable and has a simple topology. However, when the network topology changes, the static route must be manually reconfigured and cannot automatically adapt to the topological changes.
- Dynamic route: It is the route discovered by a dynamic routing protocol.

3. Route Computation

- Routing function

  Routing functions are classified into IPv4 and IPv6 routing functions. If the routing functions are disabled, a device is equivalent to a host and cannot forward routes.

- Dynamic Route

  A dynamic routing protocol learns remote routes and dynamically updates routes by exchanging routes with neighbors. If a neighbor is the next hop of a route and this neighbor fails, the route fails as well.

  Classified by the scope of action:

  - Interior Gateway Protocol (IGP): Running in an autonomous system. Protocols are Routing Information Protocol (RIP), Open Shortest Path First (OSPF), and Intermediate System to Intermediate System (IS-IS).
  - Exterior Gateway Protocol (EGP): Running between different autonomous systems. Protocol is Border Gateway Protocol (BGP).

  Classified by the route computation algorithm:

  - Distance vector routing protocols: Protocols are BGP and RIP.
  - Link state routing protocols: Protocols are OSPF and IS-IS.

- Static route

  On a network with a simple topology, you can configure only static routes to implement internetworking. Appropriate configuration and use of static routes can improve the network performance and guarantee the bandwidth for important network applications.

Whether a static route is active is computed based on the status of the local interface. When the exit interface of a static route is located at Layer-3 and is in Up status (the link status is Up and the IP address is configured), this route is active and can be used for packet forwarding.

A static route can go across VPN routing & forwarding (VRF) instances. The next hop or exit interface of a static route of VRF 1 can be configured on VRF 2.

- ARP-to-host conversion

  Enable ARP-to-host conversion on a specified interface. If the interface is a Layer-3 interface in Up status, ARP entries will be converted into 32-bit host routes on the interface.

4. Optimal Route Selection

- Management distance

  When multiple routing protocols generate routes to the same destination, the priorities of these routes can be determined based on the management distance. A smaller management distance indicates a higher priority.

- Equal-cost route

  If multiple routes to the same destination have different next hops but the same management distance, these routes are mutually equal-cost routes. Packets are distributed among these routes to implement load balancing based on the balanced forwarding policy.

  On a specific device, the total number of equal-cost routes is limited. Routes beyond the limit do not participate in packet forwarding.

  The routing protocols supporting equal-cost route are the static route, RIP, OSPF, IS-IS, and BGP.

- Floating route

  If multiple routes to the same destination have different next hops and different management distances, these routes are mutually floating routes. The route with the smallest management distance will be first selected for packet forwarding. If this route fails, a route with a larger management distance is further selected for forwarding, thus preventing communication interruption caused by a network line failure.

- Routing recursion

  Only the next hop directly connected can be used as the basis of router forwarding, but the next hop may not be directly connected in BGP and static route. In this case, route iteration is needed to query the routing table according to the next-hop address, find a route with the next hop directly connected or outbound interface, and use the route as the basis for forwarding.

1.2 Configuration Task Summary

The configuration of IP routing basic includes the following tasks:

(1) (Optional) Configuring Route Limitations The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring the Maximum Number of Equal-Cost Routes
- Disabling IPv4 Routing
- Disabling IPv6 Routing

(2) (Optional) Configuring Fast Convergence Fast Convergence The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Enabling IPv4 Recursive Routing Fast Convergence
- Enabling IPv6 Recursive Routing Fast Convergence
- Preventing a Route from Being Iterated to a Black-Hole Route
- Preventing a Route from Being Iterated to a Default Route

(3) (Optional) Configuring ARP-to-host Conversion

a Configuring the Interface to Enable ARP-to-host Conversion

b (Optional) Enabling Proxy ARP

c (Optional) Configuring BGP Redistributed ARP-host Route

(4) (Optional) Configuring ND-to-route Conversion

a Configuring the Interface to Enable ND Conversion to Route

b (Optional) Configuring Delayed Redistribution of ND-to-route

c (Optional) Configuring Tag Attribute of ND-to-route Conversion

d (Optional) Configuring to Ignore Warning Log of ND Conversion to Route

e (Optional) Configuring ND Proxy

f (Optional) Configuring BGP Redistributed ND-Route Route

(5) (Optional) Configuring BGP Route Deletion During GR

(6) (Optional) Enabling Real-Time Active Route Statistics Collection

1.3 Configuring Route Limitations

1.3.1 Overview

This function limits the number of equal-cost routes and number of static routes, or disables routing.

1.3.2 Restrictions and Guidelines

This function cannot be configured on a Layer-2 switch.

1.3.3 Configuration Tasks

The route limitation configuration tasks are as follows and optional. Select tasks for configuration according to actual condition.

- Configuring the Maximum Number of Equal-Cost Routes
- Disabling IPv4 Routing
- Disabling IPv6 Routing

1.3.4 Configuring the Maximum Number of Equal-Cost Routes

1. Overview

Run this command to configure the maximum number of next hops in the equal-cost route.

2. Restrictions and Guidelines

In load balancing mode, the number of routes on which traffic is balanced does not exceed the configured number of equal-cost routes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of equal-cost routes.

max-ecmp-paths *number*

The default value varies with the specific device model by default.

1.3.5 Disabling IPv4 Routing

1. Overview

When a device functions only as a bridge or VoIP gateway, the IPv4 routing and forwarding function is not required. In this case, the IPv4 routing function of the device can be disabled.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable IPv4 routing.

unset ip routing

The IPv4 routing function is enabled by default.

1.3.6 Disabling IPv6 Routing

1. Overview

When a device functions only as a bridge or VoIP gateway, the IPv6 routing and forwarding function is not required. In this case, the IPv6 routing function of the device can be disabled.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable IPv6 routing.

unset ipv6 unicast-routing

The IPv6 routing function is enabled by default.

1.4 Configuring Fast Convergence of Recursive Routing

1.4.1 Overview

When link discovery changes or is interrupted, the recursive routing query of dynamic route may converge 2 seconds later. The convergence time is unacceptable in some scenarios with high reliability requirements. When the fast convergence function of recursive routing is enabled, the intermediate next hop information will be added to the route sending information. Even if the intermediate link fails, it can be sensed by routing management to shorten the convergence time.

1.4.2 Configuration Tasks

The recursive routing configuration tasks are as follows and optional. Select tasks for configuration according to actual condition.

- Enabling IPv4 Recursive Routing Fast Convergence
- Enabling IPv6 Recursive Routing Fast Convergence
- Preventing a Route from Being Iterated to a Black-Hole Route
- Preventing a Route from Being Iterated to a Default Route

1.4.3 Enabling IPv4 Recursive Routing Fast Convergence

1. Overview

Enabling IPv4 recursive routing fast convergence can shorten the convergence time of IPv4 network intermediate link failure to a value less than 1 second.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IPv4 recursive routing fast convergence.

ip recur-route fastswitch-nexthop

The fast convergence function of recursive routing is disabled by default.

1.4.4 Enabling IPv6 Recursive Routing Fast Convergence

1. Overview

Enabling IPv6 recursive routing fast convergence can shorten the convergence time of IPv6 network intermediate link failure to a value less than 1 second.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IPv6 recursive routing fast convergence.

ipv6 recur-route fastswitch-nexthop

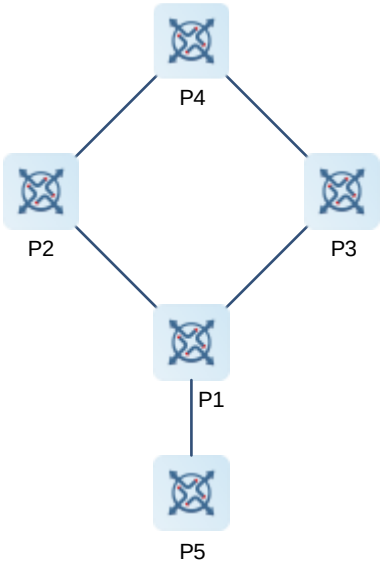
1.4.5 Preventing a Route from Being Iterated to a Black-Hole Route

1. Overview

This function allows a device to divert all traffic to the standby path when the active path fails, so as to avoid the loss of traffic caused by the import of a black-hole route.

As shown in Figure 1-1, P1 establishes BGP neighbor relationships with P2 and P3 respectively. Two paths from P1 to P2 and P3 form the active and standby paths to reach P4, where P1-P2 is the active path and P1-P3 is the standby path. In addition, P1 is configured with a black-hole route. When the link between P1-P2 fails, the expected traffic is switched to P3. However, due to the existence of a black-hole route, the router recalculates the route which is iterated to the black hole route, resulting in the loss of some traffic and failure to meet the convergence requirements. With this command, when the P1-P2 path fails, all traffic will be diverted to P3. In this case, the black-hole route is not imported, and convergence requirements can be met.

Figure 1-1 Prevent a Route from Being Iterated to a Black-Hole Route



2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure preventing a route from being iterated to a black-hole route.

ip recur-route over blackhole-route disable

By default, a route is allowed to be iterated to a black-hole route.

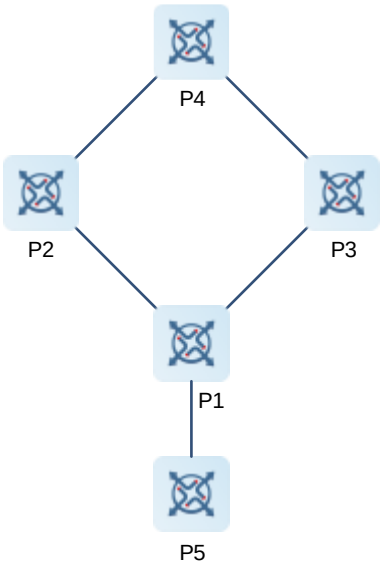
1.4.6 Preventing a Route from Being Iterated to a Default Route

3. Overview

When you do not want to select the default route for transmitting traffic, you can use this function to configure the device to not select the default route egress during recursive routing.

As shown in Figure 1-12, P1 establishes indirect EBGP neighbor relationships with P2 and P3 respectively. The traffic on P1 reaches P4 through P2 and P3 in a load balancing manner. In addition, the default route configured on P1 indicates that P1 accesses the Internet through the P1-P5 egress. When the link between P1 and P2 fails, it is expected that the traffic will be diverted to P3. However, due to the default route, some traffic will be diverted to the P1-P5 egress, which is not a service egress. This will lead to the loss of some traffic and failure to meet the requirement of 1-second convergence.

Figure 1-2 Prevent a Route from Being Iterated to a Default Route



4. Restrictions and Guidelines

- This command must be configured when the default route egress is not the expected service egress.

5. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Configure preventing a route from being iterated to a default route.
ip recur-route over blackhole-route disable

By default, a route is allowed to be iterated to a default route.

1.5 Configuring ARP-to-host Conversion

1.5.1 Overview

If other routing protocols are needed to distribute the direct route or instruct forwarding routes, the information in ARP table can be converted into host routes, and then redistributed by other routing protocols.

1.5.2 Restrictions and Guidelines

- This function cannot be configured on a Layer-2 switch.
- The interface must be a Layer-3 interface in Up status.

1.5.3 Configuration Tasks

ARP-to-host conversion configuration includes the following tasks:

(1) Configuring the Interface to Enable ARP-to-host Conversion

(2) (Optional) Enabling Proxy ARP

(3) (Optional) Configuring BGP Redistributed ARP-host Route

1.5.4 Configuring the Interface to Enable ARP-to-host Conversion

1. Overview

This function can be used to convert ARP entries into 32-bit host routes on a specified interface.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable ARP-to-host conversion on a specified interface.

ip route-config arp-to-host interface *interface-type interface-number*

The ARP-to-host conversion function is disabled on an interface by default.

1.5.5 Enabling Proxy ARP

1. Overview

After receiving an ARP request, the device uses a proxy to send an ARP response with its own Ethernet MAC address. In this case, hosts communicate with each other through Layer-3 routes.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  interface *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*
- Enter the SVI configuration mode.

  interface *vlan interface-number*
- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable the proxy ARP.

local-proxy-arp

The local proxy ARP function is disabled by default on the SVI interface.

1.5.6 Configuring BGP Redistributed ARP-host Route

1. Overview

BGP redistributes the host routes converted from the ARP table.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure BGP redistributed ARP-host route.

Redistribute-route arp-host

ARP-host routes are not redistributed by default.

1.6 Configuring ND-to-route Conversion

1.6.1 Overview

Similar to ARP-to-host conversion, the device can generate direct route information using neighbor discovery (ND) table entries. Other routing protocols can distribute routes according to the generated direct route information, and can specify the mask length of ND-to-route conversion.

1.6.2 Restrictions and Guidelines

- This function cannot be configured on a Layer-2 switch.

- The interface must be a Layer-3 interface in Up status.

1.6.3 Configuration Tasks

ND-to-route conversion configuration includes the following tasks:

(1) Configuring the Interface to Enable ND Conversion to Route

(2) (Optional) Configuring Delayed Redistribution of ND-to-route

(3) (Optional) Configuring Tag Attribute of ND-to-route Conversion

(4) (Optional) Configuring to Ignore Warning Log of ND Conversion to Route

(5) (Optional) Configuring ND Proxy

(6) (Optional) Configuring BGP Redistributed ND-Route Route

1.6.4 Configuring the Interface to Enable ND Conversion to Route

1. Overview

After receiving an ARP request, the device uses a proxy to send an ARP response with its own Ethernet MAC address. In this case, hosts communicate with each other through Layer-3 routes.

2. Restrictions and Guidelines

- If neither the IPv6 network segment nor the IPv6 prefix is specified, the ND learned on the interface generates a 128-bit route by default.

  If no IPv6 network segment is specified and only IPv6 prefix is specified, all NDs on the interface generate the specified IPv6 prefix.

  If the IPv6 network segment is specified but no IPv6 prefix is specified, this network segment generates a 128-bit route by default.

  If both the network segment and IPv6 prefix are specified, when ND is learned, the prefix will be checked to find out the network segment it matches, and the configured IPv6 prefix is generated after the specified network segment matched is found.

  If both the network segment and IPv6 prefix are specified and the IPv6 prefix of an unspecified network segment is configured at the same time, a route will be generated according to the IPv6 prefix of the network segment if the network segment is matched, and a route will be generated according to the IPv6 prefix without network segment if the network segment is not matched.

- If you configure the IPv6 address of an interface before configuring the IPv6 prefix for the network segment of this interface address to generate a route with fixed-length mask, a configuration error occurs when the prefix-length is shorter than or equal to the mask of the IPv6 address. In addition, the IPv6 prefix command for the network segment of this interface address fails. For example, if you first configure IPv6 address 10::1/64 on the interface of VLAN 40, and then configure **ipv6 route-config nd-to-route interface vlan 40 ipv6-prefix 10::1/58 prefix-length 60** in global configuration mode, the configuration fails.

- If you configure the IPv6 address of an interface before configuring the IPv6 prefix of an unspecified network segment, a configuration error occurs when the IPv6 prefix length is shorter than or equal to the mask length of any of the IPv6 addresses on the interface. In addition, the IPv6 prefix configuration command for the unspecified network segment fails. For example, if you first configure IPv6 address 10::1/64 on the interface of VLAN 40, and then configure **ipv6 route-config nd-to-route interface vlan 40 prefix-length 60** in global configuration mode, the configuration fails.

- If you configure the IPv6 prefix of an unspecified network segment before configuring the IPv6 address of this interface, syslog is displayed and the IPv6 prefix configuration of the unspecified network segment is canceled when the mask length of one of the IPv6 addresses is longer than or equal to the IPv6 prefix of the unspecified network segment. For example, if you first configure **ipv6 route-config nd-to-route interface vlan 40 prefix-length 60** in global configuration mode, and then configure address 10::1/64 on the interface of VLAN 40, syslog alarm is displayed and the configuration **ipv6 route-config nd-to-route interface vlan 40 prefix-length 60** is deleted.
- If you configure the IPv6 prefix of the specified network segment and then configure the IPv6 address of the corresponding network segment of the interface, syslog is displayed and the IPv6 prefix configuration of the specified network segment is canceled when the IPv6 address mask of the corresponding network segment is longer than or equal to the IPv6 prefix of the specified network segment. For example, if you first configure **ipv6 route-config nd-to-route interface vlan 40 ipv6-prefix 10::1/58 prefix-length 60** in global configuration mode, and then configure the address 10::1/64 on the interface of VLAN 40, the syslog alarm is displayed and the configuration **ipv6 route-config nd-to-route interface vlan 40 ipv6-prefix 10::1/58 prefix-length 60** is deleted.
- If you configure the IPv6 prefix of an unspecified network segment and the IPv6 prefix of the specified network segment, and then configure the IPv6 address of the interface, please check whether the mask of each IPv6 address is longer than or equal to the IPv6 prefix of the specified/unspecified network segment. If so, syslog is displayed and the IPv6 prefix configuration of the corresponding network segment is canceled. For example, if you first configure **ipv6 route-config nd-to-route interface vlan 40 prefix-length 62** and **ipv6 route-config nd-to-route interface vlan 40 ipv6-prefix 10::1/58 prefix-length 60** in global configuration mode, and then configure address 10::1/64 on the interface of VLAN 40, the syslog alarm is displayed and the two configuration commands are removed.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable ND conversion to a route with fixed-length mask on the specified interface.

ipv6 route-config nd-to-route interface *interface-type interface-number* [**ipv6-prefix** *X:X:X:X::X/<0-128>*] [**prefix-length** *mask-length*]

The ND-to-route conversion function is disabled on a specified interface by default.

1.6.5 Configuring Delayed Redistribution of ND-to-route

1. Overview

To prevent route jitter caused by change to an ND entry, you can delay conversion of the ND entry into a route.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure delayed redistribution of ND-to-route.

ipv6 route-config nd-to-route delay-time *delay-time*

The delayed redistribution function of the routes converted by ND is disabled by default. The delayed redistribution time is 0 seconds.

1.6.6 Configuring Tag Attribute of ND-to-route Conversion

1. Overview

This function sets the tag value of ND-to-route conversion to facilitate subsequent route control and management.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the tag attribute of ND-to-route conversion.

ipv6 route-config nd-to-route tag *tag-number*

The tag attribute function of the ND-converted route is disabled by default.

1.6.7 Configuring to Ignore Warning Log of ND Conversion to Route

1. Overview

After you configure to ignore the warning log of ND conversion to route of specified length, the alarm that the same port ND generates the same network segment route is ignored. If this command is not configured, the warning log that the same port ND generates the same network segment route is printed (only once). The alarm that different ports generate the same network segment route is not controlled by this function.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure to ignore the warning log of ND conversion to a route of specified length.

ipv6 route-config nd-to-route warning-ignore

The function of ignoring warning log of ND conversion to a route of specified length is disabled by default.

1.6.8 Configuring ND Proxy

1. Overview

When the device receives an ND request, it will send an ND response. The MAC address of the response is the local Ethernet MAC address, so that the interconnection among different hosts is implemented by Layer 3 routing.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI interface mode.

interface vlan *vlan-id*

(4) Configure the IPv6 address of the SVI interface.

ipv6 address *ipv6-address*/*prefix-length*

(5) Enable the ND proxy.

local-proxy-nd enable

The local ND proxy is disabled by default on the interface.

1.6.9 Configuring BGP Redistributed ND-Route Route

1. Overview

In BGP protocol mode, this command is used to redistribute ND-route to BGP.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter BGP configuration mode.

router-protocol bgp *as-number*

(4) Enter BGP IPv6 address family configuration mode.

address-family-mode ipv6

(5) Configure BGP redistributed ND-Route route.

Redistribute-route nd-route [**route-map** *map-tag*] [**metric** *metric-value*]

By default, BGP does not redistribute ND-Route route.

1.7 Configuring BGP Route Deletion During GR

1.7.1 Overview

When a graceful restart (GR)-enabled router is restarted on the control plane, data forwarding can continue on the forwarding plane. In addition, actions (such as neighbor relationship re-forming and route computation) performed on the control plane do not affect functions of the forwarding plane. If the function is enabled, the BGP route deletion configuration takes effect immediately during GR to prevent routing black hole.

1.7.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure direct deletion of the BGP route during GR.

ip route-config notify delete always

Deletion of the BGP route is disabled by default.

1.8 Enabling Real-Time Active Route Statistics Collection

1.8.1 Overview

The real-time active route statistics collection function is mainly applied to scenarios in which the Google remote procedure call (gRPC) periodically samples or obtains the number of active routes or the **display ip route-config summary** command is run frequently. It can effectively shorten the time required for collecting statistics on a large number of active routes.

1.8.2 Restrictions and Guidelines

- You are advised not to enable the real-time active route statistics collection function in normal scenarios, especially in scenarios with high requirements on route convergence performance.
- L2 devices do not support this function.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the real-time active route statistics collection.

ip active-route real-time summary

The real-time active route statistics collection is disabled by default.

1.9 Monitoring

Run the **display** commands to check the running status of a configured function and verify the configuration effect.

Run the **debugging** command to display the output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** command to clear the information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-2 Monitoring

Command	Purpose
cleanup ip route-config [**vrf** *vrf-name*] { ***** | *network* [*mask*] }	Clears the routing table cache.
display ip route-config [**vr**f vrf-name] [*network* [*mask* [**longer-prefix**]] | **aggregate** | **arphost** | **bgp** [**evpn**] | **connected** | **count** | [**ecmp** | **fast-reroute** | **normal**] [*network* [*mask*]] | **isis** [*tag*] | **ospf** [*process-id*] | **rip** | **static**| **tag** | **weight**]	Displays the IPv4 routing table.
display ipv6 route-config [**vrf** *vrf-name*] [**bgp** [**evpn**] | **connected** | [**fast-reroute**] *ipv6-prefix/prefix-length* [**longer-prefixes**] | **isis** | **local** | **ndroute** | **ospf** | **rip** | **static** | **tag** | **weight**]	Displays the IPv6 routing table.
display ip route-config [**vrf** *vrf-name*] **recursive** *ipv4-network/mask*	Displays the recursive information of an IP route.
display ip route-config [[**vrf** *vrf-name*] **static-route bfd**	Displays the BFD correlation information of an IP route.
display ip route-config [**vrf** *vrf-name*] **status-route**	Displays the statistics of a single routing table.
display ip route-config [**vrf** *vrf-name*] **track-route-table**	Displays the track information of IPv4 route correlation.
display ipv6 route-config [[**vrf** *vrf-name*] **static-route bfd**	Displays the BFD correlation information of an IPv6 route.

Command	Purpose
display ipv6 route-config [**vrf** *vrf-name*] **status-route**	Displays the statistics of all the IPv6 routing tables.
display nsm develop { **basic** | **bfd peer | block | client** [*client-name*] **| interface** [*interface-type interface-number*] **| memory | mom | route-num | vrf** [*vrf-name*] }	Displays route-related information.
display route-res usage [**all** | **slot** *slot-id*]	Displays the usage of routing resources.
debugging nsm kernel ucast-v4	Debugs IPv4 route management.
debugging nsm kernel ucast-v6	Debugs Ipv6 route-config management.
debugging nsm kernel frr	Debugs fast reroute management.
debugging nsm kernel default-network-ip	Debugs default network management.
debugging nsm events	Debugs internal events of route management.
debugging nsm packet send	Debugs sending of route management and routing protocol messages.
debugging nsm packet recv	Debugs receiving of route management and routing protocol messages.

Contents

1 Configuring Static Routing 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 2
1.3 Configuring Static Routes 3
1.3.1 Overview 3
1.3.2 Restrictions and Guidelines 3
1.3.3 Configuration Tasks 4
1.3.4 Configuring a Static IPv4 Route 4
1.3.5 Configuring an IPv6 Static Route 4
1.4 Configuring Static Gateway 5
1.4.1 Overview 5
1.4.2 Configuration Tasks 5
1.4.3 Configuring the IPv4 Gateway on a Layer-2 Switch 5
1.4.4 Configuring the IPv6 Gateway on a Layer-2 Switch 5
1.4.5 Configuring the Default IPv4 Gateway on a Layer-3 Switch 6
1.4.6 Configuring the Default IPv6 Gateway on a Layer-3 Switch 6
1.4.7 Configuring the Default IPv4 Network on a Layer-3 Switch 7
1.5 Configuring Routing Black Hole 7
1.5.1 Overview 7
1.5.2 Procedure 8

1.6 Configuring Route Limitations 8
1.6.1 Overview 8
1.6.2 Restrictions and Guidelines 8
1.6.3 Configuration Tasks 8
1.6.4 Configuring the Maximum Number of IPv4 Static Routes 8
1.6.5 Configuring the Maximum Number of IPv6 Static Routes 9
1.6.6 Prohibiting Static Routing Across VRFs 9
1.6.7 Configuring the Maximum Number of VRF Routes 9
1.7 Correlating a Static Route with BFD 10
1.7.1 Overview 10
1.7.2 Restrictions and Guidelines 10
1.7.3 Configuration Tasks 10
1.7.4 Configuring Correlation of IPv4 Static Routes with BFD 10
1.7.5 Configuring Correlation of IPv6 Static Routes with BFD 11
1.8 Configuring Static Fast Reroute 11
1.8.1 Overview 11
1.8.2 Restrictions and Guidelines 11
1.8.3 Procedure 11
1.9 Configuring Correlation of Static Routes with ARP 12
1.9.1 Overview 12
1.9.2 Procedure 12
1.10 Monitoring 12
1.11 Configuration Examples 13
1.11.1 Configuring an IPv4 Static Route 13

1.11.2 Configuring Correlation of IPv4 Static Routes with Track 17

1.11.3 Configuring Correlation of Static Routes with BFD 19

1.11.4 Configuring Fast Reroute 21

1 Configuring Static Routing

1.1 Introduction

1.1.1 Overview

A static route is manually configured by a network administrator. A static route is easy to configure and less demanding on the system, and therefore applicable to a small-sized network that is stable and has a simple topology. Appropriate configuration and use of static routes improves the network performance and guarantees the bandwidth for important network applications. However, when the network topology changes, the static route must be manually reconfigured and cannot automatically adapt to the topological changes.

Note

In this chapter, the term "router" refers to any network device that supports the routing function. These network devices can be Layer-3 switches, routers, or firewalls.

1.1.2 Principles

1. Static Route Basic

Whether a static route is active is computed based on the status of the local interface. When the exit interface of a static route is located at Layer-3 and is in Up status (the link status is Up and the IP address is configured), this route is active and can be used for packet forwarding.

A static route can go across VPN routing & forwarding (VRF) instances. The next hop or exit interface of a static route of VRF 1 can be configured on VRF 2.

A static route consists of two parts:

- Destination IP address and mask: The destination IP address of the packet is matched according to the destination address and mask, and the longest matching principle is adopted. The mask can be represented in dotted decimal form or using mask length.
- Outbound interface or next hop address: The next hop of a static route can be the IP address of the outbound interface or next hop router. On a point-to-point network, specifying the outbound interface indirectly determines the next hop router. However, on the Non Broadcast Multiple Access (NBMA) or broadcast network, using an outbound interface as the next hop will lead to multiple next hops. Therefore, you are advised to set the IP address for the next hop router on the NBMA and broadcast networks.

2. Default Route

In a forwarding routing table, the route with the destination network segment 0.0.0.0 and the subnet mask 0.0.0.0 is the default route. Packets that cannot be forwarded by other routes will be forwarded by the default route. The default route can be statically configured or generated by a dynamic routing protocol.

- Default gateway

  On a Layer-2 switch, run the **ip default gateway-ip** command to generate a default route.

- Static default route

On a Layer-3 switch, a static route with the network segment 0.0.0.0 and the subnet mask 0.0.0.0 is configured to generate a default route.

- Default network

  The default network is configured to generate a default route. If the **ip default-network-ip** command is configured to specify a network (a classful network, such as a Class A, B, or C network), and this network exists in the routing table, the router will use this network as the default network and the next hop of this network is the default gateway. As the network specified by the **ip default-network-ip** command is a classful one, if this command is used to identify a subnet in a classful network, the router automatically generates a static route of the classful network instead of any default route.

3. Static route reliability

When a device on a network is faulty, some routes become unreachable, resulting in traffic interruption. If connectivity of the next hop can be detected in real time, the route can be re-computed when a fault occurs, or traffic can be switched over to the standby route.

- Correlation with track

  A track object is an abstract concept. It can be used to trace whether an IP address is reachable or whether an interface is Up. If a dynamic routing protocol or a static route is correlated with the Track function, the dynamic routing protocol or the static route can quickly learn whether the next hop is reachable so as to respond quickly.

- Correlation with BFD

  The bidirectional forwarding detection (BFD) protocol provides a light-load and fast method for detecting the connectivity of the forwarding path between two adjacent routers. If a dynamic routing protocol or a static route is correlated with the BFD function, the dynamic routing protocol or the static route can quickly learn whether the next hop is reachable so as to respond quickly.

Note

The detection performance of BFD is better than that of Track.

- Correlation with ARP

  If a static route is correlated with the ARP function, when the ARP does not exist, the static route can quickly learn that the next hop is unreachable so as to respond quickly.

- Fast reroute

  Fast reroute provides a standby route. When a dynamic routing protocol or a static route detects that the next hop is unreachable, it immediately switches traffic over to the standby route to recover communication.

1.2 Configuration Task Summary

The configuration of static route includes the following tasks:

(1) Configuring Static Routes

- Configuring a Static IPv4 Route
- Configuring an IPv6 Static Route

(2) (Optional) Configuring Static Gateway The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring the IPv4 Gateway on a Layer-2 Switch
- Configuring the IPv6 Gateway on a Layer-2 Switch
- Configuring the Default IPv4 Gateway on a Layer-3 Switch
- Configuring the Default IPv6 Gateway on a Layer-3 Switch
- Configuring the Default IPv4 Network on a Layer-3 Switch

(3) (Optional) Configuring Routing Black Hole Black Hole The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

(4) (Optional) Configuring Route Limitations The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring the Maximum Number of IPv4 Static Routes
- Configuring the Maximum Number of IPv6 Static Routes
- Prohibiting Static Routing Across VRFs
- Configuring the Maximum Number of VRF Routes

(5) (Optional) Correlating a Static Route with BFD The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring Correlation of IPv4 Static Routes with BFD
- Configuring Correlation of IPv6 Static Routes with BFD

(6) (Optional) Configuring Static Fast Reroute

(7) (Optional) Configuring Correlation of Static Routes with ARP

1.3 Configuring Static Routes

1.3.1 Overview

Generate a static route in the routing table. The static route can be used to forward packets to a remote network.

1.3.2 Restrictions and Guidelines

- This function cannot be configured on a Layer-2 switch.
- If the **unset ip routing** command is configured on a Layer-3 switch, you cannot configure IPv4 static routes on this switch, and existing IPv4 static routes will also be deleted. Before the device is restarted, reconfiguring the **ip routing** command can recover the deleted IPv4 static routes. After the device is restarted, deleted IPv4 static routes cannot be recovered.
- If the **unset ipv6 unicast- routing** command is configured on a Layer-3 switch, you cannot configure IPv6 static routes on this switch, and existing IPv6 static routes will also be deleted. Before the device is restarted, reconfiguring the **ipv6 unicast- routing** command can recover the deleted IPv6 static routes. After the device is restarted, deleted IPv6 static routes cannot be recovered.
- To correlate a static route with the track function, you must run the **track** command to configure a track object.

1.3.3 Configuration Tasks

The configuration of static route includes the following tasks:

- Configuring a Static IPv4 Route
- Configuring an IPv6 Static Route

1.3.4 Configuring a Static IPv4 Route

1. Overview

In a simple network, you can configure IPv4 static routes to implement internetworking.

2. Restrictions and Guidelines

- If the static route is correlated with track and the Down status of the track object is detected, the static route is not active and does not participate in packet forwarding.
- If the static route is correlated with ARP, but no ARP information of the next hop and egress is detected, the static route is not active and does not participate in packet forwarding.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an IPv4 static route.

ip route-config [**vrf-table** *vrf-name*] *network mask* { *ipv4-address* | *interface* [*ipv4-address*] } [*distance* | **tag** *tag* | **permanent** | **track** *object-number* | **weight** *number* | **description** *description-text* | [**disabled** | **enabled**] | **bfd** *session-name*]

No static route is configured by default.

1.3.5 Configuring an IPv6 Static Route

1. Overview

In a simple network, you can configure IPv6 static routes to implement internetworking.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an IPv6 static route.

ipv6 route-config [**vrf-table** *vrf-name*] *ipv6-prefix*/*prefix-length* { *ipv6-address* [**next-hop-vrf** { *vrf-name1* | **default** }] | *interface* [*ipv6-address* [**next-hop-vrf** { *vrf-name1* | **default** }]] } [*distance* | **tag** *tag* | **track** *object-number* | **weight** *number* | **description** *description-text* | **bfd** *session-name*]

No IPv6 static route is configured by default.

1.4 Configuring Static Gateway

1.4.1 Overview

The router in an enterprise does not need to obtain all the routing information of the whole Internet. A default route is generated in the routing table and points to the operator gateway, and then the Internet can be accessed. The default route is used to forward packets that cannot be forwarded by other routes.

1.4.2 Configuration Tasks

The default static route configuration tasks are as follows and optional. Select tasks for configuration according to the actual condition.

- Configuring the IPv4 Gateway on a Layer-2 Switch
- Configuring the IPv6 Gateway on a Layer-2 Switch
- Configuring the Default IPv4 Gateway on a Layer-3 Switch
- Configuring the Default IPv6 Gateway on a Layer-3 Switch
- Configuring the Default IPv4 Network on a Layer-3 Switch

1.4.3 Configuring the IPv4 Gateway on a Layer-2 Switch

1. Overview

On a Layer-2 device, you can run the **ip default-gateway-ip** command to specify the default IPv4 route. During data interaction, the packet with a destination address not in the local network segment is sent to the default gateway, and the gateway router completes the next-step routing, achieving internetworking between the Layer-2 device and other networks.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the IPv4 gateway on a Layer-2 switch.

ip default-gateway-ip *gateway*

No default gateway is configured by default.

1.4.4 Configuring the IPv6 Gateway on a Layer-2 Switch

1. Overview

On a Layer-2 device, you can run the **ipv6 default-gateway** command to specify the default IPv6 route. During data interaction, the packet with a destination address not in the local network segment is sent to the default gateway, and the gateway router completes the next-step routing, achieving internetworking between the Layer-2 device and other networks.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the IPv6 gateway on a Layer-2 switch.

ipv6 default-gateway-ip *gateway*

No default IPv6 gateway is configured by default.

1.4.5 Configuring the Default IPv4 Gateway on a Layer-3 Switch

1. Overview

On a Layer-3 switching device and routing device, when the routing for packet sending cannot match an accurate route, the default gateway can be configured so that non-local packets are sent to the default gateway, and then the gateway router completes the next-step routing, achieving internetworking between the local device and other networks.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the default IPv4 gateway on a Layer-3 switch.

ip route-config [**vrf-table** *vrf-name*] **0.0.0.0 0.0.0.0** { *ipv4-address* | *interface* [*ipv4-address*] } [*distance*] [**tag** *tag*] [**permanent** | { **track** *object-number* | **arp** }] [**weight** *number*] [**description** *description-text*] [**disabled** | **enabled**]

ip route-config 0.0.0.0 0.0.0.0 { *ipv4-address* | *interface* [*ipv4-address*] } [*distance*] [**tag** *tag*] [**permanent** | { **track** *object-number* | **arp** }] [**weight** *number*] [**description** *description-text*] [**disabled** | **enabled**]

No default IPv4 gateway is configured on the Layer-3 switch.

1.4.6 Configuring the Default IPv6 Gateway on a Layer-3 Switch

1. Overview

On a Layer-3 switching device and routing device, when the routing for packet sending cannot match an accurate route, the default gateway can be configured so that non-local packets are sent to the default gateway, and then the gateway router completes the next-step routing, achieving internetworking between the local device and other networks.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the default IPv6 gateway on a Layer-3 switch.

ipv6 route-config [**vrf-table** *vrf-name*] **::/0** { *ipv6-address* [**next-hop-vrf** { *vrf-name1* | **default** }] | *interface* [*ipv6-address* [**next-hop-vrf** { *vrf-name1* | **default** }]] } [*distance*] [**tag** *tag*] [**weight** *number*] [**description** *description-text*]

ipv6 route::/0 { *ipv6-address* | *interface* [*ipv6-address*] } [*distance*] [**weight** *number*] [**description** *description-text*]

No default IPv6 gateway is configured on the Layer-3 switch.

1.4.7 Configuring the Default IPv4 Network on a Layer-3 Switch

1. Overview

On a Layer-3 switching device and routing device, when the routing for packet sending cannot match an accurate route, the default gateway can be configured so that non-local packets are sent to the default gateway, and then the gateway router completes the next-step routing, achieving internetworking between the local device and other networks.

2. Restrictions and Guidelines

If the network specified by the **ip default-network** command exists in the routing table, a default route is generated and the next hop to this network is the default gateway. If the network specified by the **ip default-network** command does not exist, the default route is not generated.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the default IPv4 network on a Layer-3 switch.

ip default-network-ip *network*

No default IPv4 network is configured on the Layer-3 switch.

1.5 Configuring Routing Black Hole

1.5.1 Overview

Routing black hole can avoid the problem of route loop when the route summary and default route are configured at the same time. If the 192.168.1.0/24 and 192.168.2.0/24 network segments exist on the internal network, the default route of Router A will point to the gateway, and the two internal network segments are summarized into 192.168.0.0/16 and notified to the gateway through the dynamic routing protocol. When the gateway initiates access to the address 192.168.3.1/24 not existing on the internal network to Router A, Router A will forward the access request back to the gateway according to the default route, thus causing a route loop. Null 0 is a virtual interface never in Down status, and Router A is configured with 192.168.0.0/16 to point to Null 0. Then, the request sent by the gateway to the router destination address 192.168.3.1/24 will be discarded by Router A.

Configuring a routing black hole can make full use of the forwarding performance of routers and avoid route loops. This function is not supported on Layer-2 switches.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an aggregated route.

ip route-config aggregate [**vrf-table** *vrf-name*] *network mask* **Null 0** [**tag** *tag*]

The aggregated route is disabled by default.

1.6 Configuring Route Limitations

1.6.1 Overview

This function limits the number of equal-cost routes and number of static routes, or disables routing.

1.6.2 Restrictions and Guidelines

This function cannot be configured on a Layer-2 switch.

1.6.3 Configuration Tasks

The route limitation configuration tasks are as follows and optional. Select tasks for configuration according to the actual condition.

- Configuring the Maximum Number of IPv4 Static Routes
- Configuring the Maximum Number of IPv6 Static Routes
- Prohibiting Static Routing Across VRFs
- Configuring the Maximum Number of VRF Routes

1.6.4 Configuring the Maximum Number of IPv4 Static Routes

1. Overview

This command can be used to configure the maximum number of IPv4 static routes. When the maximum number of IPv4 static routes is reached, no more IPv4 static routes can be configured.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of IPv4 static routes.

ip static-route max-route { *route-limit* | **default-vrf-table** *default-vrf-route-limit* | **vrf-table** *vrf-name vrf-route-limit* }

The configured maximum number of static routes is 1000 by default.

1.6.5 Configuring the Maximum Number of IPv6 Static Routes

1. Overview

This command can be used to configure the maximum number of IPv6 static routes. When the maximum number of IPv6 static routes is reached, no more IPv6 static routes can be configured.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of IPv6 static routes.

ipv6 static-route max-route { *route-limit* | **default-vrf-table** *default-vrf-route-limit* | **vrf-table** *vrf-name vrf-route-limit* }

The configured maximum number of static routes is 1000 by default.

1.6.6 Prohibiting Static Routing Across VRFs

1. Overview

You can use this function to prohibit static IP routing across VRFs. After this command is configured, the static IP routing across VRFs is not active and cannot participate in packet forwarding.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Prohibit static IP routing across VRFs.

unset ip route-config static- route cross-vrf

Static routing across VRFs is allowed by default.

1.6.7 Configuring the Maximum Number of VRF Routes

1. Overview

This function is used to limit the number of routes running in the default VRF. You can receive warnings only through configuration.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of routes in the default VRF.

maximum routes-number *limit* { *warn-threshold* | **warning** }

The maximum number of routes in the default VRF is not limited by default.

1.7 Correlating a Static Route with BFD

1.7.1 Overview

A static route can quickly detect a route failure with the help of BFD.

1.7.2 Restrictions and Guidelines

- This function cannot be configured on a Layer-2 switch.
- You must configure a static route before configuring this function.
- Before configuring this function, you must configure BFD session parameters using the **set-bfd period min_rx multiplier** command.

1.7.3 Configuration Tasks

The configuration tasks for correlation of static routes with BFD are as follows and optional. Select tasks for configuration according to the actual condition.

- Configuring Correlation of IPv4 Static Routes with BFD
- Configuring Correlation of IPv6 Static Routes with BFD

1.7.4 Configuring Correlation of IPv4 Static Routes with BFD

1. Overview

You can use this function to configure correlation of IPv4 static routes with BFD, quickly sensing change to the destination address link. If the Down status of the BFD session is detected, the IPv4 static route is not active and does not participate in packet forwarding.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure correlation of IPv4 static routes with BFD.

ip route-config static-route bfd [**vrf-table** *vrf-name*] *interface-type interface-number gateway* [**source** *ipv4-address*]

ip route-config static-route bfd *interface-type interface-number gateway* [**source** *ipv4-address*]

Correlation of static routes with BFD is disabled by default.

1.7.5 Configuring Correlation of IPv6 Static Routes with BFD

1. Overview

You can use this function to configure correlation of IPv6 static routes with BFD, quickly sensing change to the destination address link. If the Down status of the BFD session is detected, the IPv6 static route is not active and does not participate in packet forwarding.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure correlation of IPv6 static routes with BFD.

ipv6 route-config static-route bfd [**vrf-table** *vrf-name*] *interface-type interface-number gateway* [**source** *ipv6-address*]

ipv6 route-config static-route bfd *interface-type interface-number gateway* [**source** *ipv6-address*]

Correlation of static routes with BFD is disabled by default.

1.8 Configuring Static Fast Reroute

1.8.1 Overview

If a routing link fails on the network, the service will be interrupted and cause loss. To avoid the impact of a link failure, the static fast reroute function can be used to switch the link when a link fails.

1.8.2 Restrictions and Guidelines

- This function cannot be configured on a Layer-2 switch.
- You must configure a static route before configuring this function.
- You must configure a route map before configuring this function.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a routing policy.

route-map *route-map-name* [{ **permit** | **deny** } *sequence*]

(4) Create a matching rule.

filter interface *interface-type interface-number*

(5) Configure to define a backup route in the route map.

set-rule frr backup-interface *interface-type interface-number* **backup-nexthop** *ipv4-address*

By default, no backup egress or backup next hop is specified for the matched route.

(6) Return to the global configuration mode.

exit

(7) Configure IPv4 static fast reroute.

ip frr [**vrf-table** *vrf-name*] **static-route route-map** *route-map-name*

ip frr static-route route-map *route-map-name*

The IPv4 static fast reroute function is disabled by default.

(8) Configure IPv6 static fast reroute.

ip frr [**vrf-table** *vrf-name*] **staticroute route-map** *route-map-name*

ip frr static-route route-map *route-map-name*

The IPv6 static fast reroute function is disabled by default.

1.9 Configuring Correlation of Static Routes with ARP

1.9.1 Overview

This function is used to configure correlation of static routes with ARP so that ARP requests are actively sent when ARP fails.

1.9.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the interval for actively sending ARP requests.

ip route-config scan arp interval *request-interval*

The correlation of static routes with ARP is disabled by default, and the ARP request is sent every 5s.

(4) Configure the times of actively sending ARP requests.

ip route-config scan arp times *request-times*

By default, the correlation of static routes with ARP is disabled, and the ARP request is sent 65,535 times.

1.10 Monitoring

This section describes the **display** commands used for checking the running status of a configured function to verify the configuration effect.

Run the **debugging** command to display the output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** command to clear the information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup ip route-config arp scan times	Clears the ARP request transmission count and starts counting again.
display ip redirects-info	Displays the default gateway.
display ipv6 redirects-info	Displays the default IPv6 gateway.

1.11 Configuration Examples

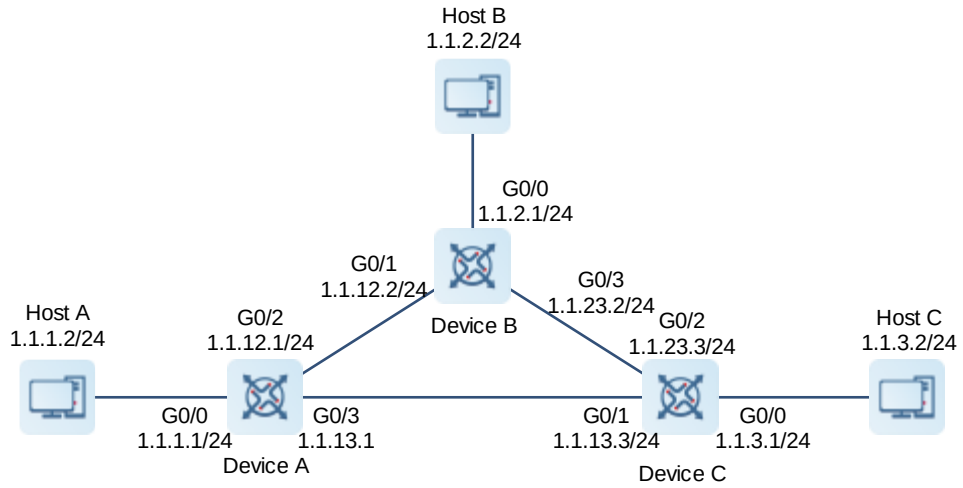
1.11.1 Configuring an IPv4 Static Route

1. Requirements

Configure static routes to make routes on the entire network reachable.

2. Topology

Figure 1-1 Configuring an IPv4 Static Route



3. Notes

- Configure an address on each interface of the device.
- Configure static routes on the device.

4. Procedure

(1) Configure an IP addresses for each interface.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/0
Device A(cfg-if-GigabitEthernet 0/0)# ip address 1.1.1.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/0)# exit
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# ip address 1.1.12.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/2)# exit
Device A(cfg)# interface gigabitethernet 0/3
Device A(cfg-if-GigabitEthernet 0/3)# ip address 1.1.13.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/3)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/0
Device B(cfg-if-GigabitEthernet 0/0)# ip address 1.1.2.1 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/0)# exit
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip address 1.1.12.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/3
Device B(cfg-if-GigabitEthernet 0/3)# ip address 1.1.23.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/3)# exit
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# interface gigabitethernet 0/0
Device C(cfg-if-GigabitEthernet 0/0)# ip address 1.1.3.1 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/0)# exit
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# ip address 1.1.13.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface gigabitethernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# ip address 1.1.23.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/2)# exit
```

(2) Enable the IPv4 static route function.

Configure Device A.

```
Device A(cfg)# ip route-config 1.1.2.0 255.255.255.0 gigabitethernet 0/2 1.1.12.2
Device A(cfg)# ip route-config 1.1.3.0 255.255.255.0 gigabitethernet 0/3 1.1.13.3
```

Configure Device B.

```
Device B(cfg)# ip route-config 1.1.1.0 255.255.255.0 gigabitethernet 0/1 1.1.12.1
Device B(cfg)# ip route-config 1.1.3.0 255.255.255.0 gigabitethernet 0/3 1.1.23.3
```

Configure Device C.

```
Device C(cfg)# ip route-config 1.1.2.0 255.255.255.0 gigabitethernet 0/2 1.1.23.2
Device C(cfg)# ip route-config 1.1.1.0 255.255.255.0 gigabitethernet 0/1 1.1.13.1
```

5. Verification

Check the routing table of the device and confirm that there are static route entries.

Check the routing table of Device A.

```
Device A# display ip route-config
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPF, B - BGP, I - IS-IS
         N1 - OSPF NSSA external-type type 1, N2 - OSPF NSSA external-type type 2
         E1 - OSPF external-type type 1, E2 - OSPF external-type type 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area, * - candidate default

Gateway of last resort is no set
C       1.1.1.0/24 is directly connected, GigabitEthernet 0/0
C       1.1.1.1/32 is local host.
S       1.1.2.0/24 [1/0] via 1.1.12.2, GigabitEthernet 0/2
S       1.1.3.0/24 [1/0] via 1.1.13.3, GigabitEthernet 0/2
C       1.1.12.0/24 is directly connected, GigabitEthernet 0/2
C       1.1.12.1/32 is local host.
C       1.1.13.0/24 is directly connected, GigabitEthernet 0/3
C       1.1.13.1/32 is local host.
```

Check the routing table of Device B.

```
Device B# display ip route-config
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPF, B - BGP, I - IS-IS
         N1 - OSPF NSSA external-type type 1, N2 - OSPF NSSA external-type type 2
         E1 - OSPF external-type type 1, E2 - OSPF external-type type 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area, * - candidate default

Gateway of last resort is no set
S       1.1.1.0/24 [1/0] via 1.1.12.1, GigabitEthernet 0/0
C       1.1.2.0/24 is directly connected, GigabitEthernet 0/0
C       1.1.2.1/32 is local host.
S       1.1.3.0/24 [1/0] via 1.1.23.3, GigabitEthernet 0/3
C       1.1.12.0/24 is directly connected, GigabitEthernet 0/1
C       1.1.12.2/32 is local host.
C       1.1.23.0/24 is directly connected, GigabitEthernet 0/3
C       1.1.23.2/32 is local host.
```

Check the routing table of Device C.

```
Device C# display ip route-config
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPF, B - BGP, I - IS-IS
         N1 - OSPF NSSA external-type type 1, N2 - OSPF NSSA external-type type 2
         E1 - OSPF external-type type 1, E2 - OSPF external-type type 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area, * - candidate default

Gateway of last resort is no set
S       1.1.1.0/24 [1/0] via 1.1.13.1, GigabitEthernet 0/2
S       1.1.2.0/24 [1/0] via 1.1.23.2, GigabitEthernet 0/2
C       1.1.3.0/24 is directly connected, GigabitEthernet 0/0
C       1.1.3.1/32 is local host.
C       1.1.13.0/24 is directly connected, GigabitEthernet 0/1
C       1.1.13.3/32 is local host.
C       1.1.23.0/24 is directly connected, GigabitEthernet 0/2
C       1.1.23.3/32 is local host.
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/0
 ip address 1.1.1.1 255.255.255.0
!
interface gigabitethernet 0/2
 ip address 1.1.12.1 255.255.255.0
!
interface gigabitethernet 0/3
 ip address 1.1.13.1 255.255.255.0
!
ip route-config 1.1.2.0 255.255.255.0 GigabitEthernet 0/2 1.1.12.2
ip route-config 1.1.3.0 255.255.255.0 GigabitEthernet 0/3 1.1.13.3
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/0
 ip address 1.1.2.1 255.255.255.0
!
interface gigabitethernet 0/1
 ip address 1.1.12.2 255.255.255.0
!
interface gigabitethernet 0/3
 ip address 1.1.23.2 255.255.255.0
!
ip route-config 1.1.1.0 255.255.255.0 GigabitEthernet 0/1 1.1.12.1
```

```
ip route-config 1.1.3.0 255.255.255.0 GigabitEthernet 0/3 1.1.23.3
!
```

- Device C configuration file

```
!
interface gigabitethernet 0/0
 ip address 1.1.3.1 255.255.255.0
!
interface gigabitethernet 0/1
 ip address 1.1.13.3 255.255.255.0
!
interface gigabitethernet 0/2
 ip address 1.1.23.3 255.255.255.0
!
ip route-config 1.1.2.0 255.255.255.0 GigabitEthernet 0/2 1.1.23.2
ip route-config 1.1.1.0 255.255.255.0 GigabitEthernet 0/1 1.1.13.1
!
```

7. Common Errors

- The link on the interface is not Up.
- No IP address is configured for the interface.

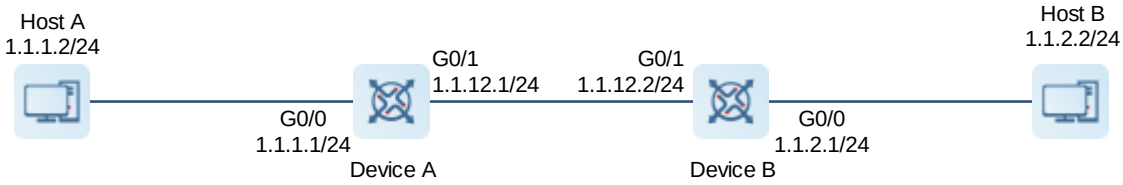
1.11.2 Configuring Correlation of IPv4 Static Routes with Track

1. Requirements

Device A and Device B implement the interworking between Host A 1.1.1.2 and Host B 1.1.2.2 through a static route. To know the link reachability in time, use correlation of IPv4 static routes with track.

2. Topology

Figure 1-2 Configuring Correlation of IPv4 Static Routes with Track



3. Notes

- Configure an address on each interface of the device.
- Configure static routes on Device A and Device B, and specify the exit interface or next hop as the interworking interface.
- Correlate static routes with track on Device A and Device B, and check the connectivity of the next hops of static routes.

4. Procedure

(1) Configure an IP address for each interface and start the unicast routing protocol.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip address 1.1.12.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/1)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip address 1.1.12.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# exit
```

(2) Configure the track function.

Configure Device A.

```
Device A(cfg)# track 2 interface gigabitethernet 0/1 line-protocol
```

Configure Device B.

```
Device B(cfg)# track 2 interface gigabitethernet 0/1 line-protocol
```

(3) Configure correlation of static routes with track.

Configure Device A.

```
Device A(cfg)# ip route-config 1.1.2.0 255.0.0.0 gigabitethernet 0/1 1.1.12.2 track 2
```

Configure Device B.

```
Device B(cfg)# ip route-config 1.1.1.0 255.0.0.0 gigabitethernet 0/1 1.1.12.1 track 2
```

5. Verification

Display the statistics of Track 2.

```
Device A# display track 2
Track 2
Interface gigabitEthernet 0/1
   The state is Up, delayed Down (5 secs remaining)
      1 change, current state last: 300 secs
   Delay up 0 secs, down 0 secs
```

Display the track correlation information and state of an IP route.

```
Device A# display ip route-config track-table
ip route-config 1.1.2.0 255.0.0.0 GigabitEthernet 0/1 1.1.12.2 track 2 up
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
  ip address 1.1.12.1 255.255.255.0
!
track 2 interface gigabitEthernet 0/1 line-protocol
```

```
!
ip route-config 1.1.2.0 255.0.0.0 gigabitEthernet 0/1 1.1.12.2 track 2
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
 ip address 1.1.12.2 255.255.255.0
!
track 2 interface gigabitethernet 0/1 line-protocol
!
ip route-config 1.1.1.0 255.0.0.0 gigabitethernet 0/1 1.1.12.1 track 2
!
```

7. Common Errors

- The static route is correlated with track, but no track object is configured.

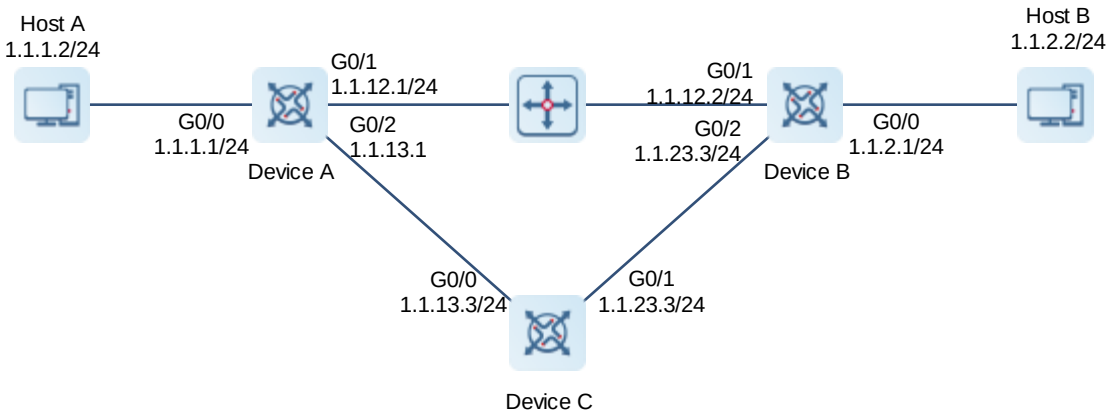
1.11.3 Configuring Correlation of Static Routes with BFD

1. Requirements

Configure static routes for Host A and Host B to realize interworking, where Device A-Device B is the primary link. After BFD is configured, if the primary link fails, packets can switch to Device C quickly to implement communication.

2. Topology

Figure 1-3 Configuring Correlation of Static Routes with BFD



3. Notes

- Configure static routes on Device A, Device B, and Device C, and specify the exit interface or next hop as the interworking interface.
- Configure a BFD session on the interconnect interface between Device A and Device B.
- Correlate static routes with BFD on Device A and Device B, and check the connectivity of the next hops of static routes.

4. Procedure

(1) Configure IP addresses of interfaces (omitted).

(2) Configure static routes to implement internetworking.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ip route-config 1.1.2.0 255.0.0.0 gigabitethernet 0/1 1.1.12.2
Device A(cfg)# ip route-config 1.1.2.0 255.0.0.0 gigabitethernet 0/2 1.1.13.3 40
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ip route-config 1.1.1.0 255.0.0.0 gigabitethernet 0/1 1.1.12.1
Device B(cfg)# ip route-config 1.1.1.0 255.0.0.0 gigabitethernet 0/2 1.1.23.3 40
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# ip route-config 1.1.1.0 255.0.0.0 gigabitethernet 0/0 1.1.13.1
Device C(cfg)# ip route-config 1.1.2.0 255.0.0.0 gigabitethernet 0/1 1.1.23.2
```

(3) Configure BFD for an interface.

Configure Device A.

```
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# set-bfd period 100 min_rx 100 multiplier 3
Device A(cfg-if-GigabitEthernet 0/1)# exit
```

Configure Device B.

```
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# set-bfd period 100 min_rx 100 multiplier 3
Device B(cfg-if-GigabitEthernet 0/1)# exit
```

(4) Configure correlation of static routes with BFD.

Configure Device A.

```
Device A(cfg)# ip route-config static-route bfd gigabitethernet 0/1 1.1.12.2
```

Configure Device B.

```
Device B(cfg)# ip route-config static-route bfd gigabitethernet 0/1 1.1.12.1
```

5. Verification

Display the states of BFD neighbors.

```
Device A# display set-bfd neighbors-table
OurAddr       NeighAddr      LD/RD     RH/RS     Holdown(mult)   State   Int
1.1.12.1      1.1.12.2       8192/0    Up        0(3    )        Up GigabitEthernet 0/1
```

Display the states of BFD neighbors.

```
Device A# display ip route-config static-route bfd
```

```
S       1.1.2.0/24 via 1.1.12.2, GigabitEthernet 0/1, BFD state is Up
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
 set-bfd period 50 min_rx 50 multiplier 3
!
ip route-config 1.1.2.0 255.0.0.0 gigabitEthernet 0/1 1.1.12.2
ip route-config 1.1.2.0 255.0.0.0 gigabitEthernet 0/2 1.1.13.3 40
!
ip route-config static-route bfd gigabitEthernet 0/1 1.1.12.2
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
 set-bfd period 50 min_rx 50 multiplier 3
!
ip route-config 1.1.1.0 255.0.0.0 gigabitethernet 0/1 1.1.12.1
ip route-config 1.1.1.0 255.0.0.0 gigabitethernet 0/2 1.1.23.3 40
!
ip route-config static-route bfd gigabitEthernet 0/1 1.1.12.1
!
```

- Device C configuration file

```
!
ip route-config 1.1.1.0 255.0.0.0 gigabitethernet 0/0 1.1.13.1
ip route-config 1.1.2.0 255.0.0.0 gigabitethernet 0/1 1.1.23.2
!
```

7. Common Errors

- The link on the interface is not Up.
- No IP address is configured for the interface.
- No BFD session parameters are configured.
- No static route is configured.

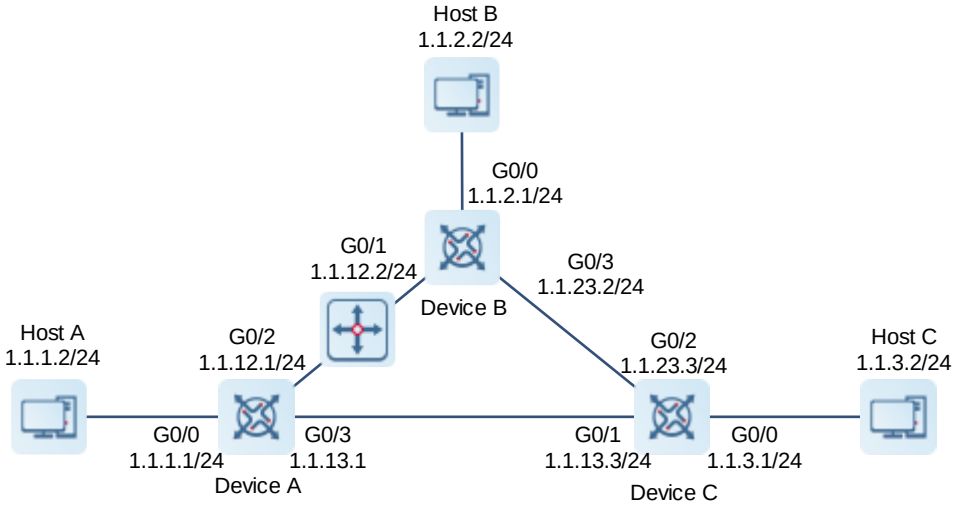
1.11.4 Configuring Fast Reroute

1. Requirements

Device A, Device B, and Device C are interconnected through static routes, and Device A accesses Host C through Device C by default. In consideration of the possible link failure of Device A and Device C, configure fast reroute to quickly switch to the backup link in case of a link failure.

2. Topology

Figure 1-4 Configuring Fast Reroute



3. Notes

- On Device A, configure a static route to the network segment of Host C, and the next hop of the exit interface is Device C.
- On Device A, configure static fast reroute. The next hop of the exit interface of the backup route is Device B.

4. Procedure

(1) Configure an IP address for the router (omitted).

(2) Configure static routes to implement internetworking between all the hosts (omitted).

(3) Configure a route map, and set the next hop of the exit interface of the backup route to Device B.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# route-map frr
Device A(cfg-routemap)# set-rule frr backup-interface gigabitethernet 0/2 backup-nexthop 1.1.12.2
```

(4) Configure static fast reroute.

```
Device A(cfg-routemap)# exit
Device A(cfg)# ip frr static-route route-map frr
```

5. Verification

Display the primary and backup routes on Device A.

```
Device A# display ip route-config frr
Codes:  C - connected, S - static, R - RIP, B - BGP
        O - OSPF, IA - OSPF inter area
        N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
        E1 - OSPF external type 1, E2 - OSPF external type 2
        i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
```

```
        ia - IS-IS inter area, * - candidate default
        Status codes: m - main entry, b - backup entry, a - active entry

Gateway of last resort is no set
S       1.1.3.0 /24 [ma] via 1.1.13.3, GigabitEthernet 0/3
                    [b]  via 1.1.12.2, GigabitEthernet 0/2
```

6. Configuration Files

- Device A configuration file

```
!
route-map frr
set-rule frr backup-interface gigabitethernet 0/2 backup-nexthop 1.1.12.2
!
ip frr static-route route-map frr
!
```

7. Common Errors

- The link on the interface is not Up.
- No static route is configured.
- The **filter** matching conditions are not configured or are not correctly configured in the road map.

Contents

1 Configuring RIP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 9
1.2 Configuration Task Summary 9
1.3 Configuring Basic Features 10
1.3.1 Overview 10
1.3.2 Restrictions and Guidelines 10
1.3.3 Configuration Tasks 11
1.3.4 Enabling RIP Process 11
1.3.5 Configuring Association between RIP Instances and VRF 11
1.3.6 Configuring Association with the Local Network 12
1.3.7 Configuring RIP Version 12
1.3.8 Configuring Split Horizon 13
1.3.9 Configuring a Passive Interface 14
1.3.10 Enabling Triggered Updates 14
1.4 Enabling RIP Route Summarization 15
1.4.1 Overview 15
1.4.2 Restrictions and Guidelines 16
1.4.3 Configuration Tasks 16

1.4.4 Configuring to Enable Automatic Route Summarization Function 16
1.4.5 Configuring RIP Route Summarization on an Interface 17
1.5 Configuring RIP Route Filtering Rules 17
1.5.1 Overview 17
1.5.2 Restrictions and Guidelines 17
1.5.3 Configuration Tasks 18
1.5.4 Configuring to Filter the Received RIP Routing Information 18
1.5.5 Configuring to Filter the Sent RIP Routing Information 19
1.6 Configuring to Adjust RIP Routing 19
1.6.1 Overview 19
1.6.2 Restrictions and Guidelines 19
1.6.3 Configuration Tasks 20
1.6.4 Configuring to Modify the Management Distance of a RIP Route 20
1.6.5 Configuring to Increase the Metric of a Received or Sent RIP Route 20
1.6.6 Configuring the Default Metric of an External Route Redistributed to RIP 21
1.7 Modifying Timers 21
1.7.1 Overview 21
1.7.2 Restrictions and Guidelines 22
1.7.3 Configuration Tasks 22
1.7.4 Configuring to Modify the Update Timer, Invalid Timer, Flush Timer, and Holddown Timer of a RIP Route 22
1.7.5 Configuring the Sending Delay Between RIP Route Update Packets 23
1.8 Configuring RIP Route Distribution 23

1.8.1 Overview 23
1.8.2 Restrictions and Guidelines 23
1.8.3 Configuration Tasks 23
1.8.4 Configuring to Advertise a Default Route to Neighbors 24
1.8.5 Configuring to Advertise a Default Route to Neighbors on an Interface 24
1.8.6 Configuring Route Redistribution 25
1.9 Controlling Interaction of RIP Packets 26
1.9.1 Overview 26
1.9.2 Restrictions and Guidelines 27
1.9.3 Configuration Tasks 27
1.9.4 Configuring to Send Unicast RIP Route Update Packets 27
1.9.5 Configuring to Send Broadcast RIPv2 Packets on an Interface 28
1.9.6 Configuring to Allow an Interface to Receive RIP Packets 29
1.9.7 Configuring to Allow an Interface to Send RIP Packets 29
1.9.8 Configuring to Allow an Interface to Send RIP Packets of a Specified Version 30
1.9.9 Configuring to Allow an Interface to Receive RIP Packets of a Specified Version 31
1.10 Configuring to Enable RIPv2 Authentication 32
1.10.1 Overview 32
1.10.2 Restrictions and Guidelines 33
1.10.3 Configuration Tasks 33
1.10.4 Configuring RIP Authentication Mode 33
1.10.5 Configuring a Key Chain for Authentication 34
1.10.6 Configuring a Key Chain for RIP Plain Text Authentication 35

1.11 Configuring BFD Correlation with RIP 36
1.11.1 Overview 36
1.11.2 Restrictions and Guidelines 36
1.11.3 Configuration Tasks 36
1.11.4 Configuring BFD Correlation with RIP Globally 36
1.11.5 Configuring BFD Correlation with RIP on an Interface 37
1.12 Configuring to Enable RIP Fast Reroute 38
1.12.1 Overview 38
1.12.2 Restrictions and Guidelines 38
1.12.3 Procedure 38
1.13 Enabling GR 39
1.13.1 Overview 39
1.13.2 Restrictions and Guidelines 39
1.13.3 Procedure 39
1.14 Configuring to Enable RIP Supernetting Routes 40
1.14.1 Overview 40
1.14.2 Restrictions and Guidelines 40
1.14.3 Procedure 40
1.15 Configuring MIB Binding 41
1.15.1 Overview 41
1.15.2 Restrictions and Guidelines 41
1.16 Configuring a Super VLAN to Enable RIP 41
1.16.1 Overview 41

1.16.2 Restrictions and Guidelines 42
1.16.3 Procedure 42
1.17 Enabling Source Address Verification 42
1.17.1 Overview 42
1.17.2 Restrictions and Guidelines 42
1.17.3 Procedure 43
1.18 Monitoring 43
1.19 Configuration Examples 44
1.19.1 Building a RIP Routing Domain 44
1.19.2 Configuring RIP to Prevent Route Loop 49
1.19.3 Configuring External Routes Imported by RIP 51
1.19.4 Configuring Increasing the Metric of a RIP Interface 53
1.19.5 Configuring RIP to Distribute a Summarized Route 54

1 Configuring RIP

1.1 Introduction

1.1.1 Overview

Routing Information Protocol (RIP) is a unicast routing protocol applied to IPv4 networks. RIP-enabled devices exchange routing information to automatically obtain routes to remote networks.

As an Interior Gateway Protocol (IGP), RIP can run only within an autonomous system (AS). RIP is a distance vector routing protocol, which can control packet interaction through UDP (Port 520) and applies to small-scale networks. Due to simplicity, ease of use, and low workload of configuration and maintenance, the RIP routing protocol is widely used for small networks.

1.1.2 Principles

1. Basic Concepts

- IGP and EGP

  IGP runs within an AS. For example, RIP is a type of IGP.

  Exterior Gateway Protocol (EGP) runs between ASs. For example, BGP is a type of EGP.

- Classful routing protocol and classless routing protocol

  Protocols can be classified based on the type of routes supported:

  Classful routing protocol: It supports classful routes. For example, RIPv1 is a classful routing protocol.

  Classless routing protocol: It supports classless routes. For example, RIPv2 is a classless routing protocol.

- RIP timer
  - Update timer: Route update packets are sent regularly, and the update interval is 30s by default.
  - Invalid timer: If an update packet for refreshing the existing route has not been received after 180s (default value), the metric for the route to set to 16, marking the route as invalid.
  - Flush timer: By default, the flush timer is set to 240s, 60s greater than the invalid timer. When the flush timer times out, the route is deleted from the routing table.
  - Hold-down timer: It prevents route loop caused by route flapping. When the device receives a route update with a metric value of 16, the route will be set to hold-down state, and the hold-down timer will be started at the same time. Before the hold-down timer expires, even if an update packet with a metric less than 16 for this route is received, the route will not be updated.

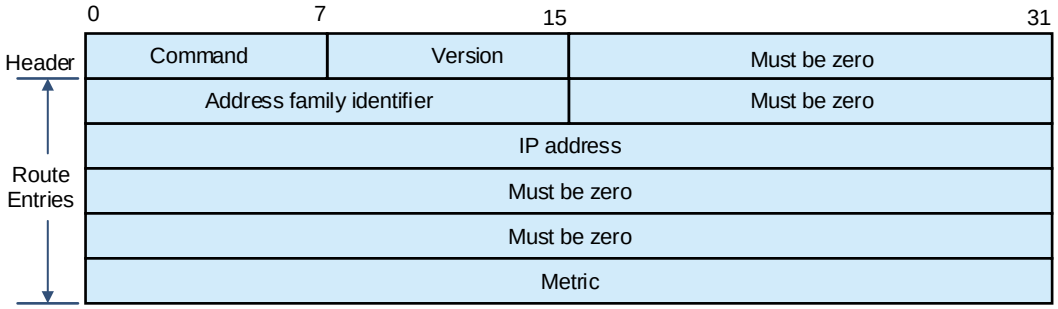
2. RIPv1 and RIPv2

RIP is available in two versions: RIPv1 and RIPv2.

- RIPv1

  RIPv1 packets are broadcast. The broadcast address is 255.255.255.255, and the UDP port ID is 520. No subnet mask is carried in RIPv1 packet, so RIPv1 cannot identify the subnet mask, and supports only classful routes. Meanwhile, RIPv1 does not support routing aggregation and discontinuous subnet.

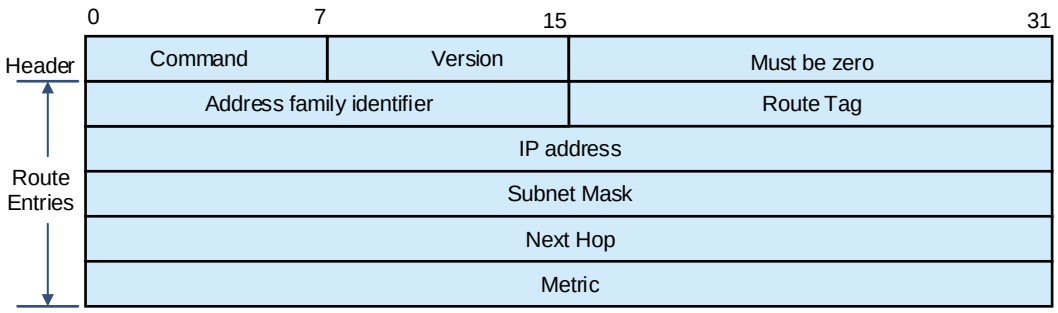
Figure 1-1 RIPv1 Packet Format



- RIPv2

  RIPv2 packets are multicast. The multicast address is 224.0.0.9, and the UDP port ID is 520. RIPv2 can identify the subnet mask, and supports classless routes, summarized route, and supernetting routes. RIPv2 supports plain text authentication and message digest 5 (MD5) authentication. During route redistribution, tags can be added to external routes, and the policy route can control RIP routes according to tags.

Figure 1-2 RIPv2 Packet Format



3. Route Control

Compared with static routing, the dynamic routing protocol shows an advantage, that is, by exchanging routing information, devices can automatically obtain routes to a remote network and update the routes in real time.

- Initialization

  After RIP is enabled on a device, the device sends a request packet to its neighbor device, asking for all routing information, that is, the routing table. After receiving the request packet, the neighbor device returns a response packet containing the local routing table. After receiving the response packet, the device updates the local routing table, and sends an update packet to the neighbor device, informing the neighbor device of the route

update information. After receiving the update packet, the neighbor device updates the local routing table, and sends the update packet to other adjacent devices. After receiving the update advertisement, all devices can obtain and retain the latest routing information.

- Periodical update

  By default, periodical update is enabled for RIP. Adjacent devices exchange complete routing information with each other every 30s, that is, the entire routing table is sent to neighbor devices. One update packet contains at most 25 routes. Therefore, a lot of update packets may be required to send the entire routing table. You can set the sending delay between update packets to avoid loss of routing information.

Note

For every non-local route, if the route is not updated within 180s (Invalid timer), the metric of the route is changed to 16 (unreachable). If the route is still not updated in the next 120s (Flush timer), the route is deleted from the routing table.

- Triggered update

  After the triggered update function is enabled, periodical update is automatically disabled. When routing information changes on a device, the device immediately sends routes related to the change (instead of the complete routing table) to the neighbor device, and use the acknowledgment and retransmission mechanisms to ensure that the neighbor device receives the routes successfully. Compared with periodical update, triggered update helps reduce flooding and accelerates route convergence.

  Events that can trigger update include router startup, interface status change, changes in routing information (such as the metric), and reception of a request packet.

- Route summarization

  When sending routing information to a neighbor device, the RIP-enabled device summarizes subnet routes that belong to the same classful network into a route, and sends the route to the neighbor device. For example, summarize 80.1.1.0/24 (metric = 2) and 80.1.2.0/24 (metric = 3) into 80.0.0.0/8 (metric = 2).

  You can manually specify the aggregation address to summarize routes into a classless route. For example, summarize 80.1.1.0/24 (metric = 2) and 80.1.2.0/24 (metric = 3) into 80.1.0.0/16 (metric = 2).

Note

Only RIPv2 supports route summarization. Route summarization can reduce the size of the routing table and improve the efficiency of routing information exchange. The best metric is selected for the summarized route.

- Supernetting route

  If the subnet mask length of a route is smaller than the natural mask length, this route is called supernetting route. For example, 80.0.0.0/6. As 80.0.0.0 is a Class A network address and the natural mask is 8 bits, 80.0.0.0/6 route is a supernetting route.

Note

Only RIPv2 supports supernetting routes.

- Default route

  In the routing table, a route to the destination network 0.0.0.0/0 is called default route.

  The default route can be learned from a neighbor device, or sent to a neighbor device.

- Route redistribution

  For RIP, other types of routes (such as direct routes, static routes, and routes of other routing protocols) are called external routes.

  External routes (excluding the default route) can be redistributed to RIP and advertised to neighbors.

- Route filtering

  Filtering conditions can be configured to limit the routing information exchanged between adjacent routers. Only the routing information that meets filtering conditions can be sent or received.

4. Routing Algorithm

RIP is a protocol based on the distance-vector algorithm. It uses the vector addition method to compute the routing information.

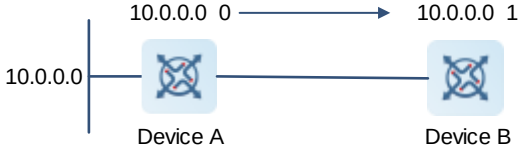
- Distance-vector algorithm

  RIP is a protocol based on the distance-vector algorithm. The distance-vector algorithm treats a route as a vector that consists of the destination network and distance. The device obtains a route from its neighbor and adds the distance vector from itself to the neighbor to the route to form its own route.

  RIP uses the hop count to evaluate the distance to the destination network. By default, the hop count from a device to its directly connected network is 0, the hop count from a device to a network that can be reached through a device is 1, and so on. That is, the metric is equal to the number of devices from the local network to the destination network. To restrict the convergence time, RIP stipulates that the metric must be an integer in the range from 0 to 15. If the metric is equal to or greater than 16, the destination network or host is unreachable. For this reason, RIP cannot be applied to a large-scale network.

  As shown in Figure 1-3, Device A is connected to the network 10.0.0.0. Device B obtains the route (10.0.0.0,0) from Device A and adds the metric 1 to the route to obtain its own route (10.0.0.0,1), and the next hop points to Device A.

Figure 1-3 Diagram of RIP Distance-Vector Algorithm

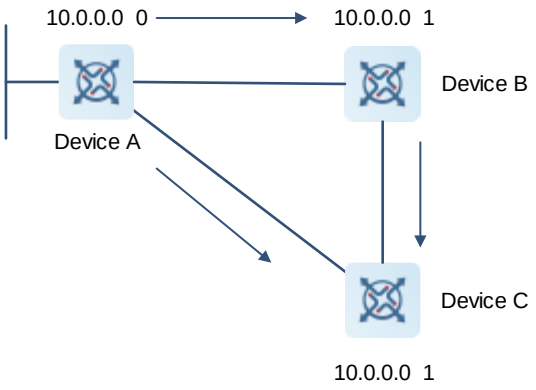


- Selecting the optimum route

RIP selects an optimum route based on the following principle: If multiple routes to the same destination network is available, a device preferentially selects the route with the smallest metric.

As shown in Figure 1-4, Device A is connected to the network 10.0.0.0. Device C obtains the route (10.0.0.0,0) from Device A and the route (10.0.0.0,1) from Device B. Device C will select the route that is obtained from Device A and add metric 1 to this route to form its own route (10.0.0.0,1), and the next hop points to Device A.

Figure 1-4 Diagram of Selecting the Optimum Route



Note

When routes coming from different sources exist on a device, the route with the smaller distance is preferentially selected.

Table 1-1 Default Distance of Routing Protocol

Route Source	Default Distance
Directly-connected network	0
Static route	1
OSPF route	110
IS-IS route	115
RIP route	120
Unreachable route	255

5. Avoiding Route Loops

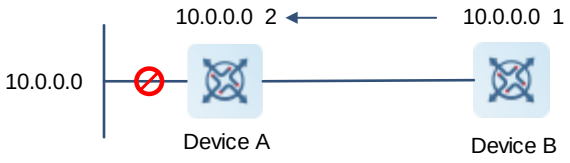
RIP uses functions such as split horizon and poison reverse to avoid route loops.

- Route loop

  A RIP route loop occurs due to inherent defects of the distance-vector algorithm.

  As shown in Figure 1-5, Device A is connected to the network 10.0.0.0, and sends an update packet every 30s. Device B receives the route 10.0.0.0 from Device A every 30s. If Device A is disconnected from 10.0.0.0, the route to 10.0.0.0 will be deleted from the routing table on Device A. Next time, the update packet sent by Device A no longer contains this route. As Device B does not receive an update packet related to 10.0.0.0, Device B determines that the route to 10.0.0.0 is valid within 180s and uses the update packet to send this route to Device A. As the route to 10.0.0.0 does not exist on Device A, the route learned from Device B is added to the routing table. Device B determines that data can reach 10.0.0.0 through Device A, and Device A determines that data can reach 10.0.0.0 through Device B. In this way, a route loop is formed.

Figure 1-5 Diagram of Route Loop



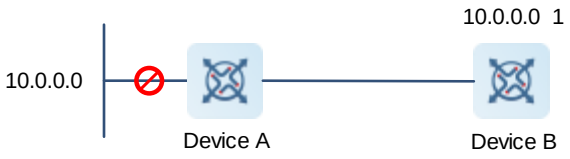
- Split horizon

  Split horizon can prevent route loops. After split horizon is enabled on an interface, a route received on this interface will not be sent out from this interface.

  As shown in Figure 1-6, after split horizon is enabled on the interface between Device A and Device B, Device B will not send the route 10.0.0.0 back to Device A. Device B will learn 180s later that 10.0.0.0 is not reachable.

  On a non broadcast multiple access (NBMA) network, there are multiple neighbors on an interface, and the device communicates with neighbors in unicast form. The routes received by the interface are distinguished by neighbors, and a route learned from a neighbor will not be sent back to this neighbor.

Figure 1-6 Diagram of Split Horizon



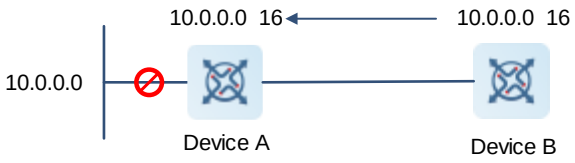
- Poison reverse

  Poison reverse can also prevent route loops. Compared with split horizon, poison reverse is more reliable, but brings more protocol packets, which makes network congestion more severe.

  After poison reverse is enabled on an interface, a route received from this interface will be sent out from this interface again, but the metric of this router will be changed to 16 (unreachable).

As shown in Figure 1-7, after learning the route 10.0.0.0 from Device A, Device B sets the metric of this route to 16 and sends the route back to Device A. After this route becomes invalid, Device B advertises the route 10.0.0.0 (metric = 16) to Device A to accelerate the process of deleting the route from the routing table.

Figure 1-7 Diagram of Poison Reverse



6. Security Measures

RIP uses functions such as authentication and source address verification to ensure protocol security.

- Authentication

  RIPv2 supports authentication, but RIPv1 does not.

  The authentication function is used to prevent unauthorized devices from accessing the RIP routing domain. After authentication is enabled on an interface, the routing information cannot be exchanged between adjacent devices if authentication fails. Authentication is classified into plain text authentication and MD5 authentication:

  Plain text authentication: The configured password is directly added to the packet in the form of plain text, so the security of this authentication mode is low.

  MD5 authentication: The configured password is added to the packet after calculated by MD5 algorithm, so the security of this authentication mode is high.

- Source address verification

  When a RIP-enabled device receives an update packet, it checks whether the source IP address in the packet and the IP address of the inbound interface are in the same network segment. If not, the device drops the packet. Source address verification is used to ensure that RIP routing information is exchanged only between adjacent routing devices.

Note

- On an unnumbered IP interface, source address verification is not performed (not configurable).
- If the triggered updates function is enabled, source address verification is automatically enabled (not configurable).
- If split horizon is disabled, source address verification is automatically enabled (not configurable).

7. Reliability Measures

Use RIP.

Correlate with Bidirectional Forwarding Detection (BFD).

Use fast reroute.

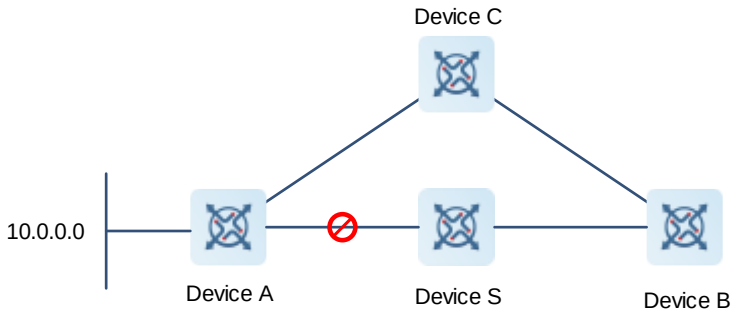
Use graceful restart (GR) and other functions to enhance reliability of the protocol.

- Correlate RTP with Bidirectional Forwarding Detection (BFD).
- Use fast reroute.

  When a link or a device is faulty on the network, packets transmitted through this route will be lost until the route is converged again.

  As shown in Figure 1-8, after the link between Device A and Device S is faulty, Device B may wait for 180s before it can detect the failure of the route (destination network: 10.0.0.0; next hop: Device A). Later, Device B may need to wait for 30s to re-obtain the route (destination network: 10.0.0.0; next hop: Device C) from Device C. Therefore, the traffic is interrupted for 210s.

Figure 1-8 Fast Reroute



  Quick detection of a route failure or fast switchover to the standby route helps shorten the traffic interruption time.

  - A BFD session can be set up between Device A and Device B, and correlated with RIP. BFD can quickly test the connectivity between adjacent devices. Once a link is faulty, RIP can detect the route failure within 1s.
  - The fast reroute function can be enabled. A backup route (destination network: 10.0.0.0; next hop: Device C) can be configured on Device B in advance. Once RIP detects a route failure, the backup route is immediately enabled.

- GR

  GR ensures uninterrupted data transmission when the protocol is restarted. If RIP is restarted on a GR-enabled device, the forwarding table before restart will be retained and a request packet will be sent to the neighbor so that the route can be learned again. During the GR period, RIP completes re-convergence of the route. After the GR period expires, RIP updates the forwarding entry and advertises the routing table to the neighbor.

8. Multiple VPN Instances

Multiple VPN instances may exist on a device.

RIP supports multiple instances. You can enable the RIP process in VPN routing and forwarding (VRF) address family mode to run RIP on VPN instances. One VRF address family is mapped to one VPN instance.

VPN instances cannot be distinguished from each other when you perform RIP operations using SNMP. You must bind the management information base (MIB) of RIP with a VPN instance before the SNMP operations take effect on the VPN instance.

1.1.3 Protocols and Standards

- RFC 1058: Routing Information Protocol
- RFC 2453: RIP Version 2

1.2 Configuration Task Summary

The RIP configuration includes the following tasks:

(1) Configuring Basic Features

a Enabling RIP Process

b (Optional) Configuring Association between RIP Instances and VRF

c Configuring Association with the Local Network

d (Optional) Configuring RIP Version

e (Optional) Configuring Split Horizon

f (Optional) Configuring a Passive Interface

g (Optional) Enabling Triggered Updates

(2) (Optional) Enabling RIP Route Summarization

a Configuring to Enable Automatic Route Summarization Function

b Configuring RIP Route Summarization on an Interface

(3) (Optional) Configuring RIP Route Filtering Rules The following configuration tasks are optional. Select tasks for configuration according to actual condition.

a Configuring to Filter the Received RIP Routing Information

b Configuring to Filter the Sent RIP Routing Information

(4) (Optional) Configuring to Adjust RIP Routing The following configuration tasks are optional. Select tasks for configuration according to actual condition.

a Configuring to Modify the Management Distance of a RIP Route

b Configuring to Increase the Metric of a Received or Sent RIP Route Metric

c Configuring the Default Metric of an External Route Redistributed to RIP

(5) (Optional) Modifying Timers The following configuration tasks are optional. Select tasks for configuration according to actual condition.

a Configuring to Modify the Update Timer, Invalid Timer, Flush Timer, and Holddown Timer of a RIP Route

b Configuring the Sending Delay Between RIP Route Update Packets

(6) (Optional) Configuring RIP Route Distribution The following configuration tasks are optional. Select tasks for configuration according to actual condition.

a Configuring to Advertise a Default Route to Neighbors

b Configuring to Advertise a Default Route to Neighbors on an Interface

c Configuring Route Redistribution

(7) (Optional) Controlling Interaction of RIP Packets The following configuration tasks are optional. Select tasks for configuration according to actual condition.

a Configuring to Send Unicast RIP Route Update Packets

b Configuring to Send Broadcast RIPv2 Packets on an Interface

c Configuring to Allow an Interface to Receive RIP Packets

d Configuring to Allow an Interface to Send RIP Packets

e Configuring to Allow an Interface to Send RIP Packets of a Specified Version

f Configuring to Allow an Interface to Receive RIP Packets of a Specified Version

(8) (Optional) Configuring to Enable RIPv2 Authentication

a Configuring RIP Authentication Mode

b Configuring a Key Chain for Authentication

c Configuring a Key Chain for RIP Plain Text Authentication

(9) (Optional) Configuring BFD Correlation with RIP

a Configuring BFD Correlation with RIP Globally

b Configuring BFD Correlation with RIP on an Interface

(10) (Optional) Configuring to Enable RIP Fast Reroute

(11) (Optional) Enabling GR

(12) (Optional) Configuring to Enable RIP Supernetting Routes

(13) (Optional) Configuring MIB Binding

(14) (Optional) Configuring a Super VLAN to Enable RIP

(15) (Optional) Enabling Source Address Verification

1.3 Configuring Basic Features

1.3.1 Overview

Build a RIP routing domain on the network. Devices in the domain obtain routes to a remote network through RIP.

1.3.2 Restrictions and Guidelines

- IPv4 addresses must be configured.
- IPv4 unicast routes must be enabled.

1.3.3 Configuration Tasks

The basic function configuration of RIP includes the following tasks:

(1) Enabling RIP Process

(2) (Optional) Configuring Association between RIP Instances and VRF

(3) Configuring Association with the Local Network

(4) (Optional) Configuring RIP Version

(5) (Optional) Configuring Split Horizon

(6) (Optional) Configuring a Passive Interface

(7) (Optional) Enabling Triggered Updates

1.3.4 Enabling RIP Process

1. Overview

This command is used to create a RIP routing process and enter the RIP routing process configuration mode.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable a RIP routing process and enter the RIP routing process configuration mode.

router-protocol rip

The RIP routing process is disabled by default.

1.3.5 Configuring Association between RIP Instances and VRF

1. Overview

This command is used to create RIP instances to be associated with VRF, and enter the IPv4 VRF address family configuration mode.

2. Restrictions and Guidelines

This function must be configured if you need to configure multiple RIP instances and associate them with VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) Create RIP instances to be associated with VRF, and enter the RIP IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

By default, the address family of RIP protocol is disabled.

(5) Exit the RIP IPv4 VRF address family configuration mode.

quit-address-family-mode

1.3.6 Configuring Association with the Local Network

1. Overview

To add an interface to the RIP domain, you need to first advertise the interface network in the RIP process.

2. Restrictions and Guidelines

- RIP can run and learn direct routes and RIP packets can be exchanged only on an interface covered by the **network-advertise**.
- If **network-advertise** 0.0.0.0 255.255.255.255 is configured, all interfaces are covered.
- If the *wildcard* parameter is not configured, the classful address range is used by default, that is, the interfaces whose addresses fall into the classful address range participate in RIP operations.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure association with the local network.

network-advertise *network-number* [*wildcard*]

1.3.7 Configuring RIP Version

1. Overview

You can run this command to define the versions of RIP packets sent or received on all interfaces. This command takes effect on the entire device.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure a RIP version.

version-num { **1** | **2** }

By default, route update packets of RIPv1 and RIPv2 can be received, but only route update packets of RIPv1 are sent.

1.3.8 Configuring Split Horizon

1. Overview

This function prevents route loop. After split horizon is enabled on an interface, a route received on this interface will not be sent out from this interface.

2. Restrictions and Guidelines

After poison reverse is enabled, split horizon is automatically disabled.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  interface aggregateport *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  interface *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*
- Enter the VXLAN routed interface configuration mode.
  interface overlayrouter *interface-number*

(4) Enable split horizon.

ip rip horizon-split-enable [**poisoned-reverse**]

By default, the split horizon function is enabled, but poison reverse is disabled.

1.3.9 Configuring a Passive Interface

1. Overview

The interface set to a passive interface suppresses RIP update packets. A passive interface defines the boundary of RIP routing domain to avoid unwanted flooding of RIP packets. If the interface connection device does not run the RIP routing protocol (such as a PC and a device running other routing protocols), you are advised to configure this interface as a passive interface.

2. Restrictions and Guidelines

You are advised to set all interfaces to passive interfaces in RIP route configuration mode, and then enter the device interconnection interface in the specific domain to disable the passive interface function.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) Configure passive interfaces.

disable-interface { **default** | *interface-type interface-number* }

The passive interface function is disabled by default.

1.3.10 Enabling Triggered Updates

1. Overview

Enable the RIP triggered updates function, after which RIP does not periodically send route update packets. This function can quickly obtain network changes, avoid route loops, and improve the convergence speed.

2. Restrictions and Guidelines

- The basic functions of RIP must be configured.
- It is recommended that split horizon with poison reverse be enabled; otherwise, invalid routing information may exist.
- This function cannot be enabled together with the function of correlating RIP with BFD.
- Ensure that the triggered updates function is enabled on every router on the same link; otherwise, the routing information cannot be exchanged properly.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  interface aggregateport *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  interface *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  interface aggregateport *interface-number.subnumber*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*
- Enter the VXLAN routed interface configuration mode.
  interface overlayrouter *interface-number*

(4) Enable triggered updates.

ip rip triggered-ext [**retransmit-timer** *timer* | **retransmit-count** *count*]

The RIP triggered updates function is disabled by default.

1.4 Enabling RIP Route Summarization

1.4.1 Overview

This function can reduce the size of the routing table, improve the routing efficiency, avoid route flapping to some extent, and improve scalability and effectiveness of the network.

1.4.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- The range of supernetting routes is larger than that of the classful network. Therefore, the automatic route summarization function is invalid for supernetting routes.
- RIPv1 performs automatic route summarization by default. If the detailed routes should be advertised, you must set the RIP version to RIPv2.

1.4.3 Configuration Tasks

The configuration for enabling RIP route summarization includes the following tasks:

(1) Configuring to Enable Automatic Route Summarization Function

(2) Configuring RIP Route Summarization on an Interface

1.4.4 Configuring to Enable Automatic Route Summarization Function

1. Overview

This function can be configured to reduce the size of the routing table, improve the routing efficiency, avoid route flapping to some extent, and improve scalability and effectiveness of the network.

2. Restrictions and Guidelines

- Route summarization is enabled by default for RIPv1 and RIPv2.
- You can disable automatic route summarization only when RIPv2 is configured.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Enable the automatic route summarization function.

auto-aggregation

The automatic summarization function of RIP routes is enabled by default.

1.4.5 Configuring RIP Route Summarization on an Interface

1. Overview

This function is used to manually summarize an address or a subnet on a specified interface.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  interface aggregateport *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  interface *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  interface aggregateport *interface-number.subnumber*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*
- Enter the VXLAN routed interface configuration mode.
  interface overlayrouter *interface-number*

(4) Configure RIP route summarization on an interface.

ip rip summary-address *ipv4-address mask*

By default, RIP route summarization is not configured on an interface.

1.5 Configuring RIP Route Filtering Rules

1.5.1 Overview

Routes that do not meet filtering criteria cannot be loaded to the routing table, or advertised to neighbors. In this way, users within the network can be prevented from accessing specified destination networks.

1.5.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.

- In regard to the filtering rules of sent routes, you must configure route redistribution first, and then filter the redistributed routes.

1.5.3 Configuration Tasks

The configuration of RIP route filtering rules includes the following tasks, and these tasks are optional. Select tasks for configuration according to actual condition.

- Configuring to Filter the Received RIP Routing Information
- Configuring to Filter the Sent RIP Routing Information

1.5.4 Configuring to Filter the Received RIP Routing Information

1. Overview

You can configure the route distribution control list to refuse to receive some specified routes. If no interface is specified, the route update packets received on all interfaces will be filtered.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ACL or prefix list. Please configure only one task.

- Configure an ACL.

  access-control-list *acl-number* { **deny** | **permit** } { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } [**time-range** *tm-range-name*] [**log**]

- Configure a prefix list.

  ip prefix-list-info *prefix-list-name* [**seq** *seq-number*] { **deny** | **permit** } *ipv4-prefix* [**ge** *minimum-prefix-length*] [**le** *maximum-prefix-length*]

(4) Enter the RIP routing configuration mode.

router-protocol rip

(5) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(6) Filter the received RIP routing information.

route-filtering { *acl-name* | *acl-number* | **gateway-filter** *prefix-list-name* | **prefix-filter** *prefix-list-name* [**gateway-filter** *prefix-list-name*] } **in** [*interface-type interface-number*]

There is no inbound distribution list by default.

1.5.5 Configuring to Filter the Sent RIP Routing Information

1. Overview

You can configure the route distribution control list to prohibit distributing some specified routes. If no interface is specified, the route update packets sent to all interfaces will be filtered.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ACL or prefix list. Please configure only one task.

- Configure an ACL.

  access-control-list *acl-number* { **deny** | **permit** } { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } [**time-range** *tm-range-name*] [**log**]

- Configure a prefix list.

  ip prefix-list-info *prefix-list-name* [**seq** *seq-number*] { **deny** | **permit** } *ipv4-prefix* [**ge** *minimum-prefix-length*] [**le** *maximum-prefix-length*]

(4) Enter the RIP routing configuration mode.

router-protocol rip

(5) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(6) Filter the sent RIP routing information.

route-filtering { [*acl-number* | *acl-name*] | **prefix-filter** *prefix-list-name* } **out** [*interface-type interface-number* **| bgp** | **connected** | **isis** [*area-tag*] **| ospf** process-id **|rip-instance | static**]]

There is no outbound distribution list by default.

1.6 Configuring to Adjust RIP Routing

1.6.1 Overview

This function can be used to change a RIP route so that the traffic passes through specified nodes or bypasses specified nodes. Change the sequence that a device selects various types of routes so as to change the priorities of RIP routes.

1.6.2 Restrictions and Guidelines

The basic functions of RIP must be configured.

1.6.3 Configuration Tasks

The configuration of RIP routing adjustment includes the following tasks, and these tasks are optional. Select tasks for configuration according to actual condition.

- Configuring to Modify the Management Distance of a RIP Route
- Configuring to Increase the Metric of a Received or Sent RIP Route Metric
- Configuring the Default Metric of an External Route Redistributed to RIP

1.6.4 Configuring to Modify the Management Distance of a RIP Route

1. Overview

This function is used to set the management distance of RIP routes and change the priority of a device in routing.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure to modify the management distance of a RIP route.

distance *distance* [*ipv4-address wildcard*]

The management distance is **120** by default.

1.6.5 Configuring to Increase the Metric of a Received or Sent RIP Route

1. Overview

You can use this function to increase the metric of a received or sent RIP route. If an interface is specified, the configuration takes effect only on the specified interface; otherwise, the configuration takes effect globally.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure to increase the metric of a received or sent RIP route.

metric-offset-list { *acl-name* | *acl-number* } { **in** | **out** } *offset* [*interface-type interface-number*]

The metric of a received or sent RIP route is not increased by default.

1.6.6 Configuring the Default Metric of an External Route Redistributed to RIP

1. Overview

The default metric of a redistributed route is 1. If you want to preferably select an internal RIP route on some networks, just increase the default metric.

2. Restrictions and Guidelines

This function needs to be used together with the routing protocol redistribution function.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure the default metric of an external route redistributed to RIP.

default-metric *metric-value*

By default, the metric of a redistributed route is 1.

1.7 Modifying Timers

1.7.1 Overview

This function can be used to change the durations of RIP timers to accelerate or slow down change to the protocol state or occurrence of an event.

1.7.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- Modifying the protocol control parameters may result in protocol running failures. Therefore, you are advised not to modify the timers.

1.7.3 Configuration Tasks

The configuration of RIP timer modification includes the following tasks, and these tasks are optional. Select tasks for configuration according to actual condition.

- Configuring to Modify the Update Timer, Invalid Timer, Flush Timer, and Holddown Timer of a RIP Route
- Configuring the Sending Delay Between RIP Route Update Packets

1.7.4 Configuring to Modify the Update Timer, Invalid Timer, Flush Timer, and Holddown Timer of a RIP Route

1. Overview

By adjusting the timers, you can reduce the convergence time and fault rectification time of the routing protocol.

2. Restrictions and Guidelines

For devices connected to the same network, the values of RIP timers must be consistent. Generally, you are not advised to modify the RIP timers unless otherwise required.

Setting timers to small values on a low-speed link brings risks because a lot of update packets consume the bandwidth. You can set timers to small values generally on the Ethernet or a link of 2 Mbps (or above) to reduce the convergence time of network routes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure to modify the update timer, invalid timer, flush timer, and holddown timer of a RIP route.

timers-set basic *update invalid flush* [**holddown** *holddown*]

By default, the update timer is 30s, the invalid timer is 180s, the flush timer is 120s, and the holddown timer is 0s.

1.7.5 Configuring the Sending Delay Between RIP Route Update Packets

1. Overview

A RIP route update packet is 512 bytes long and can contain 25 routes. If the number of routes to be updated exceeds 25, more than one update packet will be sent as fast as possible. When a high-speed device sends a lot of update packets to a low-speed device, the low-speed device may fail to process all update packets in time, causing loss of routing information. This function can be used to increase the sending delay between packets on a high-speed device so that a low-speed device can receive and process all the update packets.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure the sending delay between RIP route update packets.

updates-send-delay *delay*

The sending delay between update packets is 0 by default.

1.8 Configuring RIP Route Distribution

1.8.1 Overview

To enable the RIP routing domain to communicate with other ASs, you can inject a default route to other ASs into the RIP routing domain, or redistribute the routes of other ASs to RIP.

1.8.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- Route redistribution cannot introduce default routes of other protocols to the RIP routing domain.

1.8.3 Configuration Tasks

The configuration of RIP route distribution includes the following tasks, and these tasks are optional. Select tasks for configuration according to actual condition.

- Configuring to Advertise a Default Route to Neighbors
- Configuring to Advertise a Default Route to Neighbors on an Interface

- Configuring Route Redistribution

1.8.4 Configuring to Advertise a Default Route to Neighbors

1. Overview

RIP will not advertise the default route in the routing table to neighbors by default. If you need to advertise the default route to the neighbors, use this function.

2. Restrictions and Guidelines

You still need to run the **default-info originate** command to introduce the default route generated by **ip default-network** to RIP.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Advertise the default route to neighbors.

default-info originate [**constantly**] [**metric-value** *metric-value*] [**route-map** *route-map-name*]

1.8.5 Configuring to Advertise a Default Route to Neighbors on an Interface

1. Overview

If a default route needs to be advertised to some neighbors only, the default route to be advertised can be configured on the related interfaces, and RIP will advertise it to neighbors on the configured interfaces only.

2. Restrictions and Guidelines

- If you configure the **ip rip default-info** command for the interface, and the **default-info originate** command for the RIP process at the same time, only the default route configured for the interface is advertised.
- So far as the **ip rip default-info** command is configured for an interface, RIP does not learn the default route advertised by the neighbor.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  interface aggregateport *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  interface *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  interface aggregateport *interface-number.subnumber*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*
- Enter the VXLAN routed interface configuration mode.
  interface overlayrouter *interface-number*

(4) Advertise the default route to neighbors on an interface.

ip rip default-info { **only** | **originate** } [**metric-value** *metric-value*]

By default, default route advertisement is disabled on an interface.

1.8.6 Configuring Route Redistribution

1. Overview

The function redistributes the routes of other protocols to the RIP domain to implement internetworking between the RIP domain and other routing domains.

2. Restrictions and Guidelines

- If the **level** parameter is not carried when IS-IS route redistribution is configured, only level-2 routes can be redistributed by default. If the **level** parameter is carried during initial configuration of redistribution, the routes configured with the **level** parameter can be redistributed. If both **level 1** and **level 2** are configured, the two levels are combined and saved as **level-1-2**.
- If you configure redistribution of OSPF routes without specifying the **filter** parameter, OSPF routes of all sub-types can be distributed by default. The latest setting of the **filter** parameter is used as the initial **filter** parameter. Only routes that match the sub-types can be redistributed. You can run the **unset** form of the command to restore the default value of **filter**.
- The configuration rules for the **unset** form of the **redistribute-route** command are as follows:

- If some parameters are specified in the **unset** form of this command, default values of these parameters will be restored.
- If no parameter is specified in the **unset** form of this command, the entire command will be deleted.

- For example, if **redistribute-route isis 112 level-2** is configured, you can run the **unset redistribute-route isis 112 level-2** command to restore the default value of level-2. As **level-2** itself is the default value of the parameter, the configuration saved is still **redistribute-route isis 112 level-2** after the preceding **unset** form of the command is executed. To delete the entire command, run the **unset redistribute-route isis 112** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Redistribute routes and advertise external routes to neighbors.

redistribute-route {**bgp** | **connected** | **isis [** ***area-tag*** **] [level-1 | level-2 | level-2] | ospf** *process-id* [**filter** { **internal-type** | **external-type** [1 | 2] | **nssa-external-type** [1 | 2] } *] | **static** } [**metric-value** *metric-value* **| route-map** *route-map-name*] *

The redistribution function is not configured by default.

1.9 Controlling Interaction of RIP Packets

1.9.1 Overview

This function can be used to change the default running mechanism of RIP through configuration and control the interaction mode of RIP packets.

The following interaction modes of RIP packets can be implemented:

Allowing or prohibiting the sending of unicast RIP packets to a specified neighbor on an interface

Allowing or prohibiting the sending of unicast RIPv2 packets instead of broadcast packets to a specified neighbor on an interface

Allowing or prohibiting the receiving of RIP packets on an interface

Allowing or prohibiting the sending of RIP packets on an interface

Allowing or prohibiting the receiving of RIP packets of a specified version on an interface

Allowing or prohibiting the sending of RIP packets of a specified version on an interface

1.9.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- On an interface connecting to a neighbor device, the configured version of sent RIP packets must be the same as the version of received RIP packets.

1.9.3 Configuration Tasks

The configuration of RIP packet interaction includes the following tasks, and these tasks are optional. Select tasks for configuration according to actual condition.

- Configuring to Send Unicast RIP Route Update Packets
- Configuring to Send Broadcast RIPv2 Packets on an Interface
- Configuring to Allow an Interface to Receive RIP Packets
- Configuring to Allow an Interface to Send RIP Packets
- Configuring to Allow an Interface to Send RIP Packets of a Specified Version
- Configuring to Allow an Interface to Receive RIP Packets of a Specified Version

1.9.4 Configuring to Send Unicast RIP Route Update Packets

1. Overview

This function can be configured if you hope that only some of devices connected to an interface can receive the updated routing information or when RIP is used for a non-broadcast network.

2. Restrictions and Guidelines

Generally, you can first enable the passive interface function in RIP route configuration mode, and then specify the neighbors that can receive the routing information. This command does not affect the receiving of RIP packets. After an interface is configured as a passive interface, the interface does not send a request packet even after the device is restarted.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure to send unicast RIP route update packets to a specified neighbor.

peer *ipv4-address*

No RIP neighbor is configured by default.

1.9.5 Configuring to Send Broadcast RIPv2 Packets on an Interface

1. Overview

This function can be configured if the neighbor device does not support receiving of multicast RIPv2 packets.

2. Restrictions and Guidelines

- The default behavior depends on the configuration of the **version-num** command.
- The configuration result of this command can overwrite the default configuration of the **version-num** command.
- This command affects the behavior of sending RIP packets on the current interface, and the interface is allowed to send RIPv1 and RIPv2 packets simultaneously.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure to send broadcast RIPv2 packets on an interface.

ip rip version2-broadcast

By default, RIPv2 packets are multicast on an interface.

1.9.6 Configuring to Allow an Interface to Receive RIP Packets

1. Overview

If you want to prevent an interface from receiving RIP packets, you can enable this function.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure to allow an interface to receive RIP packets.

ip rip packet-receive enable

By default, an interface can receive RIP packets normally.

1.9.7 Configuring to Allow an Interface to Send RIP Packets

1. Overview

If you want to prevent an interface from sending RIP packets, you can disable this function.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure to allow an interface to send RIP packets.

ip rip packet-send enable

By default, the RIP packet sending function is enabled on an interface.

1.9.8 Configuring to Allow an Interface to Send RIP Packets of a Specified Version

1. Overview

To be compatible with the RIP version running on the peer device, use this function to change the version of RIP packets sent by the interface.

2. Restrictions and Guidelines

- The default behavior depends on the configuration of the **version-num** command.
- The configuration result of this command can overwrite the default configuration of the **version-num** command.
- This command affects the behavior of sending RIP packets on the current interface, and the interface is allowed

to send RIPv1 and RIPv2 packets simultaneously. If this command does not contain any parameter, the behavior of receiving RIP packets depends on the configuration of the **version-num** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure to allow an interface to send RIP packets of a specified version.

ip rip packet-send version-num [**1** | **2**] *

By default, the default behavior depends on the configuration of the **version-num** command.

1.9.9 Configuring to Allow an Interface to Receive RIP Packets of a Specified Version

1. Overview

To be compatible with the RIP version running on the peer device, use this function to change the version of RIP packets that can be received by the interface.

2. Restrictions and Guidelines

- The default behavior depends on the configuration of the **version-num** command.

- The configuration result of this command can overwrite the default configuration of the **version-num** command.
- This command affects the behavior of receiving RIP packets on the current interface, and the interface is allowed to receive RIPv1 and RIPv2 packets simultaneously. If this command does not contain any parameter, the behavior of receiving RIP packets depends on the configuration of the **version-num** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure to allow an interface to receive RIP packets of a specified version.

ip rip packet-receive version-num [**1**] [**2**]

By default, the default behavior depends on the configuration of the **version-num** command.

1.10 Configuring to Enable RIPv2 Authentication

1.10.1 Overview

This function prevents learning unauthenticated and invalid routes and advertising valid routes to unauthenticated devices, ensuring stability of the routing system, and protecting the system against intrusions.

1.10.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- Only RIPv2 supports authentication of RIP packets, and RIPv1 does not.

1.10.3 Configuration Tasks

The configuration for enabling RIPv2 authentication includes the following tasks:

(1) Configuring RIP Authentication Mode

(2) Configuring a Key Chain for Authentication

(3) Configuring a Key Chain for RIP Plain Text Authentication

1.10.4 Configuring RIP Authentication Mode

1. Overview

RIPv2 supports plain text authentication and cipher text authentication, and ensures secure transmission of RIP packets.

2. Restrictions and Guidelines

- For all devices that need to directly exchange the RIP routing information, the RIP authentication mode of these devices must be the same.
- Cipher text authentication is recommended because the plain text password is saved in the configuration file and the security is low.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable the RIP authentication mode.

ip rip enable-auth mode { **text** | **md5** }

Plain text authentication mode is used by default.

1.10.5 Configuring a Key Chain for Authentication

1. Overview

A key chain can be used to change the key dynamically to improve the security.

2. Restrictions and Guidelines

- The specified key chain must be defined using the **key chain** command in global configuration mode in advance.
- If this command and the **ip rip enable-auth mode md5** command are configured, MD5 mode authentication is adopted. If the **ip rip enable-auth mode** command is not configured, authentication is performed according to the authentication mode specified by the configured **key chain**. MD5 authentication and SM3 authentication are supported at present.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure a key chain for authentication.

ip rip enable-auth keychain *name-of-keychain*

No key chain is set by default.

1.10.6 Configuring a Key Chain for RIP Plain Text Authentication

1. Overview

This function uses a string to implement plain text encryption and improve the security of RIP packet transmission.

2. Restrictions and Guidelines

- This function takes effect only in plain text authentication mode.
- Encrypted strings can be displayed in plain text or cipher text. Cipher text is recommended.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable RIP plain text authentication and configure a key chain.

ip rip enable-auth text-password [**0** | **7**] *password-string*

No character string of plain text authentication is set by default.

1.11 Configuring BFD Correlation with RIP

1.11.1 Overview

Once a link fails, RIP can quickly detect the failure of the route. This function helps shorten the traffic interruption time.

1.11.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- The BFD correlation configured in interface configuration mode takes precedence over the global configuration.

1.11.3 Configuration Tasks

The configuration for enabling BFD correlation with RIP includes the following tasks:

- Configuring BFD Correlation with RIP Globally
- Configuring BFD Correlation with RIP on an Interface

1.11.4 Configuring BFD Correlation with RIP Globally

1. Overview

This function enables BFD correlation with RIP globally. Once a link fails, RIP can quickly detect the failure of the route, which helps shorten the traffic interruption time.

After BFD is enabled on RIP, a BFD session will be set up for the RIP routing information source (that is, the source address of RIP route update packets). Once the BFD neighbor fails, the corresponding RIP route directly enters the invalid state and is not forwarded.

2. Restrictions and Guidelines

The BFD configured on an interface takes precedence over that configured in process configuration mode.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure to enable BFD correlation with RIP globally.

bfd whole-interfaces

The BFD function is disabled by default.

1.11.5 Configuring BFD Correlation with RIP on an Interface

1. Overview

This function enables BFD correlation with RIP on an interface. Once an interface link fails, RIP can quickly detect the failure of the route, which helps shorten the traffic interruption time.

2. Restrictions and Guidelines

- By default, the BFD configuration is subject to that configured in RIP process configuration mode.
- This configuration must be performed if you need to enable or disable BFD correlation on a specified interface.
- The BFD configured on an interface takes precedence over that configured in RIP process configuration mode.
- According to the actual environment, you can select to enable BFD on the specified interface, or enable BFD globally in RIP process, or disable BFD on the specified interface.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure BFD correlation with RIP on an interface.

ip rip bfd [**disable**]

By default, BFD link detection on an interface is disabled by default.

1.12 Configuring to Enable RIP Fast Reroute

1.12.1 Overview

Once RIP detects a route failure, the router immediately switches to the second-best route. This configuration helps shorten the traffic interruption time.

1.12.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- The route map and the backup next hop must be configured.
- To accelerate convergence, set **carrier-delay** of the interface to **0** and enable BFD correlation with RIP.

1.12.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a routing policy.

route-map *route-map-name* [{ **permit** | **deny** } *sequence*]

(4) Run this command to configure the interface-based matching rule.

filter interface { *interface-type interface-number* }&<1-n>

No interface-based matching rule is configured by default.

(5) Return to the global configuration mode.

exit

(6) Enter the RIP routing configuration mode.

router-protocol rip

(7) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(8) Enable fast reroute.

frr route-map *route-map-name*

The fast reroute function is disabled by default.

1.13 Enabling GR

1.13.1 Overview

The GR period is the maximum time from restart of the RIP process to completion of the restart. During this period, RIP performs route recovery and the forwarding table remains unchanged. After the restart period expires, RIP exits the GR state and performs common RIP operations.

1.13.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- The GR period is at least twice the RIP route update period.

1.13.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Enable GR.

gr [**grace-period** *grace-period*]

The GR function is enabled by default.

1.14 Configuring to Enable RIP Supernetting Routes

1.14.1 Overview

This function is used to allow RIP to send RIP supernetting routes on a specified interface. The concept of supernetting is similar to that of subnet. Subnet means dividing a large network segment into several small network segments, while supernetting means aggregating some small network segments into a large network segment. The number of network address bits allowed by supernetting is larger than that allowed by a classful network.

1.14.2 Restrictions and Guidelines

The basic functions of RIP must be configured.

1.14.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable supernetting routes.

ip rip packet-send supernet-routes

By default, the supernetting route sending function is enabled on an interface.

1.15 Configuring MIB Binding

1.15.1 Overview

Since RIP MIB does not have RIP instances, you can only operate on the unique instance when operating RIP instances through SNMP. With this function, you can bind the MIB to the RIP instance of the specified VRF, and operate the RIP instance of the specified VRF through the SNMP protocol.

1.15.2 Restrictions and Guidelines

The RIP basic functions (with VRF parameters) must be configured.

1. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Configure binding the MIB of RIP to the specified VRF instance.

enable mib-binding

By default, the MIB is bound to the RIP instance of the default VRF.

1.16 Configuring a Super VLAN to Enable RIP

Note

This function is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.16.1 Overview

This function is used to run RIP on a super VLAN. A super VLAN is also called VLAN aggregation. Multiple sub VLANs are associated with one super VLAN. All sub VLANs share the SVI interface of the super VLAN and use it as the gateway of Layer-3 communication to save IP resources.

In normal cases, a super VLAN contains multiple sub VLANs. Multicast packets of a super VLAN are also sent to its sub VLANs. In this case, when RIP multicast packets are sent over a super VLAN containing multiple sub VLANs, the RIP multicast packets are replicated multiple times, and the device processing capability is insufficient. As a result, a large number of packets are discarded, causing protocol flapping.

In most scenarios, the RIP function does not need to be enabled on a super VLAN, and it is disabled by default. However, in some scenarios, the RIP function must be run on the super VLAN, but packets need to be sent to only one sub VLAN. In this case, you can decide to send multicast packets to a certain sub VLAN or to all sub VLANs as actually needed. You can run this command to specify a particular sub VLAN. You must be cautious when configuring packet transmission to all sub VLANs, as the large number of sub VLANs may cause a device processing bottleneck, which will lead to the neighbor flapping.

1.16.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- The designated sub VLAN must be connected to neighbors.

1.16.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI interface configuration mode.

interface vlan *interface-number*

(4) Enable RIP on a Super VLAN.

ip rip subvlan { **all |** *vlan-id* }

By default, RIP is disabled on a super VLAN.

1.17 Enabling Source Address Verification

1.17.1 Overview

You can run the command to validate the source address of a RIP route update packet. The purpose is to ensure that the RIP routing process receives only the route update packets coming from the same IP subnet neighbor.

1.17.2 Restrictions and Guidelines

- The basic functions of RIP must be configured.
- After split horizon is disabled on an interface, the RIP routing process will perform source address validation on the update packet no matter whether the **verify-update-source** command is executed in routing process configuration mode.
- For an interface not configured with an IP address, the RIP routing process does not perform source address validation for the update packet no matter whether the **verify-update-source** command is executed in routing process configuration mode.

1.17.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the RIP routing configuration mode.

router-protocol rip

(4) (Optional) Enter the IPv4 VRF address family configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(5) Enables source address verification.

verify-update-source

The source address validation function for update packets is enabled by default.

1.18 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-2 RIP Monitoring

Command	Purpose
display ip rip-instance	Displays the basic information about a RIP process.
display ip rip-instance database [**count** | *network-number network-mask* | **vrf-instance** *vrf-name* [*network-number network-mask*]]	Displays the RIP routing table.
display ip rip-instance external-type [**bgp** | **connected** | **isis-instance** [*process-id*] | **ospf-instance** *process-id* | **static**] [**vrf-instance** *vrf-name*]	Displays information about external routes redistributed by RIP.
display ip rip-instance interface [**vrf** *vrf-name*] [*interface-type interface-number*]	Displays the RIP interface information.

Command	Purpose
display ip rip-instance peer [*ipv4-address*] [**vrf** *vrf-name*]	Displays the RIP neighbor information.
debugging ip rip event	Debugs events that occur when the RIP process is running.
debugging ip rip frr	Debugs RIP fast reroute.
debugging ip rip nsm	Debugs interaction with the NSM process.
debugging ip rip packet [**recv** | **packet-send**] [**interface** *interface-type interface-number*] [**deeply**]	Debugs the sent and received packets.
debugging ip rip restart	Debugs the RIP GR process.
debugging ip rip route	Debugs the route changes of the RIP process.

1.19 Configuration Examples

1.19.1 Building a RIP Routing Domain

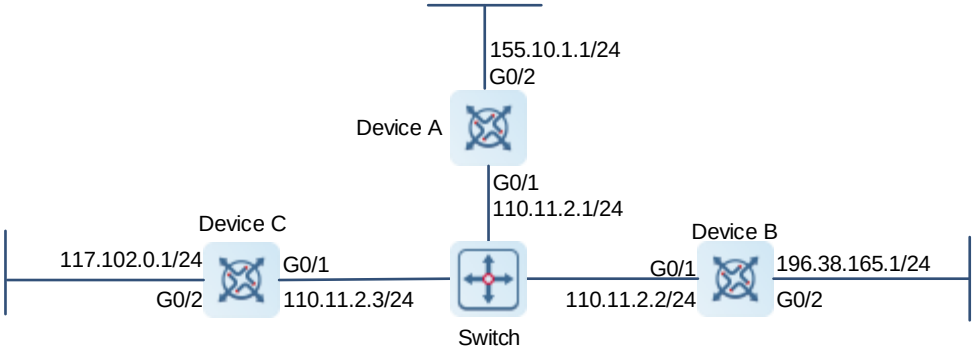
1. Requirements

Device A, Device B, and Device C implement interworking through the RIPv2 routing protocol.

To save the device performance, all interfaces of the three devices are configured as passive interfaces by default, and passive interface is disabled only for the interfaces directly connected to RIP neighbors.

2. Topology

Figure 1-9 Configuring a RIP Routing Domain



3. Notes

- Configure interface IP addresses for all the devices.
- Configure the RIP basic functions for all the devices.

4. Procedure

(1) Configure an IP address for each interface and start the unicast routing protocol.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ip address 110.11.2.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/2)# ip address 155.10.1.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/2)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ip address 110.11.2.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/2)# ip address 196.38.165.1 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/2)# exit
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/1)# ip address 110.11.2.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface gigabitethernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/2)# ip address 117.102.0.1 255.255.0.0
Device C(cfg-if-GigabitEthernet 0/2)# exit
```

(2) Configure the RIP basic functions for the devices. Except the RIP neighbor interface, configure the other interfaces as passive interfaces.

Configure Device A.

```
Device A(cfg)# router-protocol rip
Device A(cfg-router)# version 2
Device A(cfg-router)# network-advertise 0.0.0.0 255.255.255.255
Device A(cfg-router)# disable-interface reset
Device A(cfg-router)# unset disable-interface gigabitethernet 0/1
```

Configure Device B.

```
Device B(cfg)# router-protocol rip
Device B(cfg-router)# version 2
Device B(cfg-router)# network-advertise 0.0.0.0 255.255.255.255
Device B(cfg-router)# disable-interface reset
Device B(cfg-router)# unset disable-interface gigabitethernet 0/1
```

Configure Device C.

```
Device C(cfg)# router-protocol rip
Device C(cfg-router)# version 2
Device C(cfg-router)# unset auto-aggregation
Device C(cfg-router)# network-advertise 0.0.0.0 255.255.255.255
Device C(cfg-router)# disable-interface reset
Device C(cfg-router)# unset disable-interface gigabitethernet 0/1
```

5. Verification

Check the IP routing tables on Device A, Device B, and Device C. Verify that RIP has learned the routes to remote networks.

- Device A

```
Device A# display ip route-config

Codes:   C - connected, S - static, R - RIP, B - BGP
         O - OSPF, IA - OSPF inter area
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         ia - IS-IS inter area, * - candidate default

Gateway of last resort is no set
C       110.11.2.0/24 is directly connected, GigabitEthernet 0/1
C       110.11.2.1/32 is local host.
R       117.0.0.0/8 [120/1] via 110.11.2.2, 00:00:47, GigabitEthernet 0/1
C       155.10.1.0/24 is directly connected, GigabitEthernet 0/2
```

```
C       155.10.1.1/32 is local host.
C       192.168.217.0/24 is directly connected, VLAN 1
C       192.168.217.233/32 is local host.
R       196.38.165.0/24 [120/1] via 110.11.2.3, 00:19:18, GigabitEthernet 0/1
```

- Device B

```
Device B# display ip route-config

Codes:   C - connected, S - static, R - RIP, B - BGP
         O - OSPF, IA - OSPF inter area
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         ia - IS-IS inter area, * - candidate default

Gateway of last resort is no set
C       110.11.2.0/24 is directly connected, GigabitEthernet 0/1
C       110.11.2.2/32 is local host.
R       155.10.0.0/16 [120/1] via 110.11.2.1, 00:15:21, GigabitEthernet 0/1
C       196.38.165.0/24 is directly connected, GigabitEthernet 0/2
C       196.38.165.1/32 is local host.
R       117.0.0.0/8 [120/1] via 110.11.2.2, 00:00:47, GigabitEthernet 0/1
```

- Device C

```
Device C# display ip route
Codes:   C - connected, S - static, R - RIP, B - BGP
         O - OSPF, IA - OSPF inter area
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         ia - IS-IS inter area, * - candidate default

Gateway of last resort is no set
C       110.11.2.0/24 is directly connected, GigabitEthernet 0/1
C       110.11.2.3/32 is local host.
C       117.102.0.0/16 is directly connected, GigabitEthernet 0/2
C       117.102.0.1/32 is local host.
R       155.10.0.0/16 [120/1] via 110.11.2.1, 00:20:55, GigabitEthernet 0/1
R       196.38.165.0/24 [120/1] via 110.11.2.3, 00:19:18, GigabitEthernet 0/1
```

6. Configuration Files

- Device A configuration file.

```
!
```

```
interface gigabitethernet 0/1
 unset switch-port
 ip address 110.11.2.1 255.255.255.0
!
interface gigabitethernet 0/2
 unset switch-port
 ip address 155.10.1.1 255.255.255.0
!
router-protocol rip
 version 2
 disable-interface reset
 unset disable-interface gigabitethernet 0/1
 network-advertise 0.0.0.0
!
```

- Device B configuration file.

```
!
interface gigabitethernet 0/1
 unset switch-port
 ip address 110.11.2.2 255.255.255.0
!
interface gigabitethernet 0/2
 unset switch-port
 ip address 196.38.165.1 255.255.255.0
!
router-protocol rip
 version 2
 disable-interface reset
 unset disable-interface gigabitethernet 0/1
 network-advertise 0.0.0.0
!
```

- Device C configuration file.

```
!
interface gigabitethernet 0/1
 unset switch-port
 ip address 110.11.2.3 255.255.255.0
!
interface gigabitethernet 0/2
 unset switch-port
 ip address 117.102.0.1 255.255.255.0
!
router-protocol rip
 version 2
```

```
 unset auto-aggregation
 disable-interface reset
 unset disable-interface gigabitethernet 0/1
 network-advertise 0.0.0.0
!
```

7. Common Errors

- The IPv4 address is not configured on an interface.
- The RIP version is not defined on a device, or the RIP version No. on the device is different from that on other devices.
- The address range configured using the **network-advertise** command does not cover a specific interface.
- The IP address comparison bit parameter in the **network-advertise** command is not correctly configured. Value 0 indicates accurate matching, and 1 indicates that no comparison is performed.
- The interface used for interworking between devices is configured as a passive interface.

1.19.2 Configuring RIP to Prevent Route Loop

1. Requirements

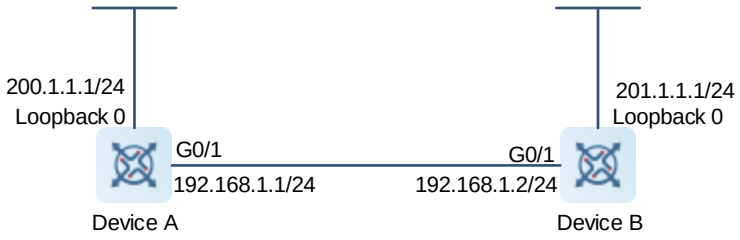
Device A, Device B, and Device C implement interworking through the RIP routing protocol.

To avoid route loop, you need to configure split horizon with poison reverse and triggered updates.

To improve reliability, enable GR, and then enable BFD on the connection interface of Device A and Device B.

2. Topology

Figure 1-10 Configuring RIP to Prevent Route Loop



3. Notes

(1) Configure interface IP addresses for all devices (omitted).

(2) Configure the RIP basic functions for all the devices.

(3) Enable RIP triggered updates and poison reverse on Device A and Device B.

4. Procedure

(1) Configure interface IP addresses for all devices (omitted).

(2) Enable the RIP process of the device and advertise the network information of each interface.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# router-protocol rip
Device A(cfg-router)# network-advertise 192.168.1.0
Device A(cfg-router)# network-advertise 200.1.1.0
Device A(cfg-router)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# router-protocol rip
Device B(cfg-router)# network-advertise 192.168.1.0
Device B(cfg-router)# network-advertise 201.1.1.0
Device B(cfg-router)# exit
```

(3) Enable RIP poison reverse and triggered updates.

Configure Device A.

```
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ip rip triggered-ext
Device A(cfg-if-GigabitEthernet 0/1)# ip rip horizon-split-enable poisoned-reverse
```

Configure Device B.

```
Device B(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ip rip triggered-ext
Device B(cfg-if-GigabitEthernet 0/1)# ip rip horizon-split-enable poisoned-reverse
```

5. Verification

Check the RIP databases of Device A and Device B, and verify that the corresponding route attributes are permanent.

- Device A

```
Device A# display ip rip-instance database
201.1.1.0/24      auto-aggregation
201.1.1.0/24
[1] via 192.168.12.2 GigabitEthernet 0/1    06:25       permanent
```

- Device B

```
Device B# display ip rip-instance database
200.1.1.0/24      auto-aggregation
200.1.1.0/24
      [1] via 192.168.12.1 GigabitEthernet 0/1    06:25       permanent
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
ip rip horizon-split-enable poisoned-reverse
  ip rip triggered-ext
!
!
router-protocol rip
network-advertise 192.168.1.0
network-advertise 200.1.1.0
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
ip rip horizon-split-enable poisoned-reverse
  ip rip triggered-ext
!
router-protocol rip
network-advertise 192.168.1.0
network-advertise 201.1.1.0
!
```

7. Common Errors

- The triggered updates function is enabled when the RIP configurations at both ends of the link are consistent.
- Both the triggered updates and BFD functions are enabled.
- The triggered updates function is not enabled on all devices on the same link.

1.19.3 Configuring External Routes Imported by RIP

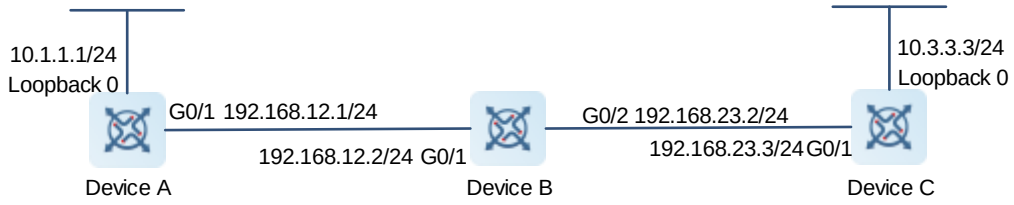
1. Requirements

Device B communicates with Device A through RIP, and Device B obtains the route 10.3.3.3 through a static route.

To enable Device A to obtain the routing information of 10.3.3.3, enable the OSPF routing protocol in Device B. The reliability of introducing external routes is low. Set the external default metric to 5.

2. Topology

Figure 1-11 Configuring External Routes Imported by RIP



3. Notes

(1) Configure an IP address for each interface and start the unicast routing protocol (omitted).

(2) Configure a static route for Device B.

(3) Configure the basic functions of RIP on Device A and Device B (omitted).

(4) Configure route redistribution on Device B and set the default metric of the redistributed route.

4. Procedure

(1) Configure an IP address for each interface and start the unicast routing protocol (omitted).

(2) Configure a static route to 10.3.3.3/24 on Device B, and set the next hop to 192.168.23.3.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ip route-config 10.3.3.3 255.255.255.0 192.168.23.3
```

(3) Configure the basic functions of RIP on Device A and Device B (omitted).

(4) Set the default metric of the redistributed route to 5 on Device B.

```
Device B(cfg)# router-protocol rip
Device B(cfg-router)# default-metric 5
```

(5) Configure the redistribution of static route to RIP on Device B.

```
Device B(cfg-router)# redistribute-route static
```

5. Verification

Check the routing table on Device A. Verify that the RIP route of 10.3.3.0/24 is available, and the metric is 6.

```
Device A# display ip route-config rip
R       10.3.3.0/24 [120/6] via 192.168.12.2, 00:06:11, GigabitEthernet 0/1
```

6. Configuration Files

Device B configuration file.

```
!
```

```
ip route-config 10.3.3.3 255.255.255.0 192.168.23.3
!
router-protocol rip
 default-metric 5
 redistribute-route static
!
```

1.19.4 Configuring Increasing the Metric of a RIP Interface

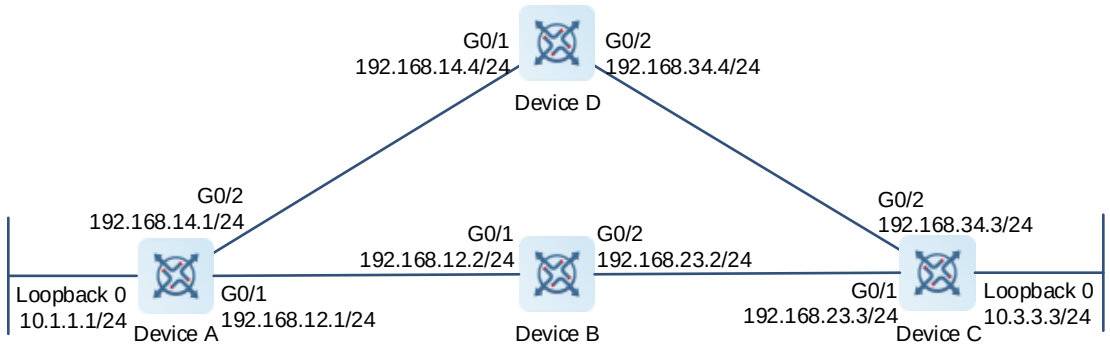
1. Requirements

Device A, Device B, Device C, and Device D implement interworking through the RIPv2 routing protocol.

There are two equal-cost paths from Device A to 10.3.3.3/24 of Device C. In consideration of the old model and low performance of Device D, increase the metric of the Gigabit Ethernet 0/2 interface of Device A so that Device A can go to 10.3.3.3 to prefer Device B.

2. Topology

Figure 1-12 Configuring Metric Increment of a RIP Interface



3. Notes

(1) Configure interface IP addresses for all the devices. (Omitted)

(2) Configure the RIP basic functions for all the devices. (Omitted)

(3) Configure ACL on Device C, and enable offset-list on the interface connected to Device B.

4. Procedure

(1) Configure ACL on Device C, and match the route 10.3.3.3/24.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# access-control-list 1 permit 10.3.3.3 0.0.0.255
```

(2) Configure Offset-List on Device C, associate with ACL 1, and enable the offset-list on the GigabitEthernet 0/2 interface.

```
Device C(cfg)# router-protocol rip
Device C(cfg-router)# metric-offset-list 1 out 3 gigabitEthernet 0/2
```

5. Verification

Check the routing table of Device A.

```
Device A# display ip route-config rip
R       10.3.3.0/24 [120/5] via 192.168.14.4, 00:06:11, GigabitEthernet 0/1
```

6. Configuration Files

Device C configuration file.

```
!
access-control-list 1 permit 10.3.3.3 0.0.0.255
!
router-protocol rip
 metric-offset-list 1 out 3 gigabitEthernet 0/2
!
```

1.19.5 Configuring RIP to Distribute a Summarized Route

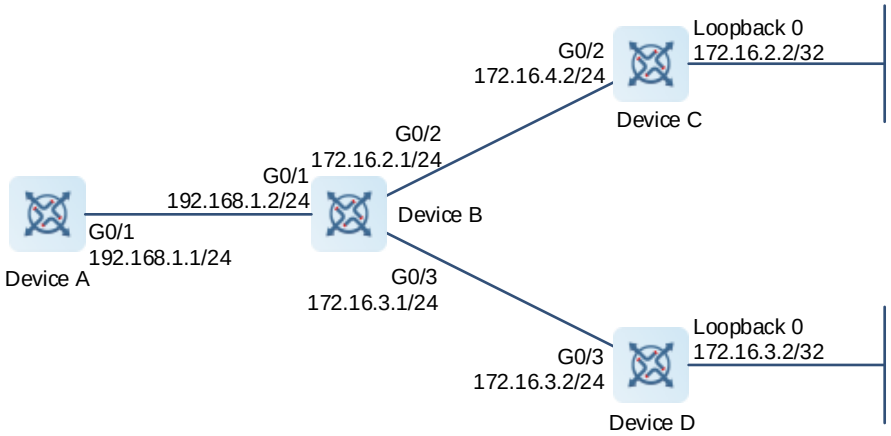
1. Requirements

Device B, implements interworking with Device A, Device C, and Device D through the RIPv2 routing protocol.

To reduce the size of the routing table of Device A, configure route summarization to 172.16.0.0/21 on Device B.

2. Topology

Figure 1-13 Configuring RIP to Distribute a Summarized Route



3. Notes

(1) Configure interface IP addresses for all devices (omitted).

(2) Configure the RIP basic functions for all the devices (omitted).

(3) Disable automatic summarization on Device B and configure route summarization.

4. Procedure

(1) Configure an IP address for each interface and start the unicast routing protocol (omitted).

(2) Configure RIP basic functions on the devices (omitted).

(3) Disable automatic summarization on Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# router-protocol rip
Device B(cfg-router)# version 2
Device B(cfg-router)# unset auto-aggregation
Device B(cfg-router)# exit
```

(4) Configure 172.16.0.0/21 route summarization on the GigabitEthernet 0/1 interface of Device B.

```
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip rip summary-address 172.16.0.0 255.255.248.0
```

5. Verification

Check the routing table of Device A, and verify that the entry 172.16.0.0/16 is generated.

```
Device A# display ip route-config rip
R 172.16.0.0/21 [120/2] via 192.168.1.2, 00:01:04, gigabitethernet 0/1
```

6. Configuration Files

Device B configuration file.

```
!
interface gigabitethernet 0/1
ip address 196.38.165.1 255.255.255.0
!
router-protocol rip
 version 2
 unset auto-aggregation
!
```

Contents

1 Configuring RIPng 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 6
1.2 Configuration Task Summary 6
1.3 Configuring RIPng Basic Functions 7
1.3.1 Overview 7
1.3.2 Restrictions and Guidelines 7
1.3.3 Configuration Tasks 7
1.3.4 Enabling an RIPng Routing Process 7
1.3.5 Configuring RIPng on an Interface 8
1.3.6 Configuring Split Horizon or Poison Reverse 9
1.3.7 Configuring a Passive Interface 9
1.4 Configuring RIPng Route Distribution 10
1.4.1 Overview 10
1.4.2 Configuration Tasks 10
1.4.3 Generating a Default Route 10
1.4.4 Configuring External Route Redistribution 11
1.5 Configuring RIPng Route Filtering Rules 11
1.5.1 Overview 11
1.5.2 Procedure 11

1.6 Configuring to Adjust RIPng Routing....12
1.6.1 Overview....12
1.6.2 Configuration Tasks....12
1.6.3 Configuring to Modify the Management Distance of an RIPng Route....12
1.6.4 Configuring to Modify the Metric Offset on an Interface....13
1.6.5 Configuring the Default Metric of a Route Redistributed to RIPng....14
1.7 Modifying Timers....14
1.7.1 Overview....14
1.7.2 Restrictions and Guidelines....14
1.7.3 Procedure....14
1.8 Enabling GR....15
1.8.1 Overview....15
1.8.2 Restrictions and Guidelines....15
1.8.3 Procedure....15
1.9 Configuring a Super VLAN to Enable RIPng....16
1.9.1 Overview....16
1.9.2 Restrictions and Guidelines....16
1.9.3 Procedure....16
1.10 Monitoring....16
1.11 Configuration Examples....17
1.11.1 Configuring Basic Functions of RIPng....17
1.11.2 Advertising Default Routes/External Routes....21

1 Configuring RIPng

1.1 Introduction

1.1.1 Overview

RIP next generation (RIPng) is a unicast routing protocol that applies to IPv6 networks. RIPng-enabled routers exchange routing information to obtain routes to remote networks.

As an Interior Gateway Protocol (IGP), RIPng can run only within an autonomous system (AS) and is applicable to small-sized networks.

Note

In this chapter, the term “router” refers to any network device that supports the routing function. These network devices can be Layer-3 switches, routers, or firewalls.

1.1.2 Principles

1. Basic Concepts

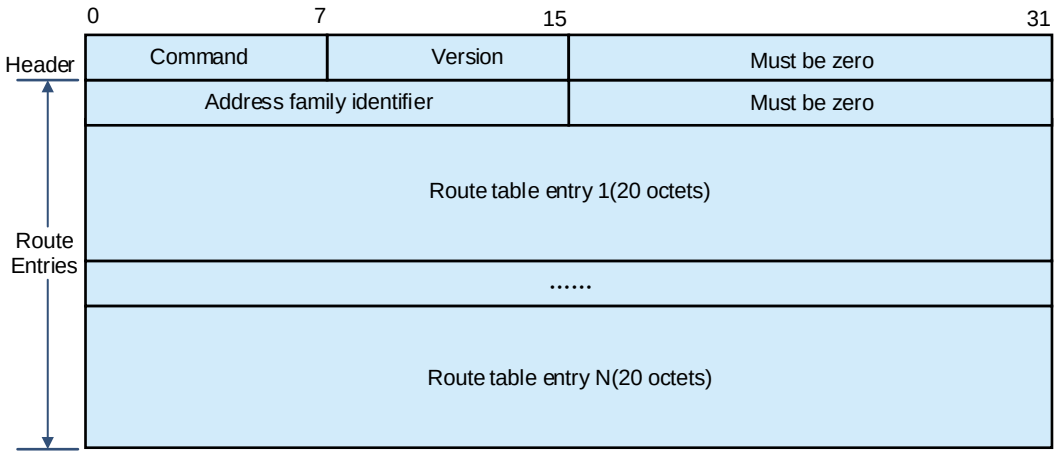
- IGP and EGP

  IGP runs within an AS. For example, RIPng is a type of IGP.

  Exterior Gateway Protocol (EGP) runs between ASs. For example, BGP is a type of EGP.

- RIPng packet format

Figure 1-1 RIPng Packet Format



RIPng packets can contain one or more routing table entries. The number of routing table entries depends on the size of the maximum transmission unit (MTU) value of the interface. Route table entry (RTE) is classified into two types:

- Next hop RTE: Located in the foremost of a set of routing table entries with the same next hop, that is,

the position of Route table entry 1 shown in Figure 1-1.

- IPv6 prefix RTE: Located behind the next hop RTE. Multiple different IPv6 prefix RTEs can exist after a next hop RTE. An IPv6 prefix RTE describes the destination IP address, prefix length, route tag, and metric of an RIPng route.

Figure 1-2 Format of Next Hop RTE

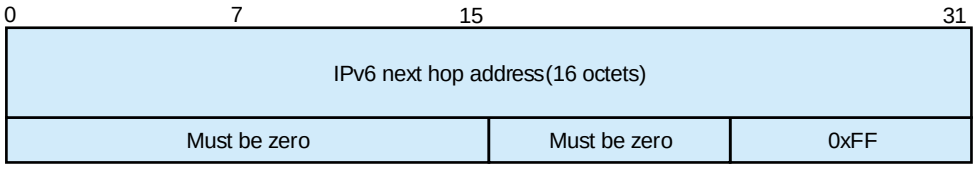
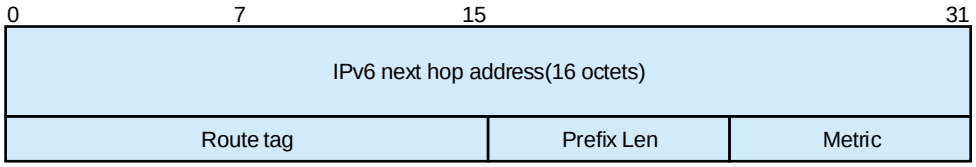


Figure 1-3 Format of IPv6 Prefix RTE



2. RIPng and RIP

RIP applies to IPv4 networks. Two RIP versions are available: RIPv1 and RIPv2.

RIPng is an extension of RIPv2 on the basis of IPv6. Both are similar in functions and configurations. The main differences are as follows:

- The UDP port ID of RIP is 520, and that of RIPng is 521.
- IPv6 features.
  - The RIPv2 multicast address is 224.0.0.9, and the RIPng multicast address is FF02::9.
  - The source address of RIPng route update packets is FE80::/10.

3. Route Control

Compared with static routing, the dynamic routing protocol shows an advantage, that is, by exchanging routing information, devices can automatically obtain routes to a remote network and update the routes in real time.

- Routing switching process

  After RIPng is enabled on a router, the router sends a request packet to its neighbor router, requesting all routing information, that is, the routing table. After receiving the request packet, the neighbor device returns a response packet containing the local routing table. After receiving the response packet, the router updates the local routing table, and sends an update packet to the neighbor router, informing the neighbor router of the route update information. After receiving the update packet, the neighbor router updates the local routing table, and sends the update packet to other adjacent routers. After receiving and processing an update advertisement, all the routers can obtain and retain the latest routing information.

- Periodical update

By default, periodical update is enabled for RIPng. Adjacent routers exchange complete routing information with each other every 30s, that is, the entire routing table is sent to neighbor routers.

Note

For every non-local route, if the route is not updated within 180s (Invalid timer), the metric of the route is changed to 16 (unreachable). If the route is still not updated in the next 120s (Flush timer), the route is deleted from the routing table.

The timer details are as follows:

- Update timer: Route update packets are sent regularly, and the update interval is 30s by default.
- Invalid timer: If an update packet for refreshing the existing route has not been received after 180s (default value), the metric for the route to set to 16, marking the route as invalid.
- Flush timer: By default, the flush timer is set to 240s, 60s greater than the invalid timer. When the flush timer times out, the route is deleted from the routing table.
- Hold-down timer: It prevents route loop caused by route flapping. When the device receives a route update with a cost of 16, the route will be set to hold-down state, and the hold-down timer will be started at the same time. Before the hold-down timer expires, even if an update packet with a metric less than 16 for this route is received, the route will not be updated.

- Default route

  In the routing table, a route to the destination network ::/0 is called default route.

  The default route can be learned from a neighbor device, or sent to a neighbor device.

- Route redistribution

  For RIPng, other types of routes (such as direct routes, static routes, and routes of other routing protocols) are called external routes.

  External routes (excluding the default route) can be redistributed to RIPng and advertised to neighbors.

- Route Filtering

  Filtering conditions can be configured to limit the routing information exchanged between adjacent routers. Only the routing information that meets filtering conditions can be sent or received.

4. Routing Algorithm

RIPng is a protocol based on the distance-vector algorithm. It uses the vector addition method to compute the routing information.

- Distance-vector algorithm

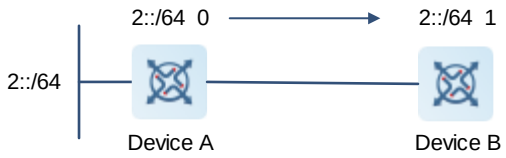
  RIPng is a protocol based on the distance-vector algorithm. The distance-vector algorithm treats a route as a vector that consists of the destination network and distance. The router obtains a route from its neighbor and adds the distance vector from itself to the neighbor to the route to form its own route.

  RIPng uses the hop count to evaluate the distance to the destination network. By default, the hop count from a router to its directly connected network is 0, the hop count from a router to a network that can be reached through a router is 1, and so on. That is, the metric is equal to the number of routers from the local network to the destination network. To restrict the convergence time, RIPng stipulates that the metric must

be an integer in the range from 0 to 15. If the metric is equal to or greater than 16, the destination network or host is unreachable. For this reason, RIPng cannot be applied on a large-scale network.

As shown in Figure 1-4, Device A is connected to the network 2::/64. Device B obtains the route (2::/64, 0) from Device A and adds the metric 1 to the route to obtain its own route (2::/64, 1), and the next hop points to Device A.

Figure 1-4 Diagram of Distance-vector Algorithm

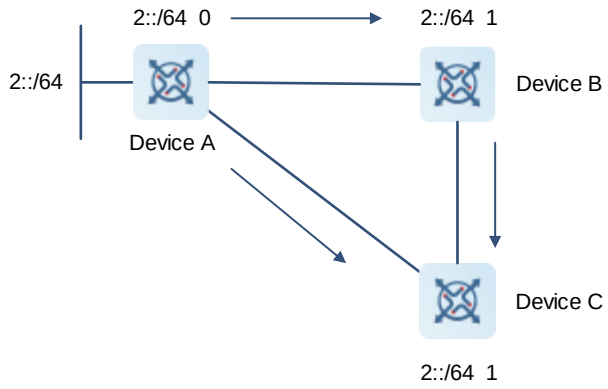


- Selecting the optimum route

  RIPng selects an optimum route based on the following principle: If multiple routes to the same destination network are available, a router preferentially selects the route with the smallest metric.

  As shown in Figure 1-5, Device A is connected to the network 2::/64. Device C obtains the route (2::/64, 0) from Device A and the route (2::/64, 1) from Device B. Device C will select the route that is obtained from Device A and add metric 1 to this route to form its own route (2::/64, 1), and the next hop points to Device A.

Figure 1-5 Diagram of Selecting the Optimum Route



Note

When routes coming from different sources exist on a router, the route with the smallest distance is preferentially selected.

Table 1-1 Distance-vector Table

Route Source	Default Distance
Directly-connected network	0
Static route	1

Route Source	Default Distance
OSPF route	110
IS-IS route	115
RIP route	120
Unreachable route	255

5. Avoiding Route Loops

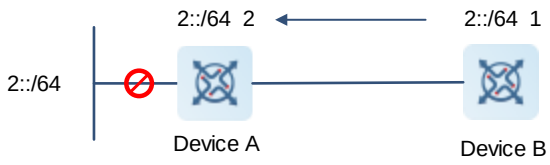
RIPng uses functions such as split horizon and poison reverse to avoid route loops.

- Route loop

  An RIPng route loop occurs due to inherent defects of the distance-vector algorithm.

  As shown in Figure 1-6, Device A is connected to the network 2::/ 64, and sends an update packet every 30s. Device B receives the route to 2::/64 from Device A every 30s. If Device A is disconnected from 2::/64, the route to 2::/64 will be deleted from the routing table on Device A. Next time, the update packet sent by Device A no longer contains this route. As Device B does not receive an update packet related to 2::/64, Device B determines that the route to 2::/64 is valid within 180s and uses the update packet to send this route to Device A. As the route to 2::/64 does not exist on Device A, the route learned from Device B is added to the routing table. Device B determines that data can reach 2::/64 through Device A, and Device A determines that data can reach 2::/64 through Device B. In this way, a route loop is formed.

Figure 1-6 Diagram of Route Loop

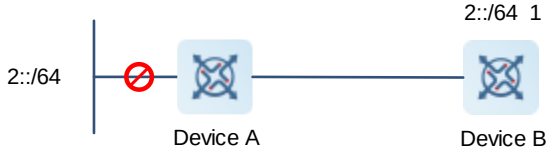


- Split horizon

  Split horizon can prevent route loops. After split horizon is enabled, a route received on this interface will not be sent out from this interface.

  As shown in Figure 1-7, after split horizon is enabled on Device B, Device B will not send the route to 2::/64 back to Device A. Device B will learn 180s later that 2::/64 is not reachable.

Figure 1-7 Diagram of Split Horizon



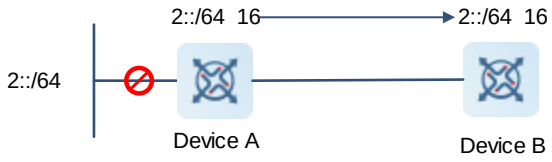
- Poison reverse

Poison reverse can also prevent route loops. Compared with slit horizon, poison reverse is more reliable, but brings more protocol packets, which makes network congestion more severe.

After poison reverse is enabled on an interface, a route received from this interface will be sent out from this interface again, but the metric of this router will be changed to 16 (unreachable).

As shown in Figure 1-8, after poison reverse is enabled on Device A, if Device A detects a disconnection from 2::/64, Device A will not delete the route to 2::/64. Instead, Device A changes the number of hops to **16**, and advertises the route through the update packet. Upon receiving the update packet, Device B learns that 2::/64 is not reachable.

Figure 1-8 Diagram of Poison Reverse



6. RIPng Authentication

IPv6 provides two authentication mechanisms, that is, authentication header (AH) and encapsulating security payload (ESP). RIPng use the authentication mechanisms to prevent routers that access the network illegally and hosts that forge RIPng packets from participating in RIPng routing switching. RIPng packets received on the RIPng interface are authenticated and decrypted. If authentication or decryption fails, the packets are dropped and do not participate in route learning.

Enabling authentication can avoid learning unauthenticated or invalid routes, thus preventing advertising valid routes to unauthenticated devices. At present, RIPng authentication is implemented by associating with the security association (SA) authentication of IPsec. SA defines a set of security algorithms and keys to ensure IP security. If a service is associated with IPsec SA authentication, the RIPng service will use an SA authentication key to authenticate and encrypt packets. For example, the definition of SM3 (OSCCA algorithm) authentication is added to SA authentication. For the RIPng service, only the SA authentication needs to be associated to encrypt/decrypt packets using SM3 authentication keys.

1.1.3 Protocols and Standards

- RFC 2080: RIPng for IPv6
- RFC 2081: RIPng Protocol Applicability Statement

1.2 Configuration Task Summary

The RIPng configuration includes the following tasks:

(1) Configuring RIPng Basic Functions

a Enabling an RIPng Routing Process

b Configuring RIPng on an Interface

c (Optional) Configuring Split Horizon or Poison Reverse

d (Optional) Configuring a Passive Interface

(2) (Optional) Configuring RIPng Route Distribution The following configuration tasks are optional. Select tasks for configuration according to actual condition.

- Generating a Default Route
- Configuring External Route Redistribution

(3) (Optional) Configuring RIPng Route Filtering Rules

(4) (Optional) Configuring to Adjust RIPng Routing The following configuration tasks are optional. Select tasks for configuration according to actual condition.

- Configuring to Modify the Management Distance of an RIPng Route
- Configuring to Modify the Metric Offset on an Interface
- Configuring the Default Metric of a Route Redistributed to RIPng Metric

(5) (Optional) Modifying Timers

(6) (Optional) Enabling GR

(7) (Optional) Configuring a Super VLAN to Enable RIPng

1.3 Configuring RIPng Basic Functions

1.3.1 Overview

This function can be used to build an RIPng routing domain on the network, and routers in the domain obtain routes to a remote network through RIPng.

1.3.2 Restrictions and Guidelines

- IPv6 addresses must be configured.
- IPv6 unicast routes must be enabled.

1.3.3 Configuration Tasks

The basic function configuration of RIPng includes the following tasks:

(1) Enabling an RIPng Routing Process

(2) Configuring RIPng on an Interface

(3) (Optional) Configuring Split Horizon or Poison Reverse

(4) (Optional) Configuring a Passive Interface

1.3.4 Enabling an RIPng Routing Process

1. Overview

This command is used to create an RIPng routing process and enter the routing configuration mode.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable an RIPng routing process and enter route configuration mode.

ipv6 router-protocol ripng

The RIPng routing process is disabled by default.

1.3.5 Configuring RIPng on an Interface

1. Overview

In RIPv2, the **network-type** command is configured in the routing configuration mode to define an IP address range. In RIPng, if the IP address of an interface belongs to this IP address range, RIPng automatically runs on this interface.

After RIPng runs on an interface, RIPng packets can be exchanged on the interface and RIPng can learn routes to the network segments directly connected to the device.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*
- Enter the virtual PPP interface configuration mode.
  - **interface virtual-ppp** *interface-number*
- Enter the virtual template interface configuration mode.
  - **interface virtual-template** *interface-number*

(4) Configure RIPng on an interface.

ipv6 rip enable

By default, an interface is not added to the RIPng process.

1.3.6 Configuring Split Horizon or Poison Reverse

1. Overview

Split horizon can prevent route loops. After split horizon is enabled, a route received on this interface will not be sent out from this interface.

After poison reverse is enabled on an interface, a route received from this interface will be sent out from this interface again, but the metric will be changed to **16** (unreachable).

2. Restrictions and Guidelines

- You are advised to enable split horizon on the interfaces of broadcast type network (such as Ethernet) and point-to-point type network (such as PPP and HDLC). (Retain the default setting.)
- It is recommended that split horizon and poison reverse be disabled on an interface to a non-broadcast multi-access network such as FR and X.25; otherwise, some devices cannot learn the complete routing information.
- If the secondary IP address is configured for an interface connected to a non-broadcast network, it is recommended that split horizon and poison reverse be disabled.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the routing configuration mode.

ipv6 router-protocol rip

(4) Configure split horizon.

horizon-split-enable [**poisoned-reverse**]

By default, split horizon is enabled and poison reverse is disabled on an interface.

1.3.7 Configuring a Passive Interface

1. Overview

A passive interface is set with the boundary of RIPng routing domain. The network segment of the passive interface belongs to the RIPng routing domain, but RIPng packets cannot be sent over the passive interface.

The interface set to a passive interface suppresses RIPng update packets. The passive interface defines the boundary of RIPng routing domain to avoid unwanted flooding of RIPng packets. If the interface connection device does not run the RIPng routing protocol (such as a PC and a device running other routing protocols), you are advised to configure this interface as a passive interface.

2. Restrictions and Guidelines

If RIPng routes need to be exchanged on an interface (such as the router interconnect interface) in the RIPng routing domain, this interface cannot be configured as a passive interface.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the routing configuration mode.

ipv6 router-protocol rip

(4) Configure passive interfaces.

disable-interface { **default** | *interface-type interface-number* }

The passive interface function is disabled by default.

1.4 Configuring RIPng Route Distribution

1.4.1 Overview

RIPng can distribute the routes of other routing protocols or default routes in the local routing domain so that the devices in the routing domain can access other routing domains.

1.4.2 Configuration Tasks

The configuration of RIPng route distribution includes the following tasks, and these tasks are optional. Select tasks for configuration according to actual condition.

- Generating a Default Route
- Configuring External Route Redistribution

1.4.3 Generating a Default Route

1. Overview

This function is used to introduce a default route to an AS edge router so that other routers in the RIPng domain access other AS domains through this AS edge router by default. In the routing table, a route to the destination network ::/0 is called default route. The default route can be learned from a neighbor device, or sent to a neighbor device. Please configure and distribute the default route according to the actual situation of networking, or specify the cost of the distributed default route.

2. Restrictions and Guidelines

- When there are no special requirements and this command is configured on the interface, an IPv6 default route is advertised to the external devices through this interface, but the route itself is not added to the route forwarding table of the device and the RIPng route database.
- To prevent occurrence of a route loop, once this command is configured on an interface, RIPng refuses to receive the default route updates advertised by neighbors.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure external route redistribution.

ipv6 ripng default-info { **only** | **originate** } [**metric-value** *metric-value*]

1.4.4 Configuring External Route Redistribution

1. Overview

This function is used to introduce external routes of the RIPng domain to the AS edge device.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the routing configuration mode.

ipv6 router-protocol rip

(4) Generate a default route.

redistribute-route { **bgp** | **connected** | **isis** [*area-tag*] **| ospf** process-id **| static** } [**metric-value** *metric-value* | **route-map** *route-map-name*] *

By default, the function of redistributing routes from other routing domains to RIPng is not configured.

1.5 Configuring RIPng Route Filtering Rules

1.5.1 Overview

Routes that do not meet filtering criteria cannot be loaded to the routing table, or advertised to neighbors. In this way, users within the network can be prevented from accessing specified destination networks.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a prefix list.

ip prefix-list-info *prefix-list-name* [**seq** *seq-number*] { **deny** | **permit** } *ipv4-prefix* [**ge** *minimum-prefix-length*] [**le** *maximum-prefix-length*]

(4) Enter the routing configuration mode.

ipv6 router-protocol ripng

(5) Configure to filter the sent and received RIP routing information.

route-filtering prefix-list *prefix-list-name* { **in** | **out** } [*interface-type interface-number*]

By default, the distribution list is disabled.

1.6 Configuring to Adjust RIPng Routing

1.6.1 Overview

Change the RIPng routes to enable the traffic pass through specified nodes or avoid passing through specified nodes. Change the sequence that a router selects various types of routes so as to change the priorities of RIPng routes.

1.6.2 Configuration Tasks

The configuration for modifying routing parameters includes the following tasks, and these tasks are optional. Select tasks for configuration according to actual condition.

- Configuring to Modify the Management Distance of an RIPng Route
- Configuring to Modify the Metric Offset on an Interface
- Configuring the Default Metric of a Route Redistributed to RIPng Metric

1.6.3 Configuring to Modify the Management Distance of an RIPng Route

1. Overview

This function is used to set the management distance of a RIPng route and change the priority of a router in routing.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the routing configuration mode.

ipv6 router-protocol rip

(4) Configure to modify the management distance of an RIPng route.

distance *distance*

The management distance is **120** by default.

1.6.4 Configuring to Modify the Metric Offset on an Interface

1. Overview

Before a route is added to the routing table, the metric of the route must be added with the metric offset set on the interface. You can control the use of a route by setting the interface metric offset.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*
- Enter the virtual PPP interface configuration mode.
  - **interface virtual-ppp** *interface-number*
- Enter the virtual template interface configuration mode.
  - **interface virtual-template** *interface-number*

(4) Configure to modify the metric offset to be added to the interface of an RIPng route.

ipv6 ripng metric-adjustment *value*

By default, the metric on an interface is **1**.

1.6.5 Configuring the Default Metric of a Route Redistributed to RIPng

1. Overview

The default metric of a redistributed route is **1**. If you want to preferably select an internal RIP route on some networks, just increase the default metric.

2. Restrictions and Guidelines

- Unless otherwise required, perform this configuration on an AS edge device to which external routes are introduced.
- If the metric is not specified during redistribution of a routing protocol process, RIPng uses the metric defined by the **default-metric** command. If the metric is specified, the metric defined by the **default-metric** command is overwritten by the specified metric.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the routing configuration mode.

ipv6 router-protocol rip

(4) Configure the default metric of a route redistributed to RIPng.

default-metric *metric*

By default, the metric for route redistribution is 1.

1.7 Modifying Timers

1.7.1 Overview

By adjusting the timers, you can reduce the convergence time and fault rectification time of the routing protocol. For devices connected to the same network, values of the three RIPng timers must be the same. Generally, you are not advised to modify the RIP timers unless otherwise required.

Setting timers to small values on a low-speed link brings risks because a lot of Update packets consume the bandwidth. You can set timers to small values generally on the Ethernet or a 2 Mbps (or above) link to reduce the convergence time of network routes.

1.7.2 Restrictions and Guidelines

- The RIPng basic functions must be configured.
- Modifying the protocol control parameters may result in protocol running failures. Therefore, you are advised not to modify the timers.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the routing configuration mode.

ipv6 router-protocol rip

(4) Configure to modify the RIPng timers.

timers-set *update invalid flush*

By default, the update timer is 30s, the invalid timer is 180s, and the flush timer is 120s.

1.8 Enabling GR

1.8.1 Overview

GR ensures uninterrupted data transmission when the protocol is restarted. The GR period is the maximum time from restart of the RIPng process to completion of GR. During this period, the forwarding table before the restart is retained, and the RIPng route is restored so as to restore the RIPng state before the restart. After the restart period expires, RIPng exits the GR state and performs common RIPng operations.

1.8.2 Restrictions and Guidelines

- The RIPng basic functions must be configured.
- The GR period is at least twice the RIP route update period.
- The **gr grace-period** command allows you to explicitly modify the GR period. Note that GR must be completed before the invalid timer of the RIPng route expires. An inappropriate GR period cannot ensure uninterrupted data forwarding during the GR process. A typical case is as follows: If the GR period is longer than the duration of the invalid timer, GR is not completed when the invalid timer expires. The route is not re-advertised to the neighbor, and forwarding of the route of the neighbor stops after the invalid timer expires, causing interruption of data forwarding on the network. Unless otherwise required, you are not advised to adjust the GR period. If it is necessary to adjust the GR period, ensure that the GR period is at least twice the RIP route update period based on the configuration of the **timers-set** command. The GR period must be also smaller than the duration of the invalid timer.
- During the RIPng GR process, ensure that the network environment is stable.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the routing configuration mode.

ipv6 router-protocol rip

(4) Enable GR.

gr [**grace-period** *grace-period*]

The GR function is enabled by default.

1.9 Configuring a Super VLAN to Enable RIPng

Note

This function is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.9.1 Overview

In normal cases, a super VLAN contains multiple sub VLANs. Multicast packets of a super VLAN are also sent to its sub VLANs. In this case, when RIP multicast packets are sent over a super VLAN containing multiple sub VLANs, the RIP multicast packets are replicated multiple times, and the device processing capability is insufficient. As a result, a large number of packets are discarded, causing protocol flapping.

In most scenarios, the RIPng function does not need to be enabled on a super VLAN, and it is disabled by default. However, in some scenarios, the RIPng function must be run on the super VLAN, but packets need to be sent to only one sub VLAN. In this case, you can decide to send multicast packets to a certain sub VLAN or to all sub VLANs as actually needed. You can run this command to specify a particular sub VLAN. You must be cautious when configuring packet transmission to all sub VLANs, as the large number of sub VLANs may cause a device processing bottleneck, which will lead to the neighbor flapping.

1.9.2 Restrictions and Guidelines

- The RIPng basic functions must be configured.
- The designated sub VLAN must be connected to neighbors.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI interface configuration mode.

interface vlan *interface-number*

(4) Enable RIPng on a super VLAN.

ipv6 rip subvlan [**all |** *vlan-id*]

By default, the RIPng function is disabled on a super VLAN.

1.10 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-2 Monitoring

Command	Purpose
display ipv6 rip	Displays information about the RIPng process.
display ipv6 rip database	Displays the RIPng routing table.
debugging ipv6 rip [interface *interface-type interface-number* | **nsm** | **restart** | **event** [**ipsec**]]	Debugs RIPng.

1.11 Configuration Examples

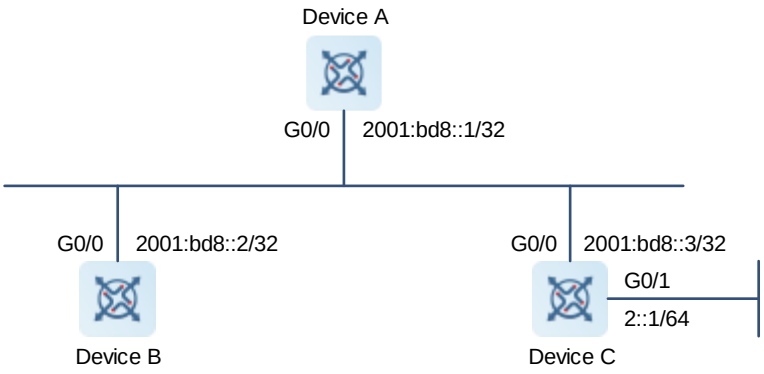
1.11.1 Configuring Basic Functions of RIPng

1. Requirements

On an IPv6 network, configure RIPng to realize internetworking.

2. Topology

Figure 1-9 Configuring Basic Functions of RIPng



3. Notes

- Configure IPv6 addresses on the related devices.
- Enable the RIPng routing function on all the devices.

4. Procedure

(1) Configure an IPv6 address for each interface.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/0
Device A(cfg-if-GigabitEthernet 0/0)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/0)# ipv6 address 2001:db8::1/32
Device A(cfg-if-GigabitEthernet 0/0)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/0
Device B(cfg-if-GigabitEthernet 0/0)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/0)# ipv6 address 2001:db8::2/32
Device B(cfg-if-GigabitEthernet 0/0)# exit
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# interface gigabitethernet 0/0
Device C(cfg-if-GigabitEthernet 0/0)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/0)# ipv6 address 2001:db8::3/32
Device C(cfg-if-GigabitEthernet 0/0)# exit
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/0)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 address 2::1/64
Device C(cfg-if-GigabitEthernet 0/1)# exit
```

(2) Enable the RIPng routing function on all the devices.

Configure Device A.

```
Device A(cfg)# ipv6 router-protocol rip
Device A(cfg-router)# exit
Device A(cfg)# interface gigabitethernet 0/0
Device A(cfg-if-GigabitEthernet 0/0)# ipv6 rip enable
```

Configure Device B.

```
Device B(cfg)# ipv6 router-protocol ripng
Device B(cfg-router)# exit
Device B(cfg)# interface gigabitethernet 0/0
Device B(cfg-if-GigabitEthernet 0/0)# ipv6 ripng enable
```

Configure Device C.

```
Device C(cfg)# ipv6 router-protocol ripng
Device C(cfg-router)# exit
Device C(cfg)# interface gigabitethernet 0/0
Device C(cfg-if-GigabitEthernet 0/0)# ipv6 ripng enable
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 ripng enable
```

5. Verification

Check the routing tables of Device A, Device B, and Device C. Verify that RIPng has learned the routes to remote networks.

- Check the routing table of Device A.

```
Device A# display ipv6 route-config
IPv6 routing table name - Default - 6 entries
```

```
Codes:  C - Connected, L - Local, S - Static
        R - RIP, O - OSPF, B - BGP, I - IS-IS, V - Overflow route
        N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
        E1 - OSPF external type 1, E2 - OSPF external type 2
        SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
        IA - Inter area

R       2::/64 [120/2] via FE80::2D0:F8FF:FEFB:D521, GigabitEthernet 0/0
C       2001:DB8::/32 via GigabitEthernet 0/0, directly connected
L       2001:DB8::1/128 via GigabitEthernet 0/0, local host
C       FE80::/10 via ::1, Null0
C       FE80::/64 via GigabitEthernet 0/0, directly connected
L       FE80::2D0:F8FF:FEFB:E7CE/128 via GigabitEthernet 0/0, local host
```

- Check the routing table of Device B.

```
Device B# display ipv6 route-config

IPv6 routing table name - Default - 6 entries
Codes:  C - Connected, L - Local, S - Static
        R - RIP, O - OSPF, B - BGP, I - IS-IS, V - Overflow route
        N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
        E1 - OSPF external type 1, E2 - OSPF external type 2
        SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
        IA - Inter area

R       2::/64 [120/2] via FE80::2D0:F8FF:FEFB:D521, GigabitEthernet 0/0
C       2001:DB8::/32 via GigabitEthernet 0/0, directly connected
L       2001:DB8::2/128 via GigabitEthernet 0/0, local host
C       FE80::/64 via GigabitEthernet 0/0, directly connected
L       FE80::2D0:F8FF:FEFB:C9BA/128 via GigabitEthernet 0/0, local host
```

- Check the routing table of Device C.

```
Device C# display ipv6 route-config

IPv6 routing table name - Default - 9 entries
Codes:  C - Connected, L - Local, S - Static
        R - RIP, O - OSPF, B - BGP, I - IS-IS, V - Overflow route
        N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
        E1 - OSPF external type 1, E2 - OSPF external type 2
        SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
        IA - Inter area

C       2::/64 via GigabitEthernet 0/1, directly connected
L       2::2/128 via GigabitEthernet 0/1, local host
C       2001:DB8::/32 via GigabitEthernet 0/0, directly connected
L       2001:DB8::3/128 via GigabitEthernet 0/0, local host
C       FE80::/10 via ::1, Null0
```

```
C        FE80::/64 via GigabitEthernet 0/0, directly connected
L        FE80::2D0:F8FF:FEFB:D521/128 via GigabitEthernet 0/0, local host
C        FE80::/64 via GigabitEthernet 0/1, directly connected
L        FE80::2D0:F8FF:FEFB:D521/128 via GigabitEthernet 0/1, local host
```

6. Configuration Files

- Device A configuration file

```
!
ipv6 router-protocol rip
!
interface gigabitethernet 0/0
 unset switch-port
 ipv6 address 2001:db8::1/32
 ipv6 rip enable
!
```

- Device B configuration file

```
!
ipv6 router-protocol rip
!
interface gigabitethernet 0/0
 unset switch-port
 ipv6 address 2001:db8::2/32
 ipv6 rip enable
!
```

- Device C configuration file

```
!
ipv6 router-protocol rip
!
interface gigabitethernet 0/0
 unset switch-port
 ipv6 address 2001:db8::3/32
 ipv6 rip enable
!
interface gigabitethernet 0/1
 unset switch-port
 ipv6 address 2::1/64
 ipv6 rip enable
!
```

7. Common Errors

- The IPv6 address is not configured on an interface.
- The interface used for interworking between devices is configured as a passive interface.

1.11.2 Advertising Default Routes/External Routes

1. Requirements

On a network adopting OSPFv3 networking, Device A and Device B are egress devices on the network, Device A is the default egress used to advertise the default route to Device B, and Device B communicates with the internal network through a static route.

2. Topology

Figure 1-10 Configuring Default Routes/External Routes



3. Notes

- Configure interface IPv6 addresses for all devices (omitted).
- Configure the RIPng basic functions on all devices (omitted).
- Redistribute the static route on Device B.
- On the GigabitEthernet 0/1 interface of Device A, configure advertisement of the default route.

4. Procedure

(1) Configure redistribution of a static route on Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ipv6 router-protocol rip
Device B(cfg-router)# redistribute-route static
```

(2) Configure redistribution of a default route on Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 rip default-info originate
```

5. Verification

Check the routing tables on Device A and Device B, and verify that Device A can learn the route 3001:10:10::/64, and Device B can learn the default route ::/0.

```
Device A# display ipv6 route-config rip

IPv6 routing table name - Default - 17 entries
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPF, B - BGP, I - IS-IS
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
```

```
            SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
            IA - Inter area
R           3001:10:10::/64 [120/2] via FE80::2D0:F8FF:FE22:334A, GigabitEthernet 0/1
Device B# display ipv6 route-config rip

IPv6 routing table name - Default - 17 entries
Codes:    C - Connected, L - Local, S - Static
            R - RIP, O - OSPF, B - BGP, I - IS-IS
            N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
            E1 - OSPF external type 1, E2 - OSPF external type 2
            SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
            IA - Inter area
R           ::/0 [120/2] via FE80::21A:A9FF:FE41:5B06, GigabitEthernet 0/1
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
 unset switch-port
 ipv6 rip default-info originate
!
```

- Device B configuration file

```
!
ipv6 router-protocol ripng
 redistribute-route static
!
```

Contents

1 Configuring OSPFv2 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Concepts 2
1.1.3 Principles 14
1.1.4 OSPF Routing 15
1.1.5 OSPF Route Advertisement 16
1.1.6 OSPF Route Summarization 17
1.1.7 OSPF Route Filtering 17
1.1.8 Enhanced Security and Reliability 17
1.1.9 Network Management Functions 20
1.1.10 VPN Extended Feature of OSPF 21
1.1.11 Protocols and Standards 24
1.2 Configuration Task Summary 25
1.3 Configuring Basic Features 28
1.3.1 Overview 28
1.3.2 Restrictions and Guidelines 28
1.3.3 Configuration Tasks 28
1.3.4 Configuring OSPF Process 28
1.3.5 Configuring Router ID 29
1.3.6 Enabling OSPF on a Network Segment 30

1.3.7 Enabling OSPF on an Interface 30
1.3.8 Creating a Virtual Link 32
1.4 Configuring OSPF Network Type 33
1.4.1 Overview 33
1.4.2 Restrictions and Guidelines 33
1.4.3 Configuration Tasks 33
1.4.4 Configuring the Interface Network Type 33
1.4.5 Configuring a Neighbor 35
1.4.6 Configuring the Interface Priority 36
1.4.7 Disabling Source Address Verification 37
1.5 Configuring Stub and NSSAs 38
1.5.1 Overview 38
1.5.2 Restrictions and Guidelines 38
1.5.3 Configuration Tasks 38
1.5.4 Configuring a Stub Area 38
1.5.5 Configuring an NSSA 39
1.6 Adjusting OSPF Routing 40
1.6.1 Overview 40
1.6.2 Restrictions and Guidelines 40
1.6.3 Configuration Tasks 40
1.6.4 Configuring the AD 41
1.6.5 Configuring the Reference Bandwidth 42
1.6.6 Configuring the Cost of an Interface 42

1.6.7 Configuring Cost Fallback Value for an AP....44
1.6.8 Configuring the Cost of the Default Route in a Stub Area or an NSSA....44
1.6.9 Configuring the Default Metric Value for Redistribution....45
1.6.10 Configuring the Maximum Metric....45
1.6.11 Configuring RFC 1583 Compatibility....47
1.6.12 Configuring Association of External Route Forwarding Addresses with Interface Configuration....47
1.7 Configuring OSPF Route Advertisement....48
1.7.1 Overview....48
1.7.2 Restrictions and Guidelines....49
1.7.3 Configuration Tasks....49
1.7.4 Configuring External Route Redistribution....49
1.7.5 Configuring VRF-based External Route Redistribution....50
1.7.6 Generating a Default Route....51
1.7.7 Configuring a Device as ASBR....52
1.8 Configuring OSPF Route Summarization....52
1.8.1 Overview....52
1.8.2 Restrictions and Guidelines....52
1.8.3 Configuration Tasks....53
1.8.4 Configuring Inter-Area Route Summarization....53
1.8.5 Configuring External Route Summarization....54
1.8.6 Configuring a Discard Route....55
1.9 Configuring OSPF Route Filtering....55

1.9.1 Overview 55
1.9.2 Restrictions and Guidelines 55
1.9.3 Configuration Tasks 56
1.9.4 Configuring a Passive Interface 56
1.9.5 Disabling Interfaces from Sending OSPF Packets 56
1.9.6 Configuring Inter-area Route Filtering 57
1.9.7 Configuring Redistributed Route Filtering 58
1.9.8 Configuring Learned Route Filtering 59
1.10 Adjusting OSPF Network Convergence Speed 60
1.10.1 Overview 60
1.10.2 Restrictions and Guidelines 60
1.10.3 Configuration Tasks 60
1.10.4 Configuring the Hello Interval 61
1.10.5 Configuring the Dead Interval 62
1.10.6 Configuring LSU Transmission Delay 63
1.10.7 Configuring LSU Retransmission Interval 64
1.10.8 Configuring LSU Generation Time 65
1.10.9 Configuring LSA Group Refresh Time 66
1.10.10 Configuring LSA Group Refresh Interval 67
1.10.11 Configuring Duplicate LSA Receiving Delay 67
1.10.12 Configuring Inter-Area Route Computation Delay 68
1.10.13 Configuring External Route Computation Delay 68
1.10.14 Configuring SPF Computation Delay 69

1.10.15 Configuring Route Calculation Suppression Time....70
1.10.16 Enabling Two-Way Maintenance....71
1.11 Configuring OSPF Authentication....71
1.11.1 Overview....71
1.11.2 Restrictions and Guidelines....71
1.11.3 Configuration Tasks....71
1.11.4 Configuring the Authentication Type of an Area....72
1.11.5 Configuring the Authentication Type of an Interface....73
1.11.6 Configuring a Plain Text Authentication Key for an Interface....74
1.11.7 Configuring an MD5 Authentication Key for an Interface....75
1.12 Enabling Fast Reroute....76
1.12.1 Overview....76
1.12.2 Restrictions and Guidelines....76
1.12.3 Configuration Tasks....77
1.12.4 Configuring Fast Reroute....77
1.12.5 Configuring Interface LFA Protection....78
1.12.6 Preventing an Interface From Becoming a Backup Interface....79
1.13 Modifying the Maximum Number of Concurrent Neighbors....80
1.13.1 Overview....80
1.13.2 Restrictions and Guidelines....80
1.13.3 Configuration Tasks....80
1.13.4 Configuring the Maximum Number of Concurrent Neighbors on the Current Process80
1.13.5 Configuring the Maximum Number of Concurrent Neighbors on All Processes....81

1.14 Configuring the GR Function 81
1.14.1 Overview 81
1.14.2 Restrictions and Guidelines 82
1.14.3 Configuration Tasks 82
1.14.4 Enabling the GR Function 82
1.14.5 Configuring the OSPF GR Helper Function 83
1.15 Correlating OSPF with BFD 84
1.15.1 Overview 84
1.15.2 Restrictions and Guidelines 84
1.15.3 Configuration Tasks 84
1.15.4 Configuring BFD Correlation with OSPF on an Interface 84
1.15.5 Configuring BFD Correlation with OSPF Globally 85
1.16 Enabling Overflow 86
1.16.1 Overview 86
1.16.2 Restrictions and Guidelines 86
1.16.3 Configuration Tasks 86
1.16.4 Configuring the Memory Overflow Function 86
1.16.5 Configuring the Database Overflow Function 87
1.16.6 Configuring the External LSA Database Overflow Function 88
1.17 Configuring VPN Extended Feature of OSPF 88
1.17.1 Overview 88
1.17.2 Restrictions and Guidelines 89
1.17.3 Configuration Tasks 89

1.17.4 Configuring Domain ID....89
1.17.5 Configuring a Domain Label....90
1.17.6 Configuring a Sham Link....91
1.17.7 Disabling Loop Detection for a VRF OSPF Process....92
1.17.8 Configuring Extcommunity Attribute for a VPN Route....93
1.17.9 Disabling Loop Detection Using the DN Bit of LSA....93
1.17.10 Disabling Loop Detection Using the Route Tag of LSA....94
1.18 Configuring Network Management Functions....95
1.18.1 Overview....95
1.18.2 Restrictions and Guidelines....95
1.18.3 Configuration Tasks....95
1.18.4 Binding the MIB with an OSPF Process....95
1.18.5 Configuring the Trap Function....96
1.18.6 Configuring the Logging Function....97
1.19 Enabling NSR....97
1.19.1 Overview....97
1.19.2 Restrictions and Guidelines....97
1.19.3 Procedure....98
1.20 Enabling the iSPF Feature....98
1.20.1 Overview....98
1.20.2 Restrictions and Guidelines....98
1.20.3 Procedure....98
1.21 Configuring to Disable MTU Verification....99

1.21.1 Overview 99
1.21.2 Restrictions and Guidelines 99
1.21.3 Procedure 99
1.22 Configuring to Enable Conflict Detection of Type-4 LSAs 100
1.22.1 Overview 100
1.22.2 Restrictions and Guidelines 100
1.22.3 Procedure 100
1.23 Enabling the Function of Securely Promoting a BDR to DR 100
1.23.1 Overview 100
1.23.2 Restrictions and Guidelines 101
1.23.3 Procedure 101
1.24 Configuring a Super VLAN to Enable OSPF 101
1.24.1 Overview 101
1.24.2 Restrictions and Guidelines 101
1.24.3 Procedure 101
1.25 Monitoring 102
1.26 Configuration Examples 104
1.26.1 Configuring Basic Functions of OSPF 104
1.26.2 Configuring a Stub Area 109
1.26.3 Configuring an NSSA 111
1.26.4 Correlating OSPF Process with BFD 114
1.26.5 Configuring Fast Reroute 116
1.26.6 Configuring Network Management Functions 118

1.26.7 Configuring the Sham Link Function 120

1.26.8 Configuring a Virtual Link..... 131

1.26.9 Selecting a DR of OSPF 133

1 Configuring OSPFv2

1.1 Introduction

1.1.1 Overview

The open shortest path first (OSPF) is an Interior Gateway Protocol (IGP) that is used within the autonomous system (AS) to allow routers to obtain a route to a remote network.

In the early stage after dynamic routing emerges, Routing Information Protocol (RIP) was mainly used on the network. RIP is a distance vector routing protocol, facing some problems such as slow convergence, routing loop, poor scalability, and use for simple networks only. OSPF Version 2 (OSPFv2) is a link-state routing protocol, which can effectively fix the above problems. Therefore, OSPFv2 is widely used on large and medium-sized networks.

Note

- OSPFv2 is applicable to IPv4, and OSPF Version 3 (OSPFv3) is applicable to IPv6. The protocol running mechanism and most configurations are the same.
- Unless otherwise specified, "OSPF" in the following descriptions refers to OSPF Version 2.

OSPF has the following characteristics:

- Wide scope of application: OSPF is applicable to a larger-scale network that supports hundreds of routers.
- Fast convergence: Once the network topology changes, advertisements can be quickly sent between routers to update routes.
- No self-loop: Only the link status information is synchronized between routers. Each router computes routes independently, and a self-loop will not occur.
- Area division: A large routing domain is divided into multiple small areas to save system resources and network bandwidth and ensure stability and reliability of routes.
- Route classification: Routes are classified into several types to support flexible control.
- Equivalent routes: OSPF supports equivalent routes.
- Authentication: OSPF supports packet authentication to ensure security of protocol interaction.
- Multicast transmission: Protocol packets are sent using the multicast address to avoid interfering with irrelevant entities and save system resources.

Note

In this chapter, the term "router" refers to any network device that supports the routing function. These network devices can be Layer-3 switches, routers, or firewalls.

1.1.2 Basic Concepts

1. Routing domain

All routers in an AS must be interconnected and use the same routing protocol. Therefore, an AS is also called a routing domain.

An AS on which OSPF runs is also called OSPF routing domain, or OSPF domain for short.

2. OSPF process

OSPF supports multiple instances, and each instance corresponds to an OSPF process. One or more OSPF processes can be started on a router. Each OSPF process runs OSPF independently, and the processes are mutually isolated. The process ID takes effect only on the local router, and does not affect exchange of OSPF packets on adjacent interfaces.

3. Router ID

The router ID uniquely identifies a router in an OSPF domain. Router IDs of any two routers cannot be the same.

If multiple OSPF processes exist on a router, each OSPF process uses one router ID. Router IDs of any two OSPF processes cannot be the same.

A router ID can be manually configured or automatically selected by the router. To prevent conflicts between Router IDs obtained by different routers through automatic selection, you are advised to select manual configuration. If no router ID is specified manually, the automatic selection rules are as follows:

(1) The largest IP address from the lookback interface addresses is selected as the router ID;

(2) If no lookback interface is configured, the largest IP address from the physical interface addresses is selected as the router ID.

4. Area

An area is a logical group of routers, and each group is identified by an area ID. As shown in Figure 1-1, OSPF supports dividing the network topology into multiple areas, which can alleviate the computing pressure of large-scale networks.

The border between areas is a router. A router may belong to one area or multiple areas. One network segment (link) can belong to only one area, or each OSPF-enabled interface must belong to a specified area.

OSPF areas can be classified into following types:

- Backbone area and normal area

  Area 0 is the backbone area, and other areas are normal areas. Normal areas must be directly connected to the backbone area.

- Stub area (OSPF stubby area)

  Configuring a stub area can reduce the number of LSAs in the area. The area border router (ABR) in the stub area will not transfer the Type-5 LSA (external route), and will advertise a Type 3 default route to the stub area.

- Totally stub area

A totally stub area is upgraded from a stub area. The router in the totally stub area will not transfer Type-3 LSA (inter-domain route) and Type-5 LSA (external route). ABR only advertises a Type 3 default route to the totally stub area.

- Not-so-stubby area (NSSA)

  An NSSA is similar to a stub area, but allows the autonomous system boundary router (ASBR). The route redistributed by ASBR will be transferred in the NSSA in the form of Type-7 LSA, and the Type-7 LSA will be converted into Type-5 LSA on ABR and then transferred to other areas. The ABR also advertises one default route to the NSSA.

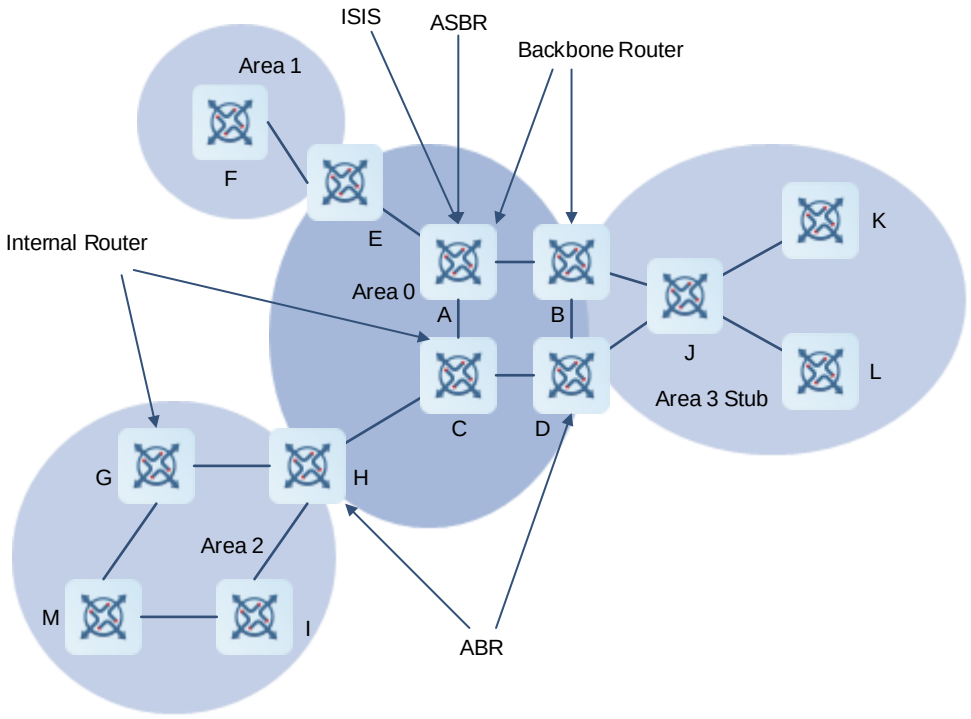
- Totally not-so-stubby area (NSSA)

  The relationship between the totally NSSA and NSSA is similar to that between the totally stub area and stub area. The totally NSSA forbids transfer of the Type-3 LSA (inter-domain route) on the basis of NSSA.

Note

- The backbone area cannot be configured as a stub area, totally stub area, NSSA, or totally NSSA.
- The stub area and totally stub area cannot contain any ASBR.
- The stub area, totally stub area, NSSA, and totally NSSA cannot contain any virtual link.

Figure 1-1 Types of OSPF Areas and Routers



5. Types of OSPF routers

The following types of routers are defined in OSPF, and assigned with different responsibilities:

- Internal router

  All interfaces of an interval router belong to the same area. See A, C, F, G, I, M, J, K, and L shown in Figure 1-1.

- ABR

  An ABR is used to connect the backbone area with a normal area. An ABR belongs to two or more areas, and one of the areas must be the backbone area. See B, D, E, and H shown in Figure 1-1.

- Backbone router

  A backbone router has at least one interface that belongs to the backbone area. All ABRs and all routers in Area 0 are backbone routers. See A, B, C, D, E, and H shown in Figure 1-1.

- ASBR

  An ASBR is used to exchange routing information with other ASs. An ASBR is not necessarily located on the border of an AS. It may be a router inside an area, or an ABR. See A shown in Figure 1-1.

6. LSA

OSPF encapsulates and advertises the description of routing information through a Link State Advertisement (LSA). Common LSA types are shown in Table 1-1.

Table 1-1 OSPF LSA Types

LSA Type	Description
Router-LSA (Type 1)	This LSA is originated by each router. It describes the link state and cost of the router, and is advertised only within the area where the originating router is located.
Network-LSA (Type 2)	This LSA is originated by a DR on the non-broadcast multiple access (NBMA) network. It describes the link state in the current network segment, and is advertised only within the area where the DR is located.
Network-summary-LSA (Type 3)	This LSA is originated by an ABR. It describes a route to another area, and is advertised to areas except totally stub areas and NSSAs.
ASBR-summary-LSA (Type 4)	This LSA is originated by an ABR. It describes a route to an ASBR, and is advertised to areas except the area where the ASBR is located.
AS-external-LSA (Type 5)	This LSA is originated by an ASBR. It describes a route to a destination outside the AS, and is advertised to all areas except stub areas and NSSAs.
NSSA LSA (Type 7)	This LSA is originated by an ASBR. It describes a route to a destination outside the AS, and is advertised only within the NSSA.

LSA Type	Description
Opaque LSA (Type 9/Type 10/Type 11)	Opaque LSAs provide a generalized mechanism to allow for the future extensibility of OSPF, wherein: ● Type-9 LSAs are advertised only within the network segment where interfaces are located. The grace LSA used to support graceful restart (GR) is one of Type-9 LSAs. ● Type-10 LSAs are advertised within an area. The LSA used to support traffic engineering (TE) is one of Type-10 LSAs. ● Type-11 LSAs are advertised within an AS. At present, there are no application examples of Type-11 LSAs.

The format of LSA packet header is described as follows.

Figure 1-2 LSA Packet Header Format

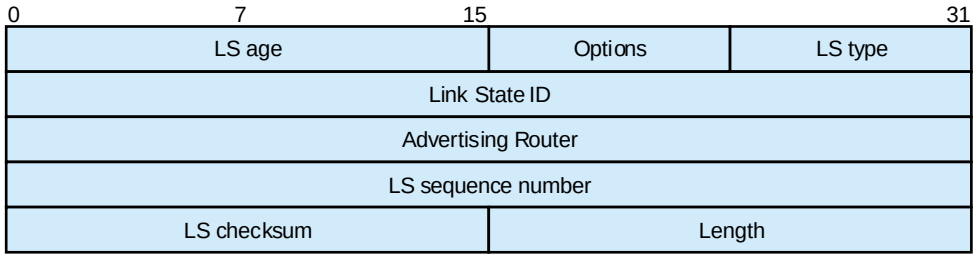


Table 1-2 Description of Fields of LSA Packet Header

Field	Description
LS age	The age of the LSA.
Options	Options of the LSA.
LS type	LSA type, as shown in Table 1-1.
Link State ID	Describes the LSA information. The content varies with the LS type.
Advertising Router	Advertises the router ID of the LSA.
LS sequence number	Serial number of the LSA. The larger the serial number is, the newer the LSA is.
LS checksum	Checksum of other parts except the LS age field.
Length	Total length of the LSA including the LSA Header, in bytes.

Table 1-3 shows the support of various areas for different types of LSAs. Stub area, NSSA, Totally Stub area, and Totally NSSA, as special forms of conventional areas, help reduce the number of LSAs transferred in the OSPF area, eliminate the propagation of a large number of Type 3 LSAs and Type 5 LSAs in the area, reduce the burden on routers, and enhance the stability of OSPF routing.

Table 1-3 Support of LSAs by Different Areas

Area	Type1 and Type 2	Type 3	Type 4	Type 5	Type 7
Non-Stub/Totally Stub and NSSA/Totally NSSAs	Allowed	Allowed	Allowed	Allowed	Blocked
Stub Area	Allowed	Allowed, including one default route	Blocked	Blocked	Blocked
Totally Stub Area	Allowed	Only one default route is allowed	Blocked	Blocked	Blocked
NSSA	Allowed	Allowed, including one default route	Allowed	Blocked	Allowed
Totally NSSA	Allowed	Only one default route is allowed	Allowed	Blocked	Allowed

Note

ABR advertises a default route to Stub/Totally Stub and NSSA/Totally NSSAs using a Type 3 LSA.

ABR converts LSA Type 7 to LSA Type 5 in NSSA/Totally NSSA, and advertise the route to the backbone area.

7. OSPF Packets

OSPF packets are encapsulated in IP packets and transmitted in multicast or unicast mode. When the interfaces running the OSPF process on any network belong to this group, the router running the OSPF process receives all the multicast packets of 224.0.0.5; when the router is used as a DR/BDR on a multi-access network, 224.0.0.6 is used as the multicast receiving address. The OSPF data packet format is shown as follows:

OSPF involves the following five types of packets, as shown in Table 1-4.

Table 1-4 Types of OSPF Packets

Packet Type	Description
Hello	Hello packets are sent periodically to discover and maintain OSPF neighbor relationships.

Packet Type	Description
DD	DD (Database Description) packets carry brief information about the local LSDB and are used to synchronize the LSDBs between OSPF neighbors.
LSR	LSR (Link State Request) packets are used to request the required LSAs from neighbors. LSR packets are sent only after DD packets are exchanged successfully between OSPF neighbors.
LSU	LSU (Link State Update) packets are used to send the required LSAs to peers.
LSAck	LSAck (Link State Acknowledgment) packets are used to acknowledge the received LSAs.

Format of OSPF packets is shown in the figure below.

Figure 1-3 Format of OSPF Packets

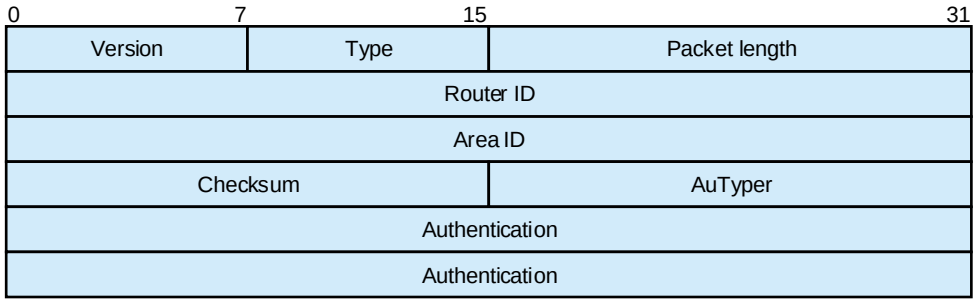


Table 1-5 Description of OSPF Packet Format Fields

Packet Field	Description
Version	Indicates the OSPF version No. The value of OSPFv2 is 2.
Type	Indicates the type of OSPF packet: 1: Indicates hello packet. 2: Indicates DD packet. 3: Indicates LSR packet. 4: Indicates LSU packet. 5: Indicates LSAck packet.
Packet length	Indicates the total length of the OSPF packet including the header, in bytes.
Router ID	Indicates the ID of the router sending the packet.
Area ID	Indicates the area of the router sending the packet.

Packet Field	Description
Checksum	Indicates the checksum of the whole packet excluding the authentication field.
AuType	Indicates the authentication type: 0: Indicates no authentication. 1: Indicates simple authentication. 2: Indicates MD5 authentication.
Authentication	The meaning of this field varies with the authentication type: No authentication: This field is not defined. Simple authentication: This field indicates the password. MD5 authentication: This field includes the key ID, MD5 authentication data length, and sequence number information.

8. OSPF route types

A mark is displayed in front of each OSPF route to indicate the type of the route. There are six types of OSPF routes:

- O: Internal route

  This type of route describes how to arrive at a destination network in the local area. The cost of this type of route is equal to the cost of the route from the local router to the destination network.

- IA: Inter-area route

  This type of route describes how to arrive at a destination network in another area. The cost of this type of route is equal to the cost of the route from the local router to the destination network.

- E1: Type-1 external route

  This type of route describes how to arrive at a destination network outside the AS. The cost of this type of route is equal to the cost of the route from the local router to the ASBR plus the cost of the route from the ASBR to the destination network. This type of route does not exist on routers in the stub or NSSA.

- E2: Type-2 external route

  This type of route describes how to arrive at a destination network outside the AS. The cost of this type of route is equal to the cost of the route from the ASBR to the destination network. This type of route does not exist on routers in the stub or NSSA.

- N1: Type-1 external route of the NSSA

  This type of route describes how to arrive at a destination network outside the AS through the ASBR in the NSSA. The cost of this type of route is equal to the cost of the route from the local router to the ASBR plus the cost of the route from the ASBR to the destination network. This type of route exists only on routers in the NSSA.

- N2: Type-2 external route of the NSSA

This type of route describes how to arrive at a destination network outside the AS through the ASBR in the NSSA. The cost of this type of route is equal to the cost of the route from the ASBR to the destination network. This type of route exists only on routers in the NSSA.

Note

Reliability of E2 and N2 routes is poor. OSPF believes that the cost of the route from the ASBR to a destination outside an AS is far greater than the cost of the route to the ASBR within the AS. Therefore, when the route cost is computed, only the cost of the route from the ASBR to a destination outside an AS is considered.

Configuring an area as a Stub/Totally Stub area or NSSA/Totally NSSA can reduce a large number of E1, E2, and IA routes in the routing table of routers in the area.

Table 1-6 Route Types in a Stub/NSSA

Area	Routes in the routing table of routers in the area
Non-Stub/Totally Stub, and NSSA/Totally NSSA	O: Route to a destination network in the local area. IA: Route to a destination network in another area. E1 or E2: Route or default route to a destination network outside the AS through any ASBR in the AS.
Stub area	O: Route to a destination network in the local area. IA: Route to a destination network in another area.
Totally Stub area	O: Route to a destination network in the local area. IA: Default route.
NSSA	O: Route to a destination network in the local area. IA: Route to a destination network in another area. N1or N2: Route or default route to a destination network outside the AS through any ASBR in the area.
Totally NSSA	O: Route to a destination network in the local area. IA: Default route. N1or N2: Route or default route to a destination network outside the AS through any ASBR in the area.

9. OSPF neighbor state

On the OSPF network, neighbor devices need to reach the adjacency state before exchanging link information. Any two devices form an OSPF neighborhood relation or adjacency relation.

- Neighborhood: If two routers can send and receive Hello packets from each other, and can successfully negotiate the parameters carried in the Hello packets, then they are neighbors to each other.
- Adjacency: When neighbor devices successfully exchange LSA and realize LSDB synchronization, then an adjacency relation is successfully established.

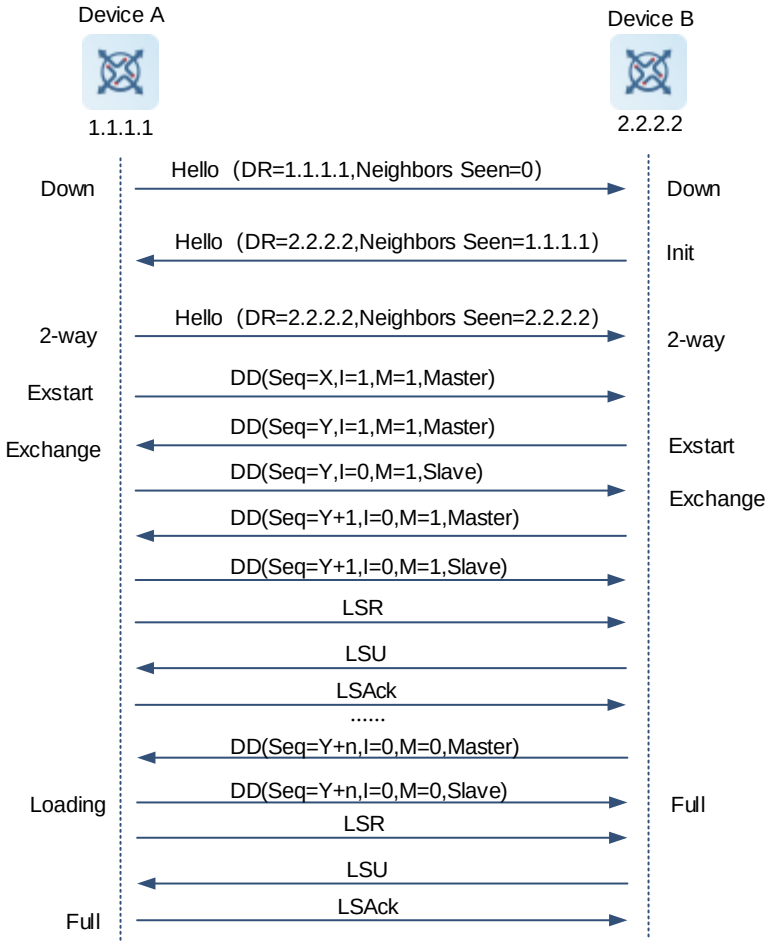
Table 1-7 shows the 8 OSPF neighbor states.

Table 1-7 OSPF Neighbor State Machine and Description

State	Description
Down	The state of the First OSPF neighbor, which indicates that the neighbor's hello packet is not received within the neighbor dead interval.
Attempt	The attempt state router sends a hello packet to the manually configured neighbor periodically. The attempt state applies to the interfaces of NBMA type only.
Init	A hello packet has been received from the neighbor, but the packet does not contain the router ID of the neighbor receiving router, that is, the peer end does not receive the hello packet sent by the local end.
Two-Way	Indicates mutual neighbors. ● When the neighbor router ID field in the received hello packet is each other's router ID, and both ends receive the hello packet sent from the peer end, a neighbor relationship is established. ● After this phase ends, DR and BDR will be selected for broadcast and non-broadcast multi-access.
Exstart	Indicates negotiation about the master/slave relationship. ● Selects the initial sequence number for forming an adjacency. ● The master/slave relationship ensures orderly transmission in the subsequent exchange of DD packets.
Exchange	Indicates exchange of DD packets. Checks whether the neighbor can provide new or updated link state information.
Loading	Indicates that, based on the information provided by DBD, the router will send LSR and interactive LSU to implement synchronization with LSDB.
Full	Indicates establishment of an adjacency. Indicates that the LSDBs of the two devices have been synchronized, and the adjacency state has been established between the local device and the neighbor device.

Figure 1-4 shows the OSPF adjacency establishment process on the broadcast network. When the neighbor state is Full, the adjacency is established.

Figure 1-4 OSPF Adjacency Establishment Process on Broadcast Network



10. DR and BDR

On the broadcast or NBMA network, many unwanted LSAs will be created when an adjacency is set up with related routers. If there are n routers, n*(n-1)/2 adjacency relationships will be set up, and n^2 LSAs will be generated on the network. In the process of setting up the adjacency and exchanging LSAs, many unwanted copies will be produced, which leads to chaos in the flooding of a multi-access network. To avoid these problems on a multi-access network, OSPF defines the concepts of DR and BDR.

After the DR and BDR are elected, all routers will set up an adjacency relationship with the DR and BDR only, send information to the DR/BDR, and the DR will broadcast the network link state. The routers other than DR and BDR are called DR Other. DR Other will no longer set up an adjacency relationship or exchange any routing information with each other, thus reducing the number of adjacency relationships between routers on the broadcast and NBMA networks, as shown Figure 1-5 and Figure 1-6.

Figure 1-5 Number of Adjacency Relationships Before DR Election

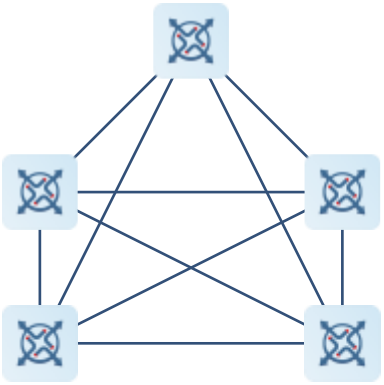
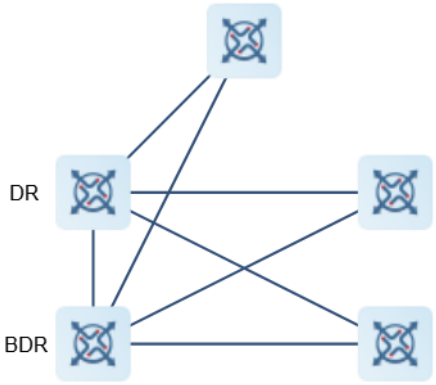


Figure 1-6 Number of Adjacency Relationships After DR Election



11. OSPF network types

A router does not necessarily need to exchange LSAs with every neighbor or set up an adjacency with every neighbor. To improve efficiency, OSPF classifies networks that use various link layer protocols into five types so that LSAs are exchanged in different ways to set up an adjacency.

By default, Ethernet and fiber distributed data interface (FDDI) belong to the broadcast type, X.25, frame relay, and ATM belong to the NBMA type, and PPP, HDLC, and LAPB belong to the P2P type.

Table 1-8 Introduction to OSPF Network Types

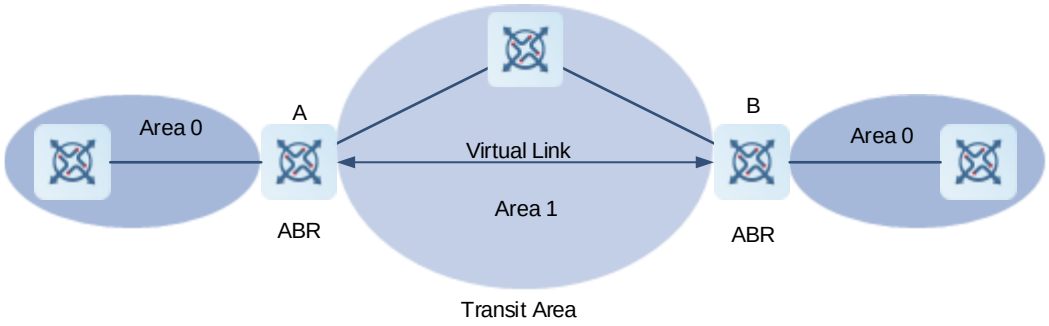
Network Type	Link Layer Protocol	Neighbor Relationship and DR Election
Broadcast	Ethernet and FDDI belong to the broadcast network type by default.	● Neighbors are discovered, and the DR and BDR are elected. ● The DR or BDR exchanges LSAs with all other routers to set up an adjacency. ● Except the DR and BDR, all other routers do not exchange LSAs with each other, and the adjacency is not set up.
NBMA	X.25, frame relay, and ATM belong to NBMA networks by default.	● Neighbors are manually configured, and the DR and BDR are elected. ● The DR or BDR exchanges LSAs with all other routers to set up an adjacency. ● Except the DR and BDR, all other routers do not exchange LSAs with each other, and the adjacency is not set up.
Point-to-Point (P2P)	PPP, HDLC, and LAPB belong to the P2P network type by default.	● Neighbors are automatically discovered, and the DR or BDR is not elected. ● LSAs are exchanged between routers at both ends of the link, and the adjacency is set up.
Point-to-Multipoint (P2MP)	Networks without any link layer protocol belong to the P2MP network type by default.	● Neighbors are automatically discovered, and the DR or BDR is not elected. ● LSAs are exchanged between any two routers, and the adjacency is set up.
P2MP broadcast	Networks without any link layer protocol belong to the P2MP network type by default.	● Neighbors are manually configured, and the DR or BDR is not elected. ● LSAs are exchanged between any two routers, and the adjacency is set up.

12. Virtual link

OSPF supports virtual links. A virtual link is a logical link that belongs to the backbone area. It is used to resolve the problems such as a discontinuous backbone area or a failure to directly connect a normal area to the backbone area on the physical network. A virtual link supports traversal of only one normal area, and this area is called transit area. Routers on both ends of a virtual link are ABRs.

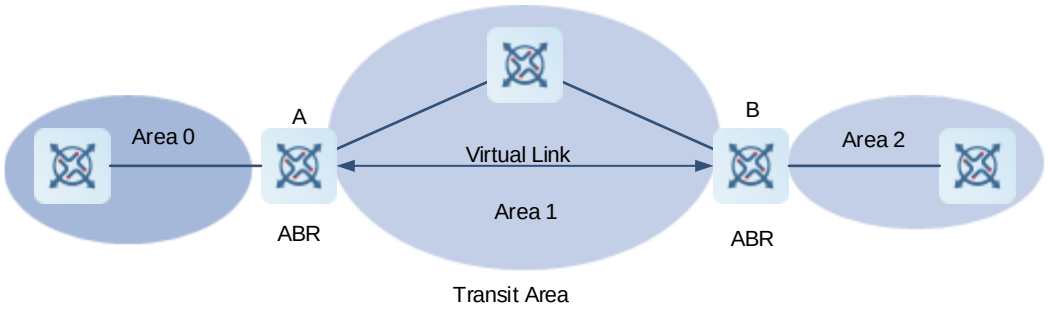
A virtual link relies on an OSPF intra-domain route, making two ABRs adjacent. The OSPF packet information exchanged by ABRs is transparent to the intermediate device.

Figure 1-7 Discontinuous Backbone Area on the Physical Network



As shown in Figure 1-7, a virtual link is set up between A and B to connect to Area 0. Area 1 is a transit area, and A and B are ABRs of Area 1.

Figure 1-8 Failure to Directly Connect a Normal Area to the Backbone Area on the Physical Network



As shown in Figure 1-8, a virtual link is set up between A and B to extend Area 0 to B so that Area 0 can be directly connected to Area 2 on B. Area 1 is a transit area, A is an ABR of Area 1, and B is an ABR of Area 1 and Area 2.

1.1.3 Principles

OSPF is a type of link-state routing protocols. Its working process consists of three phases:

(1) Neighbor discovery → Bidirectional communication

An OSPF neighbor relationship is set up between adjacent routers, and bidirectional communication is maintained.

Routers send hello packets through all OSPF-enabled interfaces or virtual links. If hello packets can be exchanged between two routers, and parameters carried in the hello packets can be successfully negotiated, the two routers become neighbors. Routers that are mutual neighbors find their own router IDs from hello packets sent from neighbors, and bidirectional communication is set up.

A hello packet includes, but is not limited to, the following information:

- Router ID of the originating router.
- Area ID, subnet mask, authentication information, Hello interval and neighbor dead interval of the originating

router interface or virtual link.

- IP addresses of the designated router (DR) and backup designated router (BDR).
- Priority of the originating router interface (used for DR/BDR election).
- Router ID of the neighbor of the originating router.

(2) Database synchronization → Full adjacency

A router uses link-state advertisements (LSAs) to advertise all its link states. LSAs are exchanged between neighbors and the link state database (LSDB) is synchronized to achieve full adjacency.

After bidirectional communication is established between neighbor routers, database description (DD), link state request (LSR), link state update (LSU), and link state acknowledgment (LSAck) packets are used to exchange LSAs and establish an adjacency relationship. The brief process is as follows:

- A router generates an LSA to describe all link states on the router.
- The LSA is exchanged between neighbors. When a router receives the LSA from its neighbor, it copies the LSA and saves the copy in the local LSDB, and then advertises the LSA to other neighbors.
- When the router and its neighbors obtain the same LSDB, full adjacency is achieved.

Note

OSPF will not generate LSDB updates and advertisements without changes in link costs or network addition/deletion. If any change takes place, the changed link states are advertised to quickly synchronize the LSDB.

(3) Shortest path tree (SPT) computation → Formation of a routing table

The router computes the shortest path to each destination network based on the LSDB and forms an OSPF routing table.

After the complete LSDB is obtained from the router, the Dijkstra algorithm is run to generate an SPT from the local router to each destination network. The SPT records the destination networks, next-hop addresses, and costs. OSPF generates a routing table based on the SPT.

If changes in link costs or network addition/deletion take place, the LSDB will be updated. The router again runs the Dijkstra algorithm, generates a new SPT, and updates the routing table.

Note

The Dijkstra algorithm is used to find the shortest path from a vertex to other vertices in a weighted directed graph.

1.1.4 OSPF Routing

- Route cost

  If redundancy links or devices exist on the network, multiple paths may exist from the local device to the destination network. OSPF selects the path with the minimum total cost to form an OSPF route. The total cost of a path is equal to the sum of the costs of individual links along the path. The total cost of a path can be

minimized by modifying the costs of individual links along the path. In this way, OSPF selects this path to form a route.

With configuration commands, you can modify the following link costs:

- Cost from an interface to a directly connected network segment and cost from the interface to a neighbor.
- Cost from an ABR to the inter-area summarization network segment and cost from the ABR to the default network segment.
- Cost from an ASBR to an external network segment and cost from the ASBR to the default network segment.

Note

Cost value refers to the cost of a link and the link quality value after calculation. Metric is the cost of a complete path and the total cost of the links passed.

- OSPF administrative distance

  The administrative distance (AD) evaluates reliability of a route. Its value is an integer ranging from 0 to 255. A smaller AD value indicates that the route is more trustworthy. If multiples exist to the same destination, the route preferentially selects a route with a smaller AD value. The route with a greater AD value becomes a floating route, that is, a standby route of the optimum route.

  By default, the route coming from one source corresponds to an AD value. The AD value is a local concept. Modifying the AD value affects route selection only on the current router.

Table 1-9 Routing Protocols and Default Administrative distances

Route Source	Directly-connected network	Static route	EBGP route	OSPF route	IS-IS route	RIP route	IBGP route	Unreachable route
Default AD	0	1	20	110	115	120	200	255

1.1.5 OSPF Route Advertisement

- Route redistribution

  Route redistribution refers to the process of introducing routes of other routing protocols, routes of other OSPF processes, static routes, and direct routes that exist on the device to an OSPF process so that these routes can be advertised to neighbors using Type-5 and Type-7 LSAs. A default route cannot be introduced during route redistribution.

  Route redistribution is often used for interworking between ASs. You can configure route redistribution on an ASBR to advertise routes outside an AS to the interior of the AS, or routes inside an AS to the exterior of the AS.

- Default route introduction

By configuring a command on an ASBR, you can introduce a default route to an OSPF process so that the route can be advertised to neighbors using Type 5 and Type 7 LSAs.

Default route introduction is often used for interworking between ASs. One default route is used to replace all the routes outside an AS.

1.1.6 OSPF Route Summarization

Route summarization is a process of summarizing routing information with the same prefix into one route, and advertising the summarized route to neighbors by replacing a large number of individual routes. Route summarization helps reduce the protocol interaction load and the size of the routing table.

By default, the ABR advertises inter-area routing information by using Type-3 LSAs within a network segment, and advertises redistributed routing information by using Type-5 and Type-7 LSAs. If continuous network segments exist, you are advised to configure route summarization.

When configuring route summarization, the summarization range may exceed the actual network scope of routes. If the data is sent to a network beyond the summarization range, a routing loop may be formed and the router processing load may increase. To prevent these problems, ABR or ASBR will automatically add a discard route to the routing table. This route will not be advertised.

1.1.7 OSPF Route Filtering

OSPF allows filtering of the learned and exchanged routes and LSAs so as to meet the security requirements in specific scenarios, for example, the routes of some network segments are not intended to be learned by other areas.

With configuration commands, you can configure route filtering for the following items:

- Interface: The interface is prevented from sending any LSAs or exchanging any LSAs with neighbors.
- Routing information advertised between areas: Only the Type-3 LSA that meets the filtering conditions can be advertised to another area.
- Routing information outside an AS: Only the Type-5 and Type-7 LSAs that meet the filtering conditions can be redistributed to the OSPF process.
- LSAs received by a router: In the OSPF routing table, only the routes that are computed based on the LSAs meeting the filtering conditions can be advertised.

1.1.8 Enhanced Security and Reliability

1. Authentication

Authentication prevents routers that illegally access the network and hosts that forge OSPF packet from participating in the OSPF process. OSPF packets received on the OSPF interface or at both ends of the virtual link are authenticated. If authentication fails, the packets are discarded and no adjacency can be set up.

Enabling authentication can avoid learning unauthenticated or invalid routes, thus preventing advertising valid routes to unauthenticated devices. On the broadcast-type network, authentication also prevents unauthenticated devices from becoming designated devices, ensuring stability of the routing system and protecting the routing system against intrusions.

Authentication is classified into following two types:

- Area authentication: This area-level authentication authenticates the packets transmitted by all the interfaces in the area. This type is configured in OSPF routing process mode.
- Interface authentication: This neighbor-level authentication authenticates the packets transmitted on the related interface. This type is configured in interface mode.

2. MTU verification

Upon receiving a DD packet, OSPF checks whether the MTU of the neighbor interface is the same as that of the local interface. If the MTU of the interface specified in the received DD packet is greater than that of the interface that receives the packet, the adjacency cannot be set up. Disabling MTU verification can avoid this problem.

3. Source address verification

Generally, the source address of a packet received by OSPF is in the same network segment as the receiving interface. The addresses at both ends of a P2P link are configured separately and are not necessarily in the same network segment. In this scenario, as the peer address information will be notified during the P2P link negotiation process, OSPF checks whether the source address of the packet is the address advertised by the peer during negotiation. If not, OSPF determines that the packet is invalid and discards this packet. In particular, OSPF never verifies the address of an interface not configured with an IP address.

In some scenarios, the source address of a packet received by OSPF may not be in the same network segment as the receiving interface, so the OSPF address verification fails. For example, the negotiated peer address cannot be obtained on a P2P link. In this scenario, source address verification must be disabled to ensure that the OSPF adjacency can be properly set up.

4. Two-way maintenance

OSPF routers periodically send hello packets to each other to maintain the adjacency. On a large network, a lot of packets may be sent or received, occupying a great proportion of CPU and memory. As a result, some packets are delayed or discarded. If the processing time of hello packets exceeds the neighbor dead interval, the adjacency will be destroyed.

If the two-way maintenance function is enabled, in addition to the hello packets, the DD, LSU, LSR, and LSAck packets can also be used to maintain the bidirectional communication between neighbors, which makes the adjacency more stable.

5. Concurrent neighbor interaction restriction

When a router simultaneously exchanges data with multiple neighbors, its performance may be affected. If the maximum number of neighbors that concurrently initiate or accept interaction with the OSPF process is restricted, the router can interact with neighbors by batches, which ensures data forwarding and other key services.

6. Overflow

OSPF requires that routers in the same area store the same LSDB. The number of routers keeps increasing on the network. Some routers, however, cannot store so much routing information due to the limited system resources. The large amount of routing information may exhaust the system resources of routers, causing failures of the routers.

The overflow function limits the number of external routes in the LSDB to control the size of the LSDB.

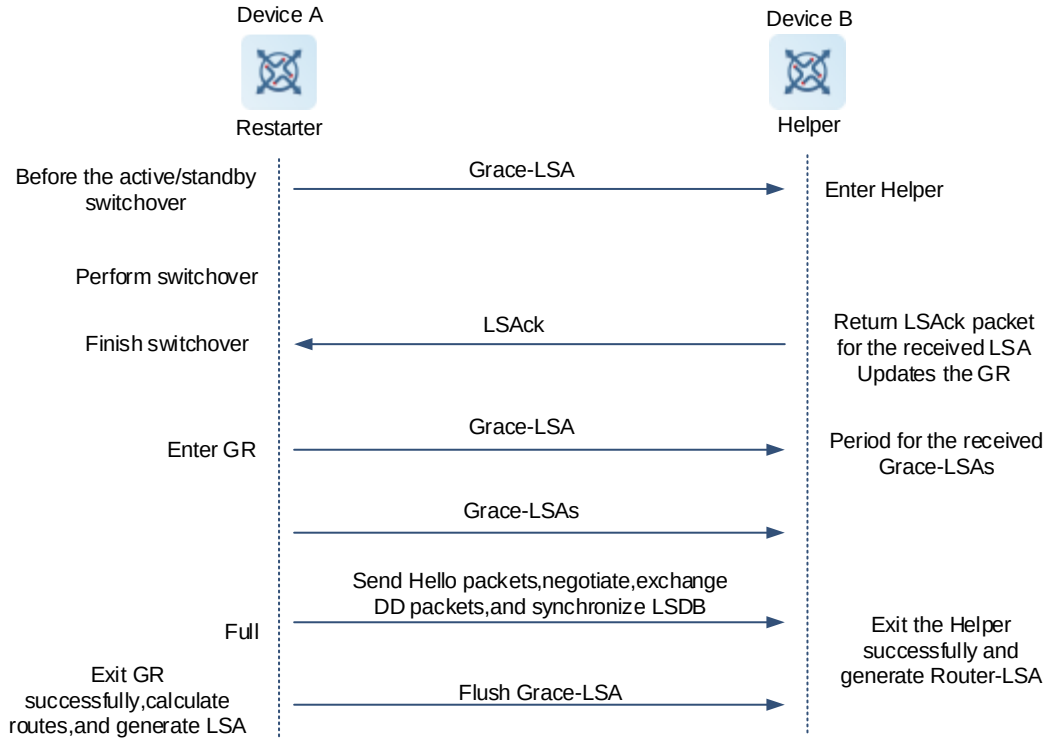
When the number of external routes on a router exceeds the upper limit, the router enters the overflow state. The router deletes the external routes generated by itself from the LSDB, and does not generate new external routes. In addition, the router discards the newly received external routes. After the overflow state timer (5s) expires, if the number of external routes is lower than the upper limit, the normal state is restored.

7. GR

The control and forwarding separated technology is widely used among routers. On a relatively stable network topology, when a GR-enabled router is restarted on the control plane, data forwarding can continue on the forwarding plane. In addition, actions (such as adjacency re-forming and route computation) performed on the control plane do not affect functions of the forwarding plane. In this way, service interruption caused by route flapping can be avoided, thus enhancing reliability of the entire network.

Currently, the GR function is used only during active/standby switchover and system upgrade.

Figure 1-9 Normal OSPF GR Process



- The GR process requires collaboration between the restarter and the helper. The restarter is the router where GR occurs. The helper is a neighbor of the restarter.
- When entering or exiting the GR process, the restarter sends a grace-LSA to the neighbor, notifying the neighbor to enter or exit the helper state.
- When the adjacency between the restarter and the helper reaches the full state, the router can exit the GR process successfully.

8. NSR

During nonstop routing (NSR), OSPF-related information is backed up from the active supervisor module of a distributed device to the standby supervisor module, or from the active host of a virtual switching unit (VSU) to the standby host. In this way, the device can automatically recover the link state and re-generate routes without the help of the neighbor devices during the active/standby switchover. The specific information of NSR active/standby synchronization includes configuration information backup, process information backup, interface information backup, neighbor information backup, and LSA information backup.

9. Fast convergence

When a link or a router on the OSPF network fails, the packets that need to be transmitted through the faulty link or the faulty router to reach the destination will be lost or cause a routing loop, and the data traffic will be interrupted. After a link fails, it takes a period of time (about 40s) before OSPF can sense the failure of the neighbor. Then, OSPF advertises the information and re-computes the SPT. During this period, traffic is interrupted. During this period, the traffic will be interrupted. The interrupted traffic cannot resume normal transmission until OSPF completes convergence according to the new topology network.

- To minimize the traffic interruption time caused by a network failure, the administrator can configure the fast hello function, correlation with BFD, or OSPF fast reroute function to speed up network convergence.
  - After the fast hello function is enabled (that is, the neighbor dead interval is set to 1s), OSPF can sense the failure of the neighbor within 1s once a link is faulty. This greatly accelerates route convergence and prevents traffic interruption.
  - BFD is used to test connectivity between devices. A link fault can be detected in as short as 150 ms. After OSPF is correlated with BFD, OSPF can sense the failure of a neighbor in as short as 150 ms once a link is faulty. This greatly accelerates route convergence and prevents traffic interruption.
  - Fast reroute prepares a backup route for OSPF. Once the OSPF senses the failure of a neighbor, the traffic is immediately switched over to the standby route, thus preventing traffic interruption.

10. iSPF feature

- The OSPF topology is area based. The SPF algorithm is run for independent computation in each area. The standard SPF algorithm re-computes the topology of the entire area each time even if only the leave nodes change in the area topology.
- When computing the network topology, the incremental shortest path first (iSPF) method corrects only the nodes on the SPT that are affected by the topology changes, and does not re-build the entire SPT. This can effectively ease the pressure on the router processors on a large network, especially when the network is not stable.

1.1.9 Network Management Functions

Functions such as management information base (MIB) and Syslog are used to facilitate OSPF management.

- MIB

  MIB is the device state information set maintained by a device. You can use the management program to view and set the MIB node.

Multiple OSPF processes can be simultaneously started on a router, but the OSPF MIB can be bound to only one OSPF process.

- Trap

  A trap message is a notification generated when the system detects a fault. This message contains the related fault information.

  If the trap function is enabled, the router can actively send the trap messages to the network management device.

- Syslog

  The syslog records the operations (such as command configuration) performed by users on routers and specific events (such as network connection failures).

  If the syslog is allowed to record the adjacency changes, the network administrator can view the logs to learn the entire process that the OSPF adjacency is set up and maintained.

1.1.10 VPN Extended Feature of OSPF

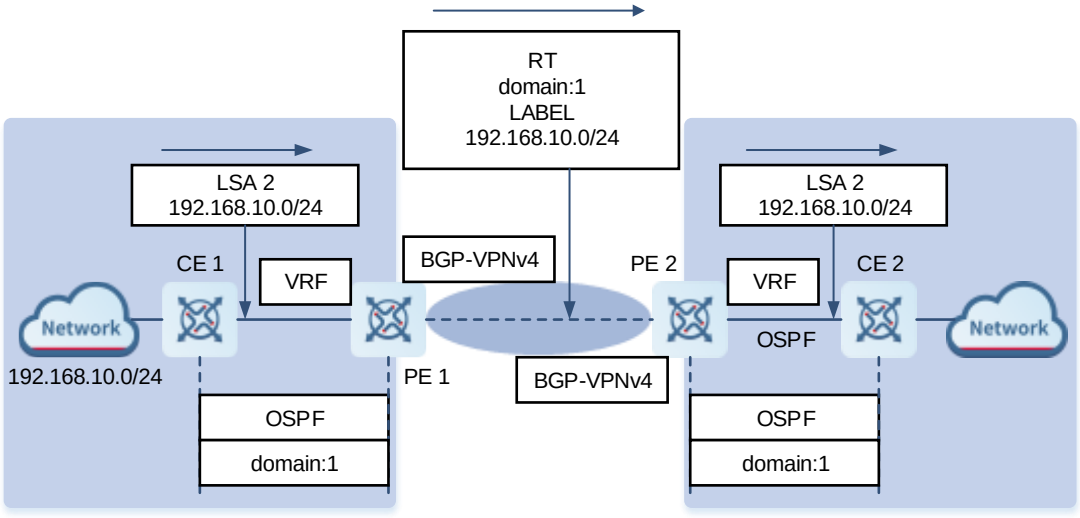
OSPF is a widely used IGP. In most of the existing application solutions, OSPF is generally selected as an internal routing protocol for VPN users. If OSPF is also used between provider edges (PE) and customer edge (CE), other routing protocols are not needed, which simplifies the configuration and management of CEs.

- Domain ID

  A domain ID identifies an OSPF domain to which a route belongs. When a CE learns an OSPF route in a VPN internal site, and this route is advertised to a PE as a Type-1/2/3 LSA, and is redistributed to the Border Gateway Protocol (BGP) domain and converted into a VPN route, the domain ID is also redistributed to the BGP domain along with the route. Then, this domain is also advertised as the extcommunity attribute of the VPN route. When another PE receives this VPN route and redistributes it to a VRF OSPF process, the domain ID is redistributed to the VRF OSPF process along with the route. If the VRF OSPF process confirms that the domain ID in the route is the same as that in the local VRF OSPF process, it advertises the route to the CE as an internal route. If the VRF OSPF process confirms that the domain ID in the route is different from that in the local VRF OSPF process, the VRF OSPF process advertises the route to the CE as an external route.

  As shown in Figure 1-10, for internal routes belonging to the same OSPF domain, CE1 advertises the routes to PE1 as Type-2 LSAs, which are converted into VPN routes and advertised to PE2. After receiving the routes, PE2 redistributes them to the VRF OSPF process. The domain IDs in the VRF OSPF process are the same as those of the VPN routes, and therefore, the VPN routes are advertised to the VPN site as internal routes.

Figure 1-10 Diagram of Domain ID

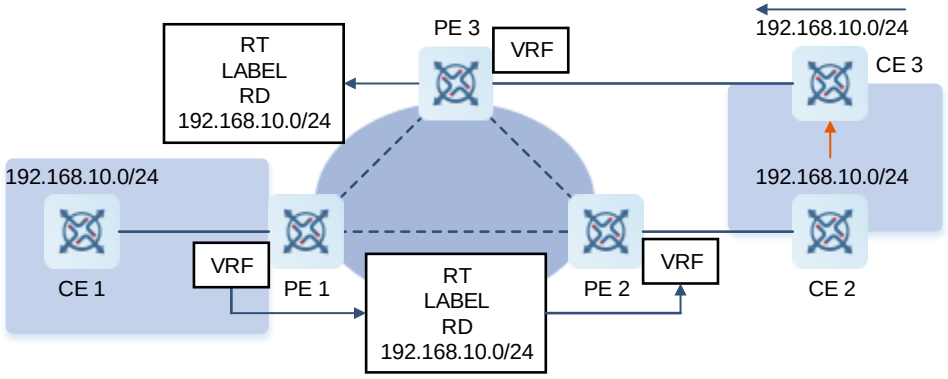


- DN bit technology

  DN bit is a bit in the optional field of the LSA header. The DN bit is a loop detection technology running the OSPF protocol between a PE and a CE. In some scenarios, loops may arise when OSPF runs between a PE and a CE, for example, multiple CEs are connected to one VPN site. If one PE advertises learned VPN routes to the VPN site, and the VPN site advertises the routes to another PE via OSPF, a loop may occur.

  As shown in Figure 1-11, PE1 advertises the 192.168.10.0/24 route to PE2 and PE3, CE2 advertises the route to CE3 via OSPF, and CE3 advertises the route to PE3, which redistributes the route to the BGP domain of PE3. PE3 selects the route that is redistributed via OSPF and converts this route into a VPNv4 route for advertisement. As a result, a loop may occur.

Figure 1-11 Diagram of DN Bit Technology



For the purpose of loop prevention, when a PE advertises a Type-3/5/7 LSA to a CE, the DN bit is set to ON in the optional field of the LSA, to prevent possible loops. When an LSA received by another PE contains the DN bit whose value is 1 in the optional field, the OSPF of the PE will not use this LSA for OSPF calculation.

- VPN router tag

  Router-tag is another loop detection technology running OSPF between a PE and a CE. When OSPF runs between a PE and a CE, the VRF OSPF process of the PE has a router tag by default, which is called VPN router tag. When a VPN route is imported into the VRF OSPF process of a PE and is converted into a Type-5/7 LSA and advertised to a CE, the LSA carries the VPN router-tag. When multiple PEs are connected to one VPN site, if the Type-5/7 LSA received by a PE contains the VPN router tag and the VPN router tag is the same as the VPN router tag of the OSPF process, the LSA will not be used for OSPF route calculation.

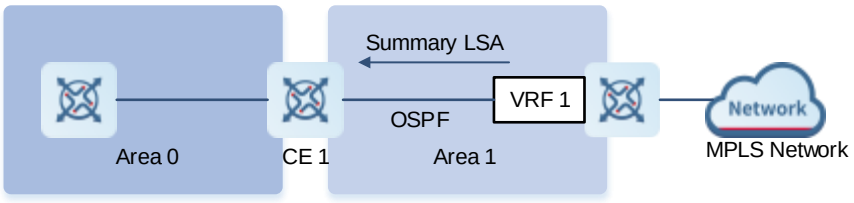
- Area deployment

  In normal cases, links between a PE and a CE may belong to any OSPF area. If links between a PE and a CE belong to a non-zero area, the PE is an ABR for the OSPF area where the CE resides. This may cause problems because the ABR running OSPF has the following features:

  - The ABR calculates Type-3 LSAs in the backbone area only.
  - The ABR forwards only Type-3 LSAs in the backbone area to a non-backbone area.

  As shown in Figure 1-12, if the link between a PE and a CE belongs to a non-zero area, the PE redistributes the VPNv4 routes advertised by Multiprotocol Extensions for BGP (MP-BGP) to the OSPF domain, restores them to type-3 LSAs, and advertises them to CE1. CE1 does not calculate LSAs in non-backbone areas. Therefore, these LSAs are not advertised to routers in Area 0 and the VPN internal sites cannot learn routes of other sites.

Figure 1-12 Diagram of Area Deployment

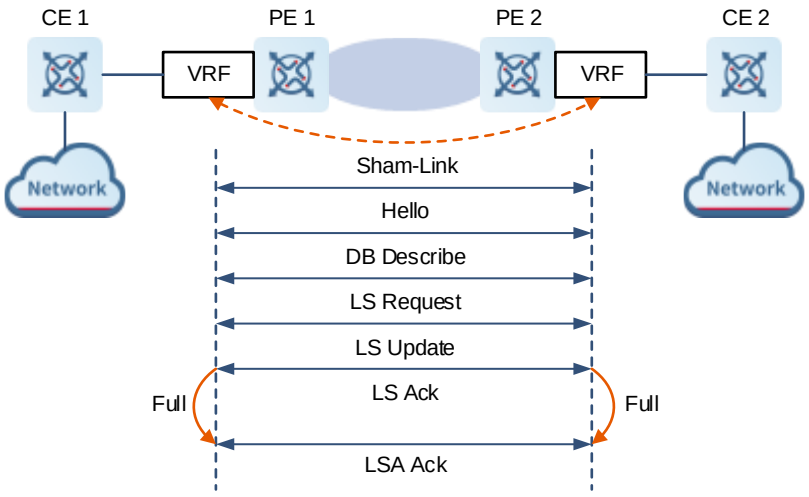


In L3VPN application, if OSPF is run between a PE and a CE to exchange VPN routes, do not deploy the backbone area at VPN internal sites if possible. If a router at a VPN internal site belongs to a backbone area in addition to a PE, at least one router at the VPN internal site must be connected to the PE and the links between the CE and PE must belong to Area 0. In this way, inter-area routes and external routes can be transmitted between the PE and the VPN site.

- Sham link

  The sham link is not a real link but a virtual link established between VRF OSPF instances of two PEs. Like a normal OSPF link, a sham link has its OSPF interfaces, and is capable of sending OSPF protocol packets, establishing neighbor relationships, and sending LSAs. When LSAs are flooded on a sham link, no types of the OSPF routes will change, as shown in Figure 1-13.

Figure 1-13 Diagram of Sham Link



The purposes of establishing sham-links between VRF OSPF processes of different PEs are as follows:

- When the Multiprotocol Interior Border Gateway Protocol (MP-IBGP) is used to carry private routes, it only transfers routes; after the routes reach the peer PE and restored, MP-IBGP imports the original OSPF routes in a best-effort manner, and the OSPF topology information cannot be communicated properly. With sham-link, an OSPF link can be established to implement actual interworking between OSPF processes at each site and establish a complete topology.
- Different sites in the same VPN exchange information via the MPLS backbone network. One link is connected between these VPN sites. The purpose of this link is that VPN sites can communicate with each other through this link when the MPLS backbone network is unavailable. This link is called backdoor link.

If two VPN sites belong to the same OSPF area and one backdoor link is connected in between, routes inside the two sites are exchanged through both the MPLS backbone network and the backdoor link. Routes exchanged through the MPLS backbone network are inter-domain routes while routes exchanged through the backdoor link are intra-domain routes. The intra-domain routes advertised by the backdoor link are prior to the inter-domain routes advertised by the MPLS backbone network. Therefore, the forwarding of the routes inside the two sites takes the backdoor link first, which is against the intention of backdoor link connection as a standby link for VPN users. In this case, a sham link is also required.

1.1.11 Protocols and Standards

- RFC 2328: OSPF Version 2
- RFC 2370: The OSPF Opaque LSA Option
- RFC 3137: OSPF Stub Router Advertisement
- RFC 3623: Graceful OSPF Restart
- RFC 3630: Traffic Engineering (TE) Extensions to OSPF Version 2
- RFC 3682: The Generalized TTL Security Mechanism (GTSM)

- RFC 3906: Calculating Interior Gateway Protocol (IGP) Routes Over Traffic Engineering Tunnels
- RFC 4576: Using a Link State Advertisement (LSA) Options Bit to Prevent Looping in BGP/MPLS IP Virtual Private Networks (VPNs)
- RFC 4577: OSPF as the Provider/Customer Edge Protocol for BGP/MPLS IP Virtual Private Networks (VPNs)
- RFC 4750: OSPF Version 2 Management Information Base

1.2 Configuration Task Summary

The OSPFv2 configuration includes the following tasks:

(1) Configuring Basic Features

- Configuring OSPF Process
- (Optional) Configuring Router ID
- Enabling OSPF on a Network Segment
- Enabling OSPF on an Interface
- (Optional) Creating a Virtual Link

(2) (Optional) Configuring OSPF Network Type

- Configuring the Interface Network Type
- (Optional) Configuring a Neighbor
- (Optional) Configuring the Interface Priority
- (Optional) Disabling Source Address Verification

(3) (Optional) Configuring Stub and NSSAs. The following tasks are optional. Select tasks for configuration according to actual condition.

- Configuring a Stub Area
- Configuring an NSSA

(4) (Optional) Adjusting OSPF Routing The following tasks are optional. Select tasks for configuration according to actual condition.

- Configuring the AD
- Configuring the Reference Bandwidth
- Configuring the Cost of an Interface
- Configuring Cost Fallback Value for an AP
- Configuring the Cost of the Default Route in a Stub Area or an NSSA
- Configuring the Default Metric Value for Redistribution Value
- Configuring the Maximum Metric
- Configuring RFC 1583 Compatibility
- Configuring Association of External Route Forwarding Addresses with Interface Configuration

(5) (Optional)Configuring OSPF Route Advertisement

- Configuring External Route Redistribution
-
- (Optional) Configuring a Device as ASBR

(6) (Optional)Configuring OSPF Route Summarization. The following tasks are optional. Select tasks for configuration according to actual condition.

- Configuring Inter-Area Route Summarization
- Configuring External Route Summarization
- Configuring a Discard Route

(7) (Optional)Configuring OSPF Route Filtering. The following tasks are optional. Select tasks for configuration according to actual condition.

- Configuring a Passive Interface
- Disabling Interfaces from Sending OSPF Packets
- Configuring Inter-area Route Filtering
- Configuring Redistributed Route Filtering
- Configuring Learned Route Filtering

(8) (Optional)Adjusting OSPF Network Convergence Speed. The following tasks are optional. Select tasks for configuration according to actual condition.

- Configuring the Hello Interval
- Configuring the Dead Interval
- Configuring LSU Transmission Delay
- Configuring LSU Retransmission Interval
- Configuring LSU Generation Time
- Configuring LSA Group Refresh Time
- Configuring LSA Group Refresh Interval
- Configuring Duplicate LSA Receiving Delay
- Configuring Inter-Area Route Computation Delay
- Configuring External Route Computation Delay
- Configuring SPF Computation Delay
- Configuring Route Calculation Suppression Time

(9) (Optional)Configuring OSPF Authentication

- Configuring the Authentication Type of an Area
- Configuring the Authentication Type of an Interface
- (Optional) Configuring a Plain Text Authentication Key for an Interface

- (Optional) Configuring an MD5 Authentication Key for an Interface

(10) (Optional)Enabling Fast Reroute

- Configuring Fast Reroute
- (Optional)Configuring Interface LFA Protection
- (Optional)Preventing an Interface From Becoming a Backup Interface

(11) (Optional)Modifying the Maximum Number of Concurrent Neighbors. The following tasks are optional. Select tasks for configuration according to actual condition.

- Configuring the Maximum Number of Concurrent Neighbors on the Current Process
- Configuring the Maximum Number of Concurrent Neighbors on All Processes

(12) (Optional)Configuring the GR Function. The following tasks are optional. Select tasks for configuration according to actual condition.

- Enabling the GR Function
- Configuring the OSPF GR Helper Function

(13) (Optional)Correlating OSPF with BFD. The following tasks are optional. Select tasks for configuration according to actual condition.

- Configuring BFD Correlation with OSPF on an Interface
- Configuring BFD Correlation with OSPF Globally

(14) (Optional)Enabling Overflow. The following tasks are optional. Select tasks for configuration according to actual condition.

- Configuring the Memory Overflow Function
- Configuring the Database Overflow Function
- Configuring the External LSA Database Overflow Function

(15) (Optional) Configuring VPN Extended Feature of OSPF. The following tasks are optional. Select tasks for configuration according to actual condition.

- Configuring Domain ID
- Configuring a Sham Link
- Configuring a Domain Label
- Disabling Loop Detection for a VRF OSPF Process
- Configuring Extcommunity Attribute for a VPN Route
- Disabling Loop Detection Using the DN Bit of LSA
- Disabling Loop Detection Using the Route Tag of LSA

(16) (Optional)Configuring Network Management Functions. The following tasks are optional. Select tasks for configuration according to actual condition.

- Binding the MIB with an OSPF Process
- Configuring the Trap Function

- Configuring the Logging Function

(17) (Optional)Enabling NSR

(18) (Optional)Enabling the iSPF Feature

(19) (Optional)Configuring to Disable MTU Verification

(20) (Optional)Configuring to Enable Conflict Detection of Type-4 LSAs

(21) (Optional)Enabling the Function of Securely Promoting a BDR to DR

1.3 Configuring Basic Features

1.3.1 Overview

Set up an OSPF routing domain on the network to provide IPv4 unicast routing service for users on the network.

1.3.2 Restrictions and Guidelines

- Ensure that the IP unitcast routing function is enabled, that is, **ip routing** is not disabled; otherwise, OSPF cannot be enabled.
- When OSPF is enabled, you are strongly advised to manually configure the router ID.
- After **ip ospf disable all** is configured, the interface neither sends or receives any OSPF packet, nor participates in OSPF computation even if the interface belongs to the advertisement range of the **network-advertise** command.

1.3.3 Configuration Tasks

The basic function configuration of OSPF includes the following tasks:

(1) Configuring OSPF Process

(2) (Optional) Configuring Router ID

(3) Enabling OSPF on a Network Segment

(4) Enabling OSPF on an Interface

(5) (Optional) Creating a Virtual Link

1.3.4 Configuring OSPF Process

1. Overview

To enable OSPF on the device, the OSPF process must be created by specifying the process ID.

2. Restrictions and Guidelines

- OSPF supports multiple instances. One or more OSPF processes can be started on a router. Each OSPF process runs OSPF independently, and the processes are mutually isolated.

- The process ID takes effect only on the local router, and does not affect exchange of OSPF packets on adjacent interfaces.
- This configuration must be performed on every router that requires OSPF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an OSPF process and enter OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

1.3.5 Configuring Router ID

1. Overview

Router ID is the unique identifier of a router in the OSPF domain, so the router ID must be unique on the entire network; otherwise it will lead to problems such as abnormality in neighbor relationship establishment and routing information error. Moreover, different OSPF processes in the same router need to use different router IDs.

2. Restrictions and Guidelines

- You are strongly advised to manually configure the router ID.
- Different OSPF processes on the same router must have different Router IDs.
- When a router ID is not manually configured, OSPF automatically selects an interface IP address. If no IP address is configured for any interface, or the configured IP addresses have been used by other OSPF processes, you must manually configure the router ID.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure a router ID.

router-id *router-id*

By default, the OSPF routing process elects the largest IP address among all the loopback interfaces as the router ID. If the loopback interfaces configured with IP addresses are not available, the OSPF process elects the largest one among the IP addresses of all its physical interfaces as the router ID.

1.3.6 Enabling OSPF on a Network Segment

1. Overview

To realize interaction of OSPF routing information between devices, interfaces of the devices must be added to the OSPF process. Through enabling OSPF on a network segment, all interfaces whose IP addresses fall within the range of the network segment join the OSPF process, and are associated with the area.

2. Restrictions and Guidelines

- All routers running OSPF must be configured to join interfaces into the OSPF process.
- When an interface IP address matches the IP address range defined by the **network-advertise area** command of multiple OSPF processes at the same time, the OSPF process that the interface joins will be determined by the optimal matching method.
- If the IP address range defined by the **network-advertise area** command only includes the secondary IP address of the interface, OSPF will not run on the interface.

When both the network-advertise area command in the OSPF process configuration mode, and the ip ospf set-area command in the interface configuration mode are configured, the latter will prevail.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Enable OSPF on the network segment, and specify an Area ID.

network-advertise *ipv4-address wildcard* **area** *area-id*

No interface IP address is configured to join the OSPF area by default.

1.3.7 Enabling OSPF on an Interface

1. Overview

To realize interaction of OSPF routing information between devices, interfaces of the devices must be added to the OSPF process. By enabling OSPF on the specified interface, all IP addresses under the interface will automatically join the OSPF process, and be associated with the area.

2. Restrictions and Guidelines

- The configuration is mandatory for each OSPF-enabled router.
- You can also add an interface to OSPF process using the **network-advertise** command on the instance. If two commands are run at the same time, the configuration on the interface takes effect first.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Enable OSPF on an interface and specify an area ID.

ip ospf *process-id* **set-area** *area-id*

OSPF is disabled on an interface by default.

(5) Allow the interface to generate OSPF packets.

unset ip ospf disable filter

By default, an interface is allowed to generate OSPF packets.

1.3.8 Creating a Virtual Link

1. Overview

In the OSPF routing domain, all the areas must be connected to the backbone area. If the backbone area is disconnected, a virtual link must be configured to connect to the backbone area; otherwise, network communication problems will occur. A virtual link must be created between two ABRs, and the area to which both ABRs belong is the transit area. A stub area or an NSSA cannot be used as a transit area. A virtual link can also be used to connect other non-backbone areas.

2. Restrictions and Guidelines

- Only the authentication key of the virtual link is defined in a virtual link. To enable OSPF packet authentication in the areas connected to the virtual link, run the **area auth-algorithm** command to enable the authentication.
- The virtual link supports the fast hello function.
  - If the fast hello function is configured for a virtual link, the hello interval field of the hello packet advertised on the virtual link is set to 0, and the hello interval field of the hello packet received on this virtual link is ignored.
  - No matter whether the fast hello function is enabled, the neighbor dead interval must be always consistent on the devices at both ends of the virtual link.
  - The **dead-period**, **minimum hello-multiplier**, and **hello-period** parameters introduced for the fast hello function cannot be configured simultaneously.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure a virtual link.

area *area-id* **vlink** *router-id* [**auth-algorithm** [**keychain** *kechain-name* | **auth-secret** | **null**] | **dead-period** { *dead-interval* | **minimum hello-multiplier** *multiplier-time* } | **hello-period** *hello-interval* | **retransmit-period** *retransmit-interval* | **transmit-lsu-delay** *transmit-delay*] * [**auth-algorithm-key** [**0** | **7**] *key* | **auth-secret-key** *key-id* **md5** [**0** | **7**] *key*]

No virtual link is configured by default.

1.4 Configuring OSPF Network Type

1.4.1 Overview

If the physical network is X.25, frame relay, or ATM, set an OSPF network type so that the network can also run OSPF and provide IPv4 unicast routing service.

OSPF classifies networks into five types according to the link layer protocol type so that LSAs are exchanged in different ways to set up an adjacency. You can configure to forcibly change the network type of an interface.

1.4.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- The network types must be consistent for the devices at both ends of the link. Otherwise, the neighbor relationship cannot be set up.
- The broadcast network sends multicast OSPF packets, automatically discovers neighbors, and elects a DR and a BDR.
- The P2P network sends multicast OSPF packets and automatically discovers neighbors.
- The NBMA network sends unicast OSPF packets. Neighbors must be manually specified, and a DR and a BDR must be elected.
- The P2MP network (without carrying the **non-broadcast** parameter) sends multicast OSPF packets. Neighbors are automatically discovered.
- The P2MP network (carrying the **non-broadcast** parameter) sends unicast OSPF packets. Neighbors must be manually specified.

1.4.3 Configuration Tasks

The OSPF network type configuration includes the following tasks:

(1) Configuring the Interface Network Type

(2) (Optional) Configuring a Neighbor

(3) (Optional) Configuring the Interface Priority

(4) (Optional) Disabling Source Address Verification

1.4.4 Configuring the Interface Network Type

1. Overview

There are five OSPF network types. By default, Ethernet and FDDI belong to the broadcast type, X.25, frame relay, and ATM belong to the NBMA type, and PPP, HDLC, and LAPB belong to the P2P type. This configuration allows you to forcibly modify the network type of the interface.

2. Restrictions and Guidelines

- The configuration is required on routers at both ends of the link.
- The broadcast type requires that the interface must have the broadcast capability.
- The P2P type requires that the interfaces are interconnected in one-to-one manner.
- The NBMA type requires full-meshed connections, and all interconnected routers can directly communicate with each other.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure a network type for the interface.

ip ospf= network-type { **broadcast** | **non-broadcast** | **point-to-multipoint** [**non-broadcast**] | **point-to-point** }

By default, the network interface of OSPF is not configured, and no interface is set to P2MP type.

1.4.5 Configuring a Neighbor

1. Overview

On the NBMA network, routers cannot send or receive broadcast packets, and cannot discover neighbors through hello packets. You need to specify neighbors manually and send unicast hello packets.

If a neighbor router becomes inactive on the NBMA network, OSPF still sends hello packets to this neighbor even if no hello packet is received within the router failure time. The interval at which the hello packet is sent is called polling interval. When running for the first time, OSPF sends hello packets only to neighbors whose priorities are not 0. In this way, neighbors with priorities set to 0 do not participate in the DR/BDR election. After a DR/BDR is elected, the DR/BDR sends the hello packets to all neighbors to set up a neighbor relationship.

2. Restrictions and Guidelines

- If the network type of an interface is set to NBMA or P2MP (carrying the **non-broadcast** parameter), neighbors must be configured.
- Neighbors are configured on routers at both ends of the NBMA or P2MP (carrying the **non-broadcast** parameter) network.
- The neighbor IP address must be the primary IP address of this neighbor interface.
- The P2MP (non-broadcast) network cannot dynamically discover neighbors because it does not have the broadcast capability. Therefore, you must run this command to manually configure neighbors for the P2MP (non-broadcast) network.
- You can use the **advertise-metric** parameter to specify the cost to reach each neighbor on the P2MP network.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf= *process-id* [**vrf-instance** *vrf-name*]

(4) Configure a neighbor.

peer *ipv4-address* [**metric** *cost* | [**poll-interval** *poll-interval* | **priority-value** *priority*] *]

No neighbor is configured by default. Neighbors must be specified for the NBMA or P2MP (non-broadcast) interfaces. The neighbor IP address must be the primary IP address of this neighbor interface. In addition, you can use the **metric** parameter to specify the cost to reach each neighbor on the P2MP network.

1.4.6 Configuring the Interface Priority

1. Overview

When the DR/BDR election occurs on the OSPF broadcast network, the router with the highest priority becomes the DR or BDR. If the priorities are the same, the router with the largest router ID becomes the DR or BDR. A router with the priority set to 0 does not participate in the DR/BDR election. The priority value of an OSPF interface is contained in the hello packet.

2. Restrictions and Guidelines

- This command is applicable only to the OSPF broadcast and NBMA interfaces.
- You must configure the interface priority if a router must be specified as a DR, or a router cannot be specified as a DR.
- Configure the interface priority on a router that must be specified as a DR, or cannot be specified as a DR.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.

- **interface overlayrouter** *interface-number*

(4) Configure the interface priority.

ip ospf priority-value *priority*

The priority value is 1 by default.

1.4.7 Disabling Source Address Verification

1. Overview

Generally, the source address of a packet received by OSPF is in the same network segment as the receiving interface. The addresses at both ends of a P2P link are configured separately and are not necessarily in the same network segment. In this scenario, as the peer address information is notified during the P2P link negotiation, OSPF checks whether the source address of the packet is the address advertised by the peer end during negotiation. If not, OSPF determines that the packet is invalid and discards it. Besides, OSPF does not verify the address of an unnumbered interface.

In some scenarios, the source address of a packet received by OSPF may not be in the same network segment as the receiving interface, so the OSPF address verification fails. For example, the negotiated peer address cannot be obtained on a P2P link. In this scenario, source address verification must be disabled to ensure that the OSPF adjacency can be properly set up.

2. Restrictions and Guidelines

- The OSPF basic functions must be configured.
- Source address verification cannot be disabled on a broadcast or NBMA network.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*

- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Disable source address verification.

ip ospf source-check-ignore

The source address verification on a P2P link is enabled by default.

1.5 Configuring Stub and NSSAs

1.5.1 Overview

Configuring a stub area as the Stub area or NSSA helps to reduce the number of LSAs in the area, reduce the amount of routing information interaction, reduce the size of the routing table, and enhance routing stability.

1.5.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- A backbone area and transmission area cannot be configured as a Stub area or NSSA.
- Routers in the Stub area cannot import external routes, while routers in the NSSA can import external routes.

1.5.3 Configuration Tasks

The configuration includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Configuring a Stub Area
- Configuring an NSSA

1.5.4 Configuring a Stub Area

1. Overview

Stub area is the stub area of OSPF. Configuring a Stub Area can reduce the number of LSAs in the area. The ABR router in the Stub area will not transmit Type 5 LSAs (external routes), but will also advertise a Type 3 default route to the Stub area.

Totally Stub area is a totally stub area, which is an upgrade of the Stub area. Routers in the Totally Stub area will not transmit Type 3 LSAs (inter-domain routes) and Type 5 LSAs (external routes). The ABR will only advertise a Type 3 default route to the Totally Stub area.

To reduce the size of the routing table of the routers in the area, you can configure a Stub area or Totally Stub area.

2. Restrictions and Guidelines

- The area must be configured as Stub area on all routers in the area.
- To configure a Totally Stub area, you only need to select the **unset-summary** keyword when running the **area stub** command on the ABR. The routing devices in the Totally Stub area can only learn the routes in the area and the internal default routes generated by the ABR.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configuring a Stub Area.

area *area-id* **stub** [**unset-summary**]

By default, the Stub area function is disabled.

1.5.5 Configuring an NSSA

1. Overview

NSSA is similar to the Stub area, but allows ASBR. ASBR redistributed routes will be transmitted in the NSSA in the form of a Type 7 LSA. On the ABR, the Type 7 LSA is converted to Type 5 LSA and then transmitted to other areas. This ABR is also called the translator. The ABR will also advertise the default route to the NSSA.

Totally NSSA area is an upgrade of NSSA. Compared with routers in the NSSA, routers in the Totally NSSA can only advertise a default inter-domain route, and other Type 3 LSAs (inter-domain routes) are not allowed to be transmitted.

You are advised to configure NSSA or Totally NSSA if you want to reduce the size of the routing table of routers in the area and the number of LSAs in the area, and to import OSPF routes outside the area.

2. Restrictions and Guidelines

- The area must be configured as NSSA on all routers in the area.
- When the **default-originate** parameter is executed on the ABR of the NSSA, the Type 7 LSA default route will be generated regardless of whether there is a default route in the route table. When this parameter is executed

on the ASBR (also not an ABR) of the NSSA, the Type 7 LSA default route will be generated only if there is a default route in the route table.

- Configuring a Totally NSSA area can further reduce the number of LSAs sent to the NSSA. If the **unset-summary** attribute is configured on the ABR, the ABR will be prevented from sending summary LSAs (Type 3 LSAs) to the NSSA.
- In one NSSA, you are advised to specify only one ABR as the translator.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configuring an NSSA.

area *area-id* **nssa** [**default-originate** [**metric-value** *metric* | **metric-class** *metric-type*] * | **unset-redistribute** | **unset-summary** | **translator** { **constantly** | **stability-interval** *stability-interval* } * | **zero-address-forwarding**] *

By default, the NSSA function is disabled.

1.6 Adjusting OSPF Routing

1.6.1 Overview

Configuring routing parameters such as administrative distance, reference bandwidth, and interface cost, can change the order in which routers select various routes to increase or decrease the priority of OSPF routes, and change OSPF routes to allow traffic to pass through or bypass certain nodes.

1.6.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- If you run the **ip ospf metric** command to configure the cost of an interface, the configured cost will automatically overwrite the cost that is computed based on the auto cost.

1.6.3 Configuration Tasks

The configuration for adjusting OSPF routing includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Configuring the AD
- Configuring the Reference Bandwidth

- Configuring the Cost of an Interface
- Configuring Cost Fallback Value for an AP
- Configuring the Cost of the Default Route in a Stub Area or an NSSA
- Configuring the Default Metric Value for Redistribution value
- Configuring the Maximum Metric
- Configuring RFC 1583 Compatibility
- Configuring Association of External Route Forwarding Addresses with Interface Configuration

1.6.4 Configuring the AD

1. Overview

AD is used to select the best path when multiple routing protocols are used to reach the same target. AD defines the reliability of routing protocol. A smaller AD indicates higher reliability and higher routing priority. The default AD values of OSPF intra-domain routes, inter-domain routes, and external routes are all 110. Different ADs can be set to control routing.

2. Restrictions and Guidelines

- Perform this configuration if you hope to change the priorities of OSPF routes on a router that runs multiple unicast routing protocols.
- Run this command to specify different ADs for different types of OSPF routes.
- Configure the **route-map** parameter and set an AD for the specific route through a policy. If the route map is configured with **set-rule distance**, then:
  - Through the matched route: the AD is set by the **set-rule distance** command.
  - Not through the matched route: the AD is set by the **distance** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure an AD.

distance { *distance* [**route-map** *map-name*] | **ospf** { **external-routes** *distance* [**route-map** *map-name*] | **inter-area-routes** *distance* [**route-map** *map-name*] | **intra-area** *distance* [**route-map** *map-name*] } * }

By default, the administrative distance is **110** for all the OSPF routes.

1.6.5 Configuring the Reference Bandwidth

1. Overview

The default value of OSPF reference bandwidth is 100 Mbps, and the cost is the reference bandwidth divided by interface bandwidth. In the Gigabit and 10 Gigabit network environments, the calculated cost value is 1, so the default bandwidth value cannot adapt to the current high-speed transport network. A router is connected to lines with different bandwidths. You are advised to adjust the configuration according to the actual bandwidth if you want to preferentially select the line with a larger bandwidth.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the reference bandwidth.

auto-gen-cost reference-bandwidth *reference-bandwidth*

By default, the reference bandwidth value of interface metric calculation is 100 Mbps.

1.6.6 Configuring the Cost of an Interface

1. Overview

Modifying the interface cost of OSPF can optimize routing. A lower cost of the link indicates a higher priority during routing.

2. Restrictions and Guidelines

- By default, the cost of an OSPF interface is equal to the reference value of the auto cost divided by the interface bandwidth.
- A router is connected with multiple lines. You are advised to configure a lower cost for a specific interface if you wish to manually specify a preferential line.
- Run the **auto-gen-cost** command to obtain the reference value of the auto cost. The default value is 100 Mbps.
- Run the **bandwidth** command to set the interface bandwidth.
- The default costs of OSPF interfaces on several typical lines are as follows:
  - 64 Kbps serial line: The cost is 1562.
  - E1 line: The cost is 48.
  - 10M Ethernet: The cost is 10.

- 100M Ethernet: The cost is 1.

- If you run the **ip ospf metric** command to configure the cost of an interface, the configured cost will automatically overwrite the cost that is computed based on the auto cost.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure the cost of an interface.

ip ospf metric *cost*

By default, the cost of an OSPF interface is equal to the reference value of the auto cost divided by the interface bandwidth (the reference bandwidth is 100 Mbps by default).

1.6.7 Configuring Cost Fallback Value for an AP

1. Overview

The bandwidth of an AP is equal to the sum of the bandwidths of all the valid member ports. When a member port fails, the bandwidth of the AP will be reduced. You can set a cost fallback value for the AP to enable OSPF to select other paths preferably. When the failed member port of the AP recovers, the cost fallback value becomes invalid, and the metric of the AP returns to normal.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

(4) Configure the cost fallback value for an AP.

ip ospf cost-fallback *cost* **threshold** *bandwidth*

The cost fallback function is disabled on an AP by default.

1.6.8 Configuring the Cost of the Default Route in a Stub Area or an NSSA

1. Overview

After an area is set to a stub area/NSSA, ABR will advertise the default route to the stub area/NSSA, ensuring that the routes from the stub area/NSSA to other areas are reachable. The default cost of a default route is 1. This function can adjust the cost of the default route and flexibly control the routes of the stub area and NSSA.

2. Restrictions and Guidelines

This command takes effect only on an ABR in a stub area or on an ABR/ASBR in an NSSA.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the cost of the default route in a stub area or an NSSA.

area *area-id* **default-metric** *cost*

By default, the cost of a route is 1.

1.6.9 Configuring the Default Metric Value for Redistribution

1. Overview

When redistributing external routes, you can configure the costs of external routes to adjust routing flexibly.

2. Restrictions and Guidelines

- This configuration is mandatory if the cost of external routes of the OSPF domain should be specified when external routes are introduced to an ASBR.
- The **default-metric** command must be used together with the **redistribute-route** command in routing process configuration mode to modify the initial metrics of all redistributed routes.
- The **default-metric** command does not take effect on external routes that are injected to the OSPF routing domain using the **default-info originate** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the default metric for redistribution.

default-metric *metric*

By default, the metric of a redistributed route is 20.

1.6.10 Configuring the Maximum Metric

1. Overview

In the scenarios of restarting a device and using a device as a candidate path, the OSPF information advertised by the current device can be configured with the maximum metric value so that the data traffic will not pass through the current device.

2. Restrictions and Guidelines

- A router may be unstable in the restart process or a period of time after the router is restarted, and users do not want to forward data traffic through this router. In this case, this configuration is recommended.
- After the **max-advertise-metric router-lsa** command is configured, the metrics of the non-stub links in the router LSAs generated by the router will be set to the maximum value (0xFFFF). If you cancel this configuration or the timer expires, the normal metrics of the links are restored.
- The **ip ospf max-advertise-metric neighborup** *seconds* command in interface configuration mode has the same function as the **max-advertise-metric router-lsa neighbor** command configured on an instance except the following: In interface configuration mode, the configuration takes effect from the interface's first neighbor in full neighbor relationship; in instance configuration mode, the configuration takes effect from the first neighbor in full neighbor relationship of the entire instance.
- By default, if the **max-metric router-lsa** command is configured, the stub links still advertise common metrics, that is, the costs of outbound interfaces. If the **include-stub-links** parameter is configured, the stub links will advertise the maximum metric.
- If an ABR is not intended to transfer inter-area traffic, use the **summary-lsa-metric** parameter to set the metric of the summary LSA to the maximum value.
- If an ASBR does not wish to transfer external traffic, use the **external-lsa** parameter to set the metric of the external LSA to the maximum value.
- The **maxadvertise-metric router-lsa** command is generally used in the following scenarios:
  - Restart a device. After the device is restarted, IGP generally converges faster, and other devices attempt to forward traffic through the restarted device. If the current device is still building the BGP routing table and some BGP routes are not learned yet, packets sent to these networks are discarded. In this case, you can use the **startup** parameter to set a delay after which the restarted device acts as the transmission mode.
  - Add a device to the network but the device is not used to transfer traffic. The device is added to the network. If a candidate path exists, the current device is not used to transfer traffic. If a candidate path does not exist, the current device is still used to transfer traffic.
  - Delete a device gracefully from the network. After the **max-metric router-lsa** command is configured, the current device advertises the maximum metric among all the routes. In this way, other devices on the network can select the backup path for data transmission before the device is shut down.
  - In the earlier OSPF version (RFC1247 or earlier), the links with the maximum metric (0xFFFF) in the LSAs do not participate in the SPF computation, that is, no traffic is sent to routers that generate these LSAs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the maximum metric.

max-advertise-metric router-lsa [**external-lsa** [*max-metric-value*] | **neighborup** [*full-interval-time*] | **include-stub-links** | **startup** [*startup-interval-time*] | **summary-lsa-metric**

By default, the LSAs of normal metric are advertised.

1.6.11 Configuring RFC 1583 Compatibility

1. Overview

When there are multiple paths to an ASBR or the forwarding address of an external route, RFC 1583 and RFC 2328 define different routing rules. If RFC 1583 compatibility is configured, a path in the backbone area or an inter-area path is preferentially selected. If RFC 1583 compatibility is not configured, a path in a non-backbone area is preferentially selected.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure compatible RFC 1583.

reference rfc1583

The RFC1583 rule function is enabled by default.

1.6.12 Configuring Association of External Route Forwarding Addresses with Interface Configuration

1. Overview

Setting the forwarding address (FA) of relevant flag bit can associate the interface configuration with the external route forwarding address to calculate the reachability and metric.

2. Restrictions and Guidelines

- In the case of interconnection with a Cisco device, the FA configuration compatible with the Cisco device must be enabled.
- When the FA of an external route is 0, external route computation of OSPF will compute the reachability and metric to the advertiser. When the FA is not 0, the reachability and metric to the forwarding address will be

computed.

- When the following conditions are met simultaneously in default configuration, the ASBR of OSPF fills the forwarding address field of the external route with digits other than 0:
  - OSPF is enabled on the next hop interface for connecting the ASBR to external network.
  - The next hop interface for connecting the ASBR to the external network is in the network scope advertised in the OSPF protocol.
- When the following conditions are met simultaneously after association of forwarding address rules with interface configuration is configured, the ASBR of OSPF fills the forwarding address field of the external route with digits other than 0:
  - OSPF is enabled on the next hop interface for connecting the ASBR to external network.
  - The next hop interface for connecting the ASBR to the external network is in the network scope advertised in the OSPF protocol.
  - The next hop interface for connecting the ASBR to the external network is not configured as a passive interface.
  - The next hop interface for connecting the ASBR to the external network is not P2P or P2MP type of OSPF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf- *process-id* [**vrf-instance** *vrf-name*]

(4) Configure association of external route forwarding addresses with interface configuration.

reference forward-address

By default, the external route forwarding address is set according to the default rules.

1.7 Configuring OSPF Route Advertisement

1.7.1 Overview

The function aims to introduce unicast routes for other AS domains or default routes to other AS domains to the OSPF domain and provide the unicast routing service to other AS domains for users in the OSPF domain. The router that introduces routes will automatically become an ASBR.

OSPF has two types of external routes. The metric of the Type-1 external route changes, but the metric of the Type-2 external route is fixed. If two parallel paths to the same destination network have the same route metric, the

priority of Type-1 route is higher than that of Type-2 route. Therefore, the display ip route-config command displays only the Type-1 route.

1.7.2 Restrictions and Guidelines

The OSPF basic functions must be configured.

1.7.3 Configuration Tasks

The configuration of OSPF route advertisement includes the following tasks:

(1) Configuring External Route Redistribution

(2) Configuring VRF-based External Route Redistribution

(3) (Optional) Configuring a Device as ASBR

1.7.4 Configuring External Route Redistribution

1. Overview

Configure external route redistribution if external routes of the OSPF domain need to be introduced to an ASBR.

2. Restrictions and Guidelines

- Perform this configuration on an ASBR.
- After this command is configured, the router becomes an ASBR, imports related routing information to the OSPF domain, and advertises the routing information as Type-5 LSAs to other OSPF routers in the domain.
- If the **level** parameter is not carried when IS-IS route redistribution is configured, only Level-2 routes can be redistributed by default. The **level** parameter is carried during initial configuration of redistribution, the routes configured with the **level** parameter can be redistributed. If both **level 1** and **level 2** are configured, the two levels are combined and saved as **level-1-2**.
- If you configure redistribution of OSPF routes without specifying the **match** parameter, OSPF routes of all sub-types can be distributed by default. The latest setting of the **match** parameter is used as the initial **match-rule** parameter. Only routes that match the sub-types can be redistributed. You can run the **unset** form of the command to restore the default value of **match-rule**.
- If **route-map** is specified, the **match-rule** filtering rules specified in **route-map** are applicable to the original parameters of redistribution. For redistribution of OSPF or IS-IS routes, **route-map** is used for filtering only when the redistributed routes meet the criteria specified by **match-rule** or **level**.
- The **set metric-value** value range of the associated **route-map** is from 0 to 16777214. If the value exceeds this range, routes cannot be introduced.
- The configuration rules for the **unset** form of the **redistribute-route** command are as follows:
  - If some parameters are specified in the **unset** form of this command, default values of these parameters will be restored.
  - If no parameter is specified in the **unset** form of this command, the entire command will be deleted.

- For example, if **redistribute-route isis 112 level-2** is configured, you can run the **unset redistribute-route isis 112 level-2** command to restore the default value of level-2. As **level-2** itself is the default value of the parameter, the configuration saved is still **redistribute-route isis 112 level-2** after the preceding **unset** form of the command is executed. To delete the entire command, run the **unset redistribute-route isis 112** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure external route redistribution.

redistribute-route { **arp**-**host** | **bgp** | **connected** | **isis** [*area-tag*] [**level-1** | **level-1-2** | **level-2**] | **ospf** *process-id* [**match-rule** { **externa**l [**1** | **2**] | **internal-type** | **nssa-external-route** [**1** | **2**] }] | **rip** | **static** } [**metric-value** *metric-value* | **metric-class** { **1** | **2** } | **route-map** *route-map-name* | **subnets** | **tag-value** *tag-value*]

The route redistribution function is not configured by default.

1.7.5 Configuring VRF-based External Route Redistribution

1. Overview

Configure VRF-based external route redistribution if external routes of the OSPF domain need to be introduced to an ASBR.

2. Restrictions and Guidelines

- Perform this configuration on an ASBR.
- After this command is configured, the router becomes an ASBR, imports related routing information into the OSPF domain, and advertises routing information as Type-5 LSAs to other OSPF routers in the domain. When this function is configured together with the **redistribute-route** command configured on an instance, the parameter configuration of the instance is used for LSA generation.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure VRF-based external route redistribution.

router-protocol ospf redistribute-route { **bgp** | **connected** | **Isis** | **ospf** | **rip** | **static** } [**vrf-instance** *vrf-name*] [**route-map** *route-map-name*]

VRF-based route redistribution is not configured by default.

1.7.6 Generating a Default Route

1. Overview

A default route needs to be introduced to an ASBR so that other routers in the OSPF domain access other AS domains through this ASBR by default.

2. Restrictions and Guidelines

- When the **redistribute-route** or **default-info** command is executed, the OSPF router automatically becomes an ASBR. The ASBR, however, does not automatically generate or advertise a default route to all routers in the OSPF routing domain. To enable an ASBR to generate a default route, configure the **default-info originate** command.
- The metric of the external default route can only be defined in the **default-info originate** command, instead of the **default-metric** command.
- A router in a stub area cannot generate an external default route.
- The **set metric-value** value range of the associated **route-map** is from 0 to 16777214. If the value exceeds this range, routes cannot be introduced.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Generate a default route.

default-info originate [**constantly** | **metric-value** *metric* | **metric-class** *type* | **route-map** *map-name*] *

No default route is generated by default.

If the **constantly** parameter is specified, the OSPF routing process advertises an external default route to neighbors regardless of whether a default route exists. This default route, however, is not displayed on the local router. To confirm whether the default route is generated, run the **display ip ospf database-info** command to display the OSPF LSDB. The external link with the ID 0.0.0.0 describes the default route. On an OSPF neighbor, you can run the **display ip route-config** command to see the default route.

1.7.7 Configuring a Device as ASBR

1. Overview

Configure a device as ASBR in the AS domain or AS boundary.

2. Restrictions and Guidelines

After the redistribute-route or default-info command is executed, the OSPF router automatically becomes an ASBR. If you want the device to become an ASBR without configuring the above command, configure the asbr enable command. If the asbr enable command is deleted, but the redistribute-route or default-info command is still configured, the device is still an ASBR.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure a device as ASBR.

asbr enable

ASBR is disabled by default.

1.8 Configuring OSPF Route Summarization

1.8.1 Overview

Summarize routes to reduce interaction of routing information and the size of routing table, and enhance stability of routes. Shield or filter routes.

1.8.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- The address range of the route to be summarized may exceed the actual network range in the routing table. If data is sent to a network beyond the summarization range, a routing loop may be formed and the router processing load may increase. To prevent these problems, a discard route must be added to the routing table to shield or filter routes.

1.8.3 Configuration Tasks

The configuration of OSPF route summarization includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Configuring Inter-Area Route Summarization
- Configuring External Route Summarization
- Configuring a Discard Route

1.8.4 Configuring Inter-Area Route Summarization

1. Overview

On a large-scale OSPF network, configuring route summarization can reduce the size of a routing table and the LSA information, thus reducing the system loss.

2. Restrictions and Guidelines

- Perform this configuration when routes of the OSPF area need to be summarized.
- Unless otherwise required, perform this configuration on an ABR in the area where routes to be summarized are located.
- This command can be executed only on the ABR. It is used to combine or summarize multiple routes of an area into one route, and then advertise the route to other areas. Combination of the routing information occurs only on the boundary of an area. Routers inside the area can learn specific routing information, whereas routers in other areas can learn only one summarized route. You can set **advertise** or **not-advertise** to determine whether to advertise this summarized route to shield and filter routes. By default, the summarized route is advertised. You can use the **advertise-metric** parameter to set the metric of the summarized route.
- You can configure the route summarization command for multiple areas. This simplifies routes in the entire OSPF routing domain, and improves the network forwarding performance, especially for a large-sized network.
- When multiple routes for summarization are configured and have an inclusive relationship with each other, the range of routes to be summarized is determined based on the longest match principle.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure inter-area route summarization.

area *area-id* **summary** *ipv4-address mask* [**advertise** | **not-advertise**] [**metric** *cost* | **inherit-minimum**]

The inter-area route summarization function is disabled by default.

1.8.5 Configuring External Route Summarization

1. Overview

When a lot of OSPF domain routes need to be introduced, configuring route summarization can reduce the size of a routing table and the LSA information, thus reducing the system loss.

2. Restrictions and Guidelines

- Perform this configuration when routes outside the OSPF domain need to be summarized.
- Unless otherwise required, perform this configuration on an ASBR, to which routes that need to be summarized are introduced.
- When routes are redistributed from other routing processes and injected to the OSPF routing process, every route is advertised to the OSPF routers using an external LSA respectively. If the injected routes are in a continuous address space, the ASBR can advertise only one summarized route to significantly reduce the size of the routing table.
- The **area summary** command summarizes the routes between OSPF areas, whereas **summary-address** summarizes external routes of the OSPF routing domain.
- When configured on the NSSA ABR translator, **summary-address** summarizes redistributed routes and routes obtained based on the LSAs that are converted from Type 7 to Type 5. When configured on the ASBR (not an NSSA ABR translator), **summary-address** summarizes only redistributed routes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure external route summarization.

summary-address *ipv4-address mask* [**not-advertise** | [**tag** *tag-value* | **advertise-metric** *cost* | **distribute-delay** *interval* | **nssa-only**] *]

The route summarization function is disabled by default.

1.8.6 Configuring a Discard Route

1. Overview

The address range of summarized routes may exceed the actual network range in the routing table. If data is sent to a network beyond the summarization range, a routing loop may be formed and the router processing load may increase. To prevent these problems, a discard route must be added to the routing table on the ABR or ASBR. This route is automatically generated, and is not advertised.

The discard route means adding a route whose egress is Null 0 to the routing table, and it will not generate the corresponding LSA, so it will not be received by neighbors. The packets matching the route are discarded.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure a discard route.

discard-route-reinstall { **external-routes** | **internal** }

The discard route adding function is enabled by default.

1.9 Configuring OSPF Route Filtering

1.9.1 Overview

Routes that do not meet filtering conditions cannot be loaded to the routing table, or advertised to neighbors. Network users cannot access specified destination network.

1.9.2 Restrictions and Guidelines

The OSPF basic functions must be configured.

- There are three ways to implement route filtering:
  - Run the **disable-interface** command to configure the interface as a passive interface so that the interface no longer receives and sends Hello packets.
  - Run the **route-filtering out** command when the ASBR imports external routes, and you want to filter the imported external routes.
  - If it is necessary to restrict users' access to the specified destination network, run the **route-filtering in** command to filter the routes calculated by the received LSA. Only the routes that meet the filtering conditions can be used for packet forwarding. It does not affect the link state database, nor does it affect the

neighbor's routing table.

1.9.3 Configuration Tasks

The configuration of OSPF route filtering includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Configuring a Passive Interface
- Disabling Interfaces from Sending OSPF Packets
- Configuring Inter-area Route Filtering
- Configuring Redistributed Route Filtering
- Configuring Learned Route Filtering

1.9.4 Configuring a Passive Interface

1. Overview

To prevent other routers on the network from learning the routing information of the local router, you can configure a specified network interface of the local router as the passive interface, or a specified IP address of a network interface as the passive address. A passive interface cannot send and receive OSPF packets.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure a passive interface.

disable-interface { **default** | *interface-type interface-number* | *interface-type interface-number ipv4-address* }

By default, the passive modes of all interfaces are disabled, and all interfaces are allowed to send and receive OSPF packets.

1.9.5 Disabling Interfaces from Sending OSPF Packets

1. Overview

Enable this function on an interface to prevent sending the LSA update packet from this interface. After this function is enabled, the local router does not advertise the LSA update packet to neighbors, but still sets up an adjacency with neighbors and receives LSAs from neighbors.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Disable the interface from sending OSPF packets.

ip ospf outgoing-lsa filter out

By default, the function of diffusing LSA packets to the outside is enabled on an interface, that is, any LSA update packet can be sent from the interface.

1.9.6 Configuring Inter-area Route Filtering

1. Overview

- Configuring inter-area route filtering on an ABR can filter Type3 LSAs, which can restrict route learning between different areas, and restrict users' access to the designated OSPF area network.

2. Restrictions and Guidelines

- This command can be configured only on an ABR.
- Unless otherwise required, perform this configuration on an ABR in the area where filtered routes are located.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ACL or prefix list. Please configure only one task.

- Configure an ACL.

  access-control-list *acl-number* { **deny** | **permit | comment | list-comment | sequence number**} { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } [**time-range** *tm-range-name*] [**log**]

- Configure a prefix list.

  ip prefix-list-info *prefix-list-name* [**seq** *seq-number*] { **deny** | **permit | description** } *ipv4-prefix* [**ge** *minimum-prefix-length*] [**le** *maximum-prefix-length*]

(4) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(5) Configure inter-area route filtering.

area *area-id* **filter-rule-list** { **access** *acl-number* | **prefix** *prefix-list-name* } { **in** | **out** }

By default, the filtering list function is not configured on an ABR.

1.9.7 Configuring Redistributed Route Filtering

1. Overview

The redistributed route filtering function refers to filtering routes redistributed from other protocols to OSPF. The route-filtering out command is similar to the redistribute-route route-map command, but it does not redistribute routes, and is generally used together with the redistribute-route command.

2. Restrictions and Guidelines

- Perform this configuration if external routes introduced by the ASBR need to be filtered.
- Unless otherwise required, perform this configuration on an ASBR to which filtered routes are introduced.
- The ACL and prefix list filtering rules are mutually exclusive in the configuration. In other words, if the ACL is used for filtering routes coming from a certain a source, the prefix list cannot be configured to filter the same routes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ACL or prefix list. Please configure only one task.

- Configure an ACL.

  access-control-list *acl-number* { **deny** | **permit | comment | list-comment | sequence number** } { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } [**time-range** *tm-range-name*] [**log**]

- Configure a prefix list.

  ip prefix-list *prefix-list-name* [**seq** *seq-number*] { **deny** | **permit** } *ipv4-prefix* [**ge** *minimum-prefix-length*] [**le** *maximum-prefix-length*]

(4) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(5) Configure redistributed route filtering.

route-filtering { [*acl-name* | *acl-number*] | **prefix-filter** *prefix-list-name* } out [**arp-host** | **bgp** | **connected** | **isis** [*area-tag*] **| ospf** *process-id* **| rip | static**]

By default, the filtering function of redistributed routes is disabled, that is, all the redistributed routes get past.

1.9.8 Configuring Learned Route Filtering

1. Overview

This function filters the routes that are computed based on received LSAs. Only the routes meeting the filtering conditions can be forwarded.

2. Restrictions and Guidelines

Filtering routes by using the route-filtering in command affects forwarding of local routes only, but does not affect route computation based on LSAs. Therefore, if route filtering is configured on the ABR, Type-3 LSAs will still be generated and advertised to other areas because routes can still be computed based on LSAs. As a result, black hole routes are generated. In this case, you can run the area filter-rule-list or area summary (containing the not-advertise parameter) command on the ABR to prevent generation of black hole routes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ACL or prefix list. Please configure only one task.

- Configure an ACL.

  access-control-list *acl-number* { **deny** | **permit | comment | list-comment | sequence number** } { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } [**time-range** *tm-range-name*] [**log**]

- Configure a prefix list.

  ip prefix-list-info *prefix-list-name* [**seq** *seq-number*] { **deny** | **permit | descripton** } *ipv4-prefix* [**ge** *minimum-prefix-length*] [**le** *maximum-prefix-length*]

(4) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(5) Configure learned route filtering.

route-filtering { [*acl-name* | *acl-number*] | **gateway-filter** *prefix-list-name* | **prefix-filter** *prefix-list-name* [**gateway** *prefix-list-name*] | **route-map** *route-map-name* } **in** [*interface-type interface-number*]

By default, the filtering function of the routes calculated based on the received LSA is disabled, that is, all these routes get passed.

1.10 Adjusting OSPF Network Convergence Speed

1.10.1 Overview

This function aims to modify protocol control parameters to change the protocol running status, thus adjusting the convergence speed of OSPF network.

1.10.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- The neighbor relationship maintenance time cannot be shorter than the hello interval.

1.10.3 Configuration Tasks

The configuration for adjusting the convergence speed of OSPF network includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Configuring the Hello Interval
- Configuring the Dead Interval
- Configuring LSU Transmission Delay
- Configuring LSU Retransmission Interval
- Configuring LSU Generation Time
- Configuring LSA Group Refresh Time

- Configuring LSA Group Refresh Interval
- Configuring Duplicate LSA Receiving Delay
- Configuring Inter-Area Route Computation Delay
- Configuring External Route Computation Delay
- Configuring SPF Computation Delay

1.10.4 Configuring the Hello Interval

1. Overview

The hello packet interval is contained in the hello packet. A shorter hello interval indicates that OSPF can detect topology changes more quickly, but the network traffic increases.

2. Restrictions and Guidelines

- This configuration is performed on routers at both ends of a link.
- The hello packet interval must be the same on all routers in the same network segment. If you want to manually modify the neighbor dead interval, ensure that the neighbor dead interval is longer than the hello packet interval.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*

- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure the hello interval.

ip ospf hello-period *hello-interval*

By default, the hello packet interval is 10s for Ethernet, HDLC encapsulation interfaces, and frame relay point-to-point sub-interfaces; the hello packet interval is 30s for non-frame relay point-to-point sub-interfaces and X.25 interfaces.

1.10.5 Configuring the Dead Interval

1. Overview

This configuration can be adjusted if you wish to accelerate OSPF convergence when a link fails. The failure determining interval of an OSPF neighbor is contained in the hello packet. If OSPF does not receive a hello packet from a neighbor within the neighbor dead interval, it declares that the neighbor is invalid and deletes this neighbor record from the neighbor list.

2. Restrictions and Guidelines

- This configuration is performed on routers at both ends of a link.
- By default, the dead interval is four times the hello packet interval. If the hello packet interval is modified, the dead interval is modified automatically.
- Run this command to manually modify the interval for OSPF to judge failure of a neighbor with caution. Pay attention to the following two issues:
  - The neighbor dead interval cannot be smaller than the hello packet sending interval;
  - The neighbor dead interval must be the same on all routers in the same network segment.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.

- **interface aggregateport** *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure the dead interval.

ip ospf dead-period { *dead-period* | **minimum hello-multiplier** *multiplier* }

By default, the fast hello function is disabled, and the neighbor dead interval is four times the sending interval of hello packet.

1.10.6 Configuring LSU Transmission Delay

1. Overview

On a low speed network, a device can be configured with the delay before sending LSU packets.

2. Restrictions and Guidelines

- Before an LSU packet is transmitted, the **Age** fields in all LSAs in this packet will increase based on the amount specified by the LSU transmission delay. Considering the transmission delay and line propagation delay on the interface, you need to set the LSU transmission delay to a greater value for a low-speed line or interface. The LSU packet transmission delay of a virtual link is defined by the **transmit-lsu-delay** parameter in the **area vlink** command.
- If the value of the **Age** field of an LSA reaches 3600, the packet will be retransmitted or a retransmission request will be sent. If the LSA is not updated in time, the expired LSA will be deleted from the LSDB.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure the LSU transmission delay.

ip ospf transmit-lsu-delay *transmit-delay*

The LSU packet transmission delay is 1s by default.

1.10.7 Configuring LSU Retransmission Interval

1. Overview

When a network is congested, frequent LSU retransmission will aggravate the congestion. The retransmission delay can be set to a greater value on a serial line or virtual link to prevent unwanted retransmission. The LSU retransmission interval of a virtual link is defined by the retransmit-period parameter in the area vlink command.

After a router finishes sending an LSU packet, this packet is still kept in the transmit buffer queue. If an acknowledgment from the neighbor is not received within the LSU retransmission interval, the router retransmits the LSU packet.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure the LSU retransmission interval.

ip ospf retransmit-period *retransmit-interval*

By default, the LSU retransmission interval is 5s.

1.10.8 Configuring LSU Generation Time

1. Overview

The LSU generation time can be adjusted according to different requirements for network performance and convergence speed.

2. Restrictions and Guidelines

- If a high convergence requirement is raised when a link changes, you can set *delay-time* to a smaller value. You can also appropriately increase the values of *delay-time, hold-time* and *max-wait-time* to reduce the CPU usage.
- When configuring this command, the value of *hold-time* cannot be smaller than the value of *delay-time*, and the value of *max-wait-time* cannot be smaller than the value of *hold-time*.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the exponential backoff algorithm for generating LSAs.

timers-set throttling lsa all *delay-time hold-time max-wait-time*

By default, the minimum delay of LSA generation is 0 ms, the minimum interval between the first update and the second update of LSA is 5000 ms, and the maximum interval between consecutive LSA updates is 5000 ms.

1.10.9 Configuring LSA Group Refresh Time

1. Overview

Every LSA has its time to live (LSA age). When the LSA age reaches 3600s, a refreshment is needed to prevent normal LSAs from being cleared because their ages reaching the maximum. If LSA update and aging computation are performed for every LSA, the device will consume a lot of CPU resources. To use CPU resources effectively, you can refresh LSAs by group on the device. Allocate the LSAs received within a certain time (pacing interval) to the same group. The device will maintain a refresh timer for this group. After the timer expires, the LSAs in the group will be refreshed uniformly.

2. Restrictions and Guidelines

If the total number of LSAs does not change, a larger group pacing interval indicates that more LSAs need to be processed after timeout, which will lead to a high load of the CPU in a certain period of time. For CPU stability, the number of LSAs processed upon each timeout cannot be too large. If the number of LSAs is large, you are advised to reduce the group pacing interval. For example, if there are 10,000 LSAs in the database, you can reduce the pacing interval; if there are 40 to 100 LSAs, you can set the pacing interval to 10-20 minutes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the LSA group refresh time.

timers-set packet-pacing lsa-batch *pacing-interval*

The group pacing interval of LSA is 30s by default.

1.10.10 Configuring LSA Group Refresh Interval

1. Overview

If a router has a large number of routing entries, synchronizing LSDB will send a large number of LSU packets, easily causing network congestion. The LSU packets can be grouped and sent at an interval to ensure stability of the network.

2. Restrictions and Guidelines

- If the number of LSAs is large and the device load is heavy in an environment, properly configuring *transimit-time* and *transimit-count* can limit the number of LS-UPD packets flooded on the network.
- If the CPU usage is not high and the network bandwidth load is not heavy, reducing the value of *transimit-time* and increasing the value of *transimit-count* can accelerate the environment convergence.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the LSA group refresh interval.

timers-set packet-pacing lsa-transmit *transmit-interval transmit-count*

By default, the interval for sending LSU group is 40 ms, and the number of LS-UPD packets in each group is 1.

1.10.11 Configuring Duplicate LSA Receiving Delay

1. Overview

To prevent excessive usage of network bandwidth and device resources caused by frequent network connection or routing turbulence, you can configure the device to not process duplicate LSAs within a specified time according to the network connection and device resources in the current network.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure duplicate LSA receiving delay.

timers-set lsa repeat-delay *arrival-time*

By default, the delay for receiving a duplicate LSA is 1000 ms.

1.10.12 Configuring Inter-Area Route Computation Delay

1. Overview

In case of route flapping, frequent inter-area route computation will consume a lot of device performance, and delaying the process of inter-area route computation can reduce the device load.

2. Restrictions and Guidelines

- This delay cannot be modified if strict requirements are raised for the network convergence time.
- If a lot of inter-area or external routes exist on the network and the network is not stable, you can adjust the corresponding delays.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure inter-area route computation delay.

timers-set throttling route inter-area *ia-delay*

By default, the waiting time for inter-area route calculation is 0 ms.

1.10.13 Configuring External Route Computation Delay

1. Overview

In case of route flapping, frequent external route computation will consume a lot of device performance, and delaying the process of external route computation can reduce the device load.

2. Restrictions and Guidelines

- This delay cannot be modified if strict requirements are raised for the network convergence time.

- If a lot of inter-area or external routes exist on the network and the network is not stable, you can adjust the corresponding delays.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the external route computation delay.

timers-set throttling route ase *ase-delay*

By default, the waiting time for external route calculation is 0 ms.

1.10.14 Configuring SPF Computation Delay

1. Overview

Changes to LSDB will trigger SPF computation. Frequent network jitter will consume a lot of CPU resources. Setting a reasonable delay for SPF computation can avoid occupying excessive router memory and bandwidth resources.

2. Restrictions and Guidelines

- Compared with the **timers-set spf** command, the **timers-set throttling spf** command supports more flexible settings to accelerate the convergence speed of SPF computation and further reduce the system resources consumed by SPF computation when the topology continuously changes. Therefore, you are advised to run the **timers-set throttling spf** command to adjust the SPF computation time.
- The value of *spf-holdtime* cannot be smaller than that of *spf-delay*; otherwise, *spf-holdtime* will be automatically set to the value of *spf-delay*.
- The value of *spf-max-waittime* cannot be smaller than that of *spf-holdtime*; otherwise, *spf-max-waittime* will be automatically set to the value of *spf-holdtime*.
- The configurations of **timers-set throttling spf** and **timers-set spf** are mutually overwritten.
- When neither **timers-set spf** nor **timers-set throttling spf** is configured, the default value of **timers-set throttling spf** prevails.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the SPF computation delay. Select one of the following items to configure.

- **timers-set throttling spf** *spf-delay spf-holdtime spf-max-waittime*

By default, the delay for SPF computation is 50 ms, the minimum interval for two SPF calculations is 200 ms, and the maximum interval between two SPF calculations is 5000 ms.

- **timers-set spf** *spf-delay spf-holdtime*

- By default, the **timers-set spf** command does not take effect. The SPF calculation delay is subject to the default configuration of the **timers-set throttling spf** command.

1.10.15 Configuring Route Calculation Suppression Time

1. Overview

When devices are started, the devices have different route calculation time due to differences in device performance and database size after a neighbor is in full neighbor relationship. This may cause abnormal traffic interruption in the routing table. Therefore, it is necessary to suppress route calculation of an interface after the interface is in full neighbor relationship state, so as to ensure that calculated routes are the latest.

2. Restrictions and Guidelines

- After route calculation suppression is enabled on an interface, the interface is suppressed from participating in route calculation when the first neighbor of the interface is in full neighbor relationship. The interface recovers after the suppression period expires. This function is usually used in combination with the **max-metric** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) The following example configures the route calculation suppression time after the first neighbor of an interface is in full neighbor relationship.

ip ospf suppress neighborup *second*

Route calculation is not suppressed by default.

1.10.16 Enabling Two-Way Maintenance

1. Overview

On a large network, a lot of packets may be sent or received, occupying too much CPU and memory. As a result, some packets are delayed or discarded. If the time required for processing hello packets exceeds the neighbor dead interval, the corresponding adjacency times out and is removed. If the two-way maintenance function is enabled, in addition to the hello packets, the DD, LSU, LSR, and LSAck packets from a neighbor can also be used to maintain the bidirectional communication between neighbors when a large number of packets exist on the network. This prevents termination of the adjacency caused by delayed or discarded hello packets. The function is enabled by default. You are advised to retain the default configuration.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Enable two-way maintenance.

two-way-maintain

By default, the two-way maintenance function is enabled for an OSPF process.

1.11 Configuring OSPF Authentication

1.11.1 Overview

Authentication prevents routers that illegally access the network and hosts that forge OSPF packet from participating in the OSPF process. All routers connected to the OSPF network must be authenticated to ensure stability of OSPF and protect OSPF against intrusions.

1.11.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- If authentication is configured for an area, the configuration takes effect on all interfaces that belong to this area.
- If authentication is configured for both an interface and the area to which the interface belongs, the configuration for the interface takes effect preferentially.

1.11.3 Configuration Tasks

The OSPF authentication configuration includes the following tasks:

- Configuring the Authentication Type of an Area
- Configuring the Authentication Type of an Interface
- (Optional) Configuring a Plain Text Authentication Key for an Interface
- (Optional) Configuring an MD5 Authentication Key for an Interface

1.11.4 Configuring the Authentication Type of an Area

1. Overview

OSPF area authentication can improve the security of an OSPF area. The consistency in the authentication type and password is a requisite for establishing a neighbor relationship. If the same authentication type must be enabled on all interfaces in an area, you are advised to make the configuration through the area authentication type. If the router accesses the network that requires authentication, then area authentication or interface authentication must be configured.

2. Restrictions and Guidelines

- The device supports three authentication types:
  - Type 0: No authentication is required. When the OSPF authentication command is not run to enable OSPF authentication, the authentication type carried in OSPF data packets is type 0.
  - Type 1: The authentication type is plain text authentication if this command is configured but does not contain the **auth-secret** parameter.
  - Type 2: The authentication type is MD5 authentication if this command is configured and contains the **auth-secret** parameter.
- All routers in the same OSPF area must use the same authentication type. If authentication is enabled, the authentication key must be configured on interfaces that are connected to neighbors.
- If keychain authentication is configured, the authentication key must be configured using the **key chain** command. Currently, keychain supports plain-text authentication.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the authentication type of an area.

area *area-id* **auth-algorithm** [**keychain** *keychain-name* | **message-digest**]

The OSPF area authentication function is disabled by default.

1.11.5 Configuring the Authentication Type of an Interface

1. Overview

- OSPF interface authentication can improve the communication security of neighbors. The consistency in the authentication type and password is a requisite for establishing a neighbor relationship. This configuration is mandatory if different authentication types are enabled on different interfaces in the same area. If the router accesses the network that requires authentication, then area authentication or interface authentication must be configured.

2. Restrictions and Guidelines

- If the **ip ospf auth-algorithm** command does not contain any option, plain text authentication is enabled. If the authentication type is set to **null**, authentication is disabled forcibly. When authentication is configured for both an interface and the area to which the interface belongs, the authentication type configured for the interface is used preferentially.
- If keychain authentication is configured, the key and authentication type configured for keychain are used. Currently, keychain supports plain text authentication, MD5 authentication, and SM3 authentication.
- If area authentication has been configured, the interface authentication mode is determined by the area authentication configuration when the **unset** option is used to restore the interface authentication mode to the default value.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*

- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure the authentication type of an interface.

ip ospf auth-algorithm [**keychain** *keychain-name* | **message-digest** | **null**]

By default, no authentication mode is set on an interface. In this case, the authentication type of the related area is used on the interface.

1.11.6 Configuring a Plain Text Authentication Key for an Interface

1. Overview

This configuration is required when a router accesses a network that requires plain text authentication. The key information will be inserted to the headers of all OSPF packets. If the keys are inconsistent, two directly connected devices cannot set up an OSPF neighbor relationship and cannot exchange routing information.

2. Restrictions and Guidelines

- Different keys can be configured for different interfaces, but all the routers connected to the same physical network segment must be configured with the same key.
- You must set the authentication type to plain text authentication on the related interface or in the area in advance.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure plain text authentication for an interface.

ip ospf auth-algorithm

(5) Configure a plain text authentication key for an interface.

ip ospf auth-algorithm-key [**0** | **7**] *key*

The authentication key is disabled by default.

1.11.7 Configuring an MD5 Authentication Key for an Interface

1. Overview

The MD5 cipher text authentication mode can be used to enhance security to a certain extent.

2. Restrictions and Guidelines

- MD5 authentication features high security, and therefore is recommended. You must configure either plain text authentication or MD5 authentication. The HMAC-SHA256 algorithm features the highest security and consumes the maximum performance, while the MD5 algorithm features the lowest security and consumes the minimum performance. Select one according to the actual situation.
- This configuration is required if a router accesses a network that requires MD5 authentication.
- You must set the authentication type to cipher text authentication on the related interface or in the area in advance.
- Authentication succeeds only when both *key-id* and key are matched.
- Multiple *key-id* values are configured for the same interface, and the last *key-id* configured takes effect when packets are sent, but all *key-id* values can participate in the authentication when packets are received.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure cipher text authentication for an interface.

ip ospf auth-algorithm message-digest

(5) Configure an MD5 authentication key for an interface.

ip ospf auth-secret-key *key-id* **md5** [**0** | **7**] *key*

The cipher text key function is disabled by default.

1.12 Enabling Fast Reroute

1.12.1 Overview

After the fast reroute function is enabled, once OSPF detects a route failure, the router can immediately switch to the second-best route, helping shorten the traffic interruption time.

1.12.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.

- The LAF configuration for fast reroute is mutually exclusive with the virtual link configuration.
- The **carrier-delaytime** value must be set to **0** for an interface.

1.12.3 Configuration Tasks

The configuration for enabling fast reroute includes the following tasks:

- Configuring Fast Reroute
- (Optional) Configuring Interface LFA Protection
- (Optional) Preventing an Interface From Becoming a Backup Interface

1.12.4 Configuring Fast Reroute

1. Overview

Fast reroute will prepare a backup route for OSPF. Once OSPF detects a neighbor failure, the router will immediately forward the traffic to the backup route to avoid traffic interruption. This configuration is required if you want to increase the OSPF network convergence speed to the millisecond level.

2. Restrictions and Guidelines

- This configuration is performed on a router that has multiple paths to a destination network.
- It is recommended that automatic computation of loop-free alternate (LFA) for the backup path be disabled if any of the following cases exists on the network:
  - Virtual links exist.
  - Alternative ABRs exist.
  - An ASBR is also an ABR.
  - Multiple ABSRs advertise the same external route.
- When the OSPF fast reroute function is used, it is recommended that BFD be enabled at the same time so that the device can quickly detect any link failure and therefore shorten the forwarding interruption time. If the interface is Up or Down, to shorten the forwarding interruption time during OSPF fast reroute, you can configure **carrier-delaytime 0** in Layer-3 interface configuration mode to achieve the fastest switchover speed.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure fast reroute.

frr { **lfa** [**downstream-paths**] | **route-map** *route-map-name* }

The fast reroute function is disabled by default.

1.12.5 Configuring Interface LFA Protection

1. Overview

A backup route is generated for the primary route based on the LFA protection mode specified in interface configuration mode. By default, LFA protection is enabled on each OSPF interface, and a failure of the primary link does not affect data forwarding on the backup route.

2. Restrictions and Guidelines

This command does not take effect if frr route-map is configured.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure interface LFA protection.

ip ospf frr protection { **node | link-node | disable** }

The LFA link protection function is enabled by default.

1.12.6 Preventing an Interface From Becoming a Backup Interface

1. Overview

This configuration is mandatory if you hope that data traffic is not switched over to a specified path after the best path fails. After the best path fails, the traffic will be switched over to another second-best path, but a new best path will be selected based on the interface costs after OSPF converges again. If you do not want to forward the data traffic to the specified path after the optimal path fails, you must disable an interface as a backup interface for fast rerouting. When the optimal path fails, the traffic will switch to other suboptimal paths. After the OSPF protocol finally converges again, a new optimal path will be selected according to the interface cost.

2. Restrictions and Guidelines

- If the remaining bandwidth of an interface is small or if the interface and its primary interface may fail at the same time, the interface cannot be used as a backup interface. Therefore, you need to run this command in interface configuration mode to prevent this interface from becoming a backup interface during OSPF fast reroute computation. After this command is executed, a backup interface will be selected from the other interfaces. If the remaining bandwidth of an interface is small or an interface may fail with the primary interface at the same time, the interface is not suitable as a backup interface. In this case, you need to disable the interface as a backup interface in the interface configuration mode of the interface. After an interface is disabled as a backup interface, the backup interface will be selected from other interfaces during OSPF fast rerouting calculation.
- This command does not take effect if **frr route-map** is configured.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Prevent an interface from becoming a backup interface.

ip ospf frr no-eligible-backup

By default, all interfaces can be used as a backup interface of OSPF fast reroute.

1.13 Modifying the Maximum Number of Concurrent Neighbors

1.13.1 Overview

Modifying the maximum number of concurrent neighbors can control the maximum number of concurrent neighbors on the OSPF process to ease the pressure on the device.

1.13.2 Restrictions and Guidelines

The OSPF basic functions must be configured.

1.13.3 Configuration Tasks

The configuration of modifying the maximum number of concurrent neighbors includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Configuring the Maximum Number of Concurrent Neighbors on the Current Process
- Configuring the Maximum Number of Concurrent Neighbors on All Processes

1.13.4 Configuring the Maximum Number of Concurrent Neighbors on the Current Process

1. Overview

When the performance of a router is affected because the router exchanges data with multiple neighbors, you can use this command to restrict the maximum of neighbors with which each OSPF process can concurrently initiate or accept interaction. This configuration is recommended if you want to set up an OSPF adjacency more quickly when a router is connected to a lot of other routers.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the maximum number of concurrent neighbors on the current process.

max-concurrent-dd *neighbor-num*

The maximum number of neighbors of a single OSPF routing process is 5 by default.

1.13.5 Configuring the Maximum Number of Concurrent Neighbors on All Processes

1. Overview

When the performance of a router is affected because the router exchanges data with multiple neighbors, you can configure this command to restrict the maximum of neighbors with which all OSPF processes can concurrently initiate or accept interaction. This configuration is recommended if you want to set up an OSPF adjacency more quickly when a router is connected to a lot of other routers.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of concurrent neighbors on all processes.

router-protocol ospf max-concurrent-dd *max-neighbor*

The maximum number of neighbors of all OSPF routing processes is 10 by default.

1.14 Configuring the GR Function

1.14.1 Overview

When a GR-enabled router is restarted on the control plane, data forwarding can be still guided on the forwarding plane. In addition, actions such as neighbor relationship re-forming and route computation performed on the control plane do not affect functions of the forwarding plane. In this way, service interruption caused by route flapping can be avoided, thus enhancing reliability of the entire network.

1.14.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- The neighbor router must support the GR helper.
- The GR time cannot be shorter than the neighbor relationship maintenance time of the neighbor router.

1.14.3 Configuration Tasks

The configuration for enabling the GR function includes the following tasks:

(1) Enabling the GR Function

(2) (Optional) Configuring the OSPF GR Helper Function

1.14.4 Enabling the GR Function

1. Overview

The GR function is configured based on the OSPF process. You can configure different parameters for different OSPF processes based on the actual conditions. This command is used to configure the GR restarter capability of a device. The grace period is the maximum time of the entire GR process, during which link state is rebuilt so that the original state of the OSPF process is restored. After the GR period expires, OSPF exits the GR state and performs common OSPF operations.

The precondition for successful execution of GR and uninterrupted forwarding is that the topology remains stable. If the topology changes, OSPF quickly converges without waiting for further execution of GR, thus avoiding long-time forwarding black hole. When the topology detection function is enabled, the device can detect the topology changes during GR.

2. Restrictions and Guidelines

- You are advised to retain the default configuration.
- If the topology detection function is disabled, the topology change in the GR process cannot be converged in time, which may cause a long-time forwarding black hole.
- If the topology detection function is enabled, forwarding may be interrupted when the topology changes, but the interruption time is far shorter than that when the topology detection function is disabled.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Enable the opaque LSA processing capability.

capability opaque

The opaque LSA processing capability is enabled by default.

(5) Enable the GR function.

gr [**grace-period** *grace-period* | **inconsistent-lsa-checking | helper**]

The GR capability is enabled by default.

1.14.5 Configuring the OSPF GR Helper Function

1. Overview

When a neighbor router implements GR, GR Helper advertises a Grace-LSA to all neighbor routers. If the GR helper function is enabled on the local router, the local router becomes the GR helper upon receiving the Grace-LSA, and helps the neighbor to complete GR. The disable option indicates that the GR helper is not provided for any device that implements GR.

2. Restrictions and Guidelines

- You are advised to retain the default configuration.
- After a device becomes a GR helper, the network changes are not detected by default. If any change takes place on the network, the network topology converges after GR is completed. If you hope that network changes can be quickly detected during the GR process, you can configure **check-rigid** to check Type-1 to Type-5 and Type-7 LSAs that indicate the network information or **internal-lsa-checking** to check Type-1 to Type-3 LSAs that indicate internal routes of the AS domain. When the network scale is large, it is recommended that you disable the LSA checking options (strict-lsa-checking and internal-lsa-checking) because regional network changes may trigger termination of GR and consequently reduce the convergence of the entire network.
- The **disable** option indicates that the GR helper is not provided for any device that implements GR.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the OSPF GR helper function.

gr helper { **disable** | **internal-lsa-checking** | **check-rigid** }

The GR helper capability is enabled by default. After the GR helper is enabled on the device, LSA changes are not checked.

1.15 Correlating OSPF with BFD

1.15.1 Overview

After OSPF is correlated with BFD, once a link is faulty, OSPF can quickly detect the failure of the route. This configuration helps shorten the traffic interruption time.

1.15.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- The BFD parameters must be configured for the interface in advance.
- If correlation with BFD is configured for both the OSPF process and an interface, the interface-based configuration takes effect preferentially.
- The configuration must be performed on routers at both ends of the link.

1.15.3 Configuration Tasks

The configuration for correlating OSPF with BFD includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Configuring BFD Correlation with OSPF on an Interface
- Configuring BFD Correlation with OSPF Globally

1.15.4 Configuring BFD Correlation with OSPF on an Interface

1. Overview

Correlating OSPF with BFD on a specified interface can quickly detect the interface fault and accelerate OSPF convergence.

2. Restrictions and Guidelines

- The BFD configured on an interface takes precedence over that configured in OSPF process configuration mode.
- Based on the actual environment, you can run the **ip ospf set-bfd** command to enable BFD on a specified interface for link detection, or run the **bfd whole-interfaces** command in OSPF process configuration mode to enable BFD on all interfaces of the OSPF process, or run the **ip ospf set-bfd disable** command to disable BFD on a specified interface.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure BFD function on an OSPF-enabled interface.

ip ospf set-bfd [**disable**]

By default, the BFD function is disabled on an interface, and the BFD configuration is subject to that configured in RIP process configuration mode.

1.15.5 Configuring BFD Correlation with OSPF Globally

1. Overview

OSPF dynamically discovers neighbors through the hello packets. After OSPF enables the BFD function, a BFD session will be set up to achieve the full adjacency, and the BFD mechanism is used to detect the neighbor state. Once a neighbor failure is detected through BFD, OSPF performs network convergence immediately.

2. Restrictions and Guidelines

You can also run the ip ospf set-bfd [disable] command in interface configuration mode to enable or disable the BFD function on a specified interface, and this configuration takes precedence over the bfd whole-interfaces command used in OSPF process configuration mode.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Enable the BFD function on all interfaces in the OSPF process.

bfd whole-interfaces

By default, the BFD function is disabled on all the interfaces.

1.16 Enabling Overflow

1.16.1 Overview

When the overflow function is enabled, new routes are not loaded to routers when the router memory is insufficient or the usage of the database space reaches the upper limit.

1.16.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- After a router enters the overflow state, you can run the **cleanup ip ospf instance** command, or stop and then restart the OSPF protocol to exit the overflow state.

1.16.3 Configuration Tasks

The configuration of enabling overflow includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Configuring the Memory Overflow Function
- Configuring the Database Overflow Function
- Configuring the External LSA Database Overflow Function

1.16.4 Configuring the Memory Overflow Function

1. Overview

The OSPF process enters the overflow state to discard newly-learned external routes. This behavior can effectively prevent the memory usage from increasing.

2. Restrictions and Guidelines

- This configuration is recommended if a large number of routes exist in the domain and may cause insufficiency of the router memory.
- After the overflow function is enabled, the OSPF process enters the overflow state and discards newly-learned

external routes, which may cause a routing loop on the entire network. To reduce the occurrence probability of this problem, OSPF generates a default route to the null interface, and this route always exists in the overflow state.

- You can run the **cleanup ip ospf instance** command to reset the OSPF process so that the OSPF process can exit the overflow state. You can run the **unset** form of the command to prevent the OSPF process from entering the overflow state when the memory is insufficient. This, however, may lead to over-consumption of the memory resource, after which the OSPF process will stop and delete all the learned routes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the memory overflow function.

overflow memory-lack

By default, the OSPF process is allowed to enter the overflow state when the memory is insufficient.

1.16.5 Configuring the Database Overflow Function

1. Overview

This function controls the maximum number of routes in LSDB to avoid abnormal operation of a router due to insufficient memory.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the database overflow function.

overflow database *max-lsa* [**hard** | **soft**]

The maximum number of LSAs is not configured by default.

1.16.6 Configuring the External LSA Database Overflow Function

1. Overview

This function controls the maximum number of external routes in LSDB to avoid abnormal operation of a router due to insufficient memory.

2. Restrictions and Guidelines

- This configuration is recommended if the ASBR introduces a large number of external routes and the router memory may be insufficient.
- When using the overflow function, ensure that the same *max-dbsize* is configured on all routers in the OSPF backbone area and common areas; otherwise, the following problems may occur:
  - The LSDBs throughout the network are inconsistent, and the neighbor relationship fails to reach the full state.
  - Routes are incorrect, including routing loops.
  - AS external LSAs are frequently retransmitted.
- If the maximum number of external LSAs is set, when the number of external LSAs exceeds the maximum, the normal state will not be restored.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the external LSA database overflow function.

overflow database external *max-dbsize wait-time*

The maximum number of external LSAs is not configured by default.

1.17 Configuring VPN Extended Feature of OSPF

1.17.1 Overview

When OSPF is enabled on a VPN client network, OSPF runs between the PE and CE to simplify the configuration and management of CE.

1.17.2 Restrictions and Guidelines

The router ID of Label Distribution Protocol (LDP) must consist of 32 bits.

1.17.3 Configuration Tasks

The configuration of VPN extended feature of OSPF includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Configuring Domain ID
- Configuring a Domain Label
- Configuring a Sham Link
- Disabling Loop Detection for a VRF OSPF Process
- Configuring Extcommunity Attribute for a VPN Route
- Disabling Loop Detection Using the DN Bit of LSA
- Disabling Loop Detection Using the Route Tag of LSA

1.17.4 Configuring Domain ID

1. Overview

Domain ID is used to indicate the domain information of an OSPF process.

A VRF OSPF process can be configured with multiple domain IDs, but only one is the primary domain ID, and the remaining ones are secondary domain IDs. When an OSPF route is converted into a VPN route for advertising, only the primary domain ID is included in a VPN route.

2. Restrictions and Guidelines

- This function takes effect only for the OSPF process associated with VRF.
- The domain IDs of different VRF OSPF processes can be the same and do not influence each other.
- VRF OSPF processes that belong to the same VPN must be configured with the same domain ID to ensure the correctness of route advertising.
- You are advised to configure all the VRF OSPF processes that belong to the same VPN with the same domain ID.

3. Prerequisites

Before configuring the VPN feature, complete the VRF configuration.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* **vrf-instance** *vrf-name*

(4) Configure a domain ID.

domain-id { *ipv4-address* [secondary] | null | type { 0005 | 0105 | 0205 | 8005 } value *hex-value* [secondary] }

By default, the domain ID value of an OSPF process is **null**, and the type value is 0x0005.

1.17.5 Configuring a Domain Label

1. Overview

Domain label is the VPN routing tag of the VRF OSPF process, and is used to detect loops between a PE and a CE running the OSPF protocol. If a VPN site is connected to multiple PEs, the VPN route is learned from the PE through MP-BGP. If the route is advertised to the VPN site through a Type-5 or Type-7 LSA, it may be learned and advertised by other PE routers connected to the VPN site, which may cause a loop. To prevent the preceding loop, the VRF OSPF processes connected to the same VPN site on the PEs are configured with the same VPN route tag. When a VRF OSPF process sends a Type-5 or Type-7 LSA to a VPN site, the VPN route tag information will also be attached to the LSA. When other PE sites receive such a Type-5 or Type-7 LSA, if they detect that the VPN route tag in the LSA is consistent with that of the local OSPF process, the LSA will not participate in the OSPF calculations. Generally, the OSPF processes belonging to the same VPN site are configured with the same tag value.

The VPN route tag occupies four bytes in an OSPF packet. If a VRF OSPF process is not configured with this command, when the OSPF process advertises a Type-5 or Type-7 LSA, the first two bytes of the VPN route tag are set to 0xD000 by default, and the last two bytes are the AS number of local BGP. For example, if the AS number of the local BGP is 1, the hexadecimal form of the VPN route tag is 0xD0000001.

2. Restrictions and Guidelines

This function takes effect only for the OSPF process associated with VRF and BGP redistributed routes.

3. Prerequisites

Before configuring the VPN feature, complete the VRF configuration.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *ospf-id* **vrf-instance** *vrf-name*

(4) Configure a VPN route tag.

domain-tag *tag*

By default, the default value of a VRF OSPF process is the AS number of the local BGP protocol.

1.17.6 Configuring a Sham Link

1. Overview

A sham link is used in an environment with backdoor links between VPN sites. If you still want VPN data to be transmitted through the MPLS backbone network in this case, you can establish a sham link between the VRF OSPF processes of two PEs. The VRF OSPF processes of two PEs can set up an OSPF neighbor relationship through the sham link and distribute LSA packets on the sham link.

2. Restrictions and Guidelines

- The area ID of sham link configured for the two PEs must be consistent.
- The combination of source address and destination address of a sham link configured on one PE must be equal to the combination of destination address and source address of a sham link configured on the other PE.
- Both the source address and destination address used to establish a sham link on the PEs must be a 32-bit loopback address bound to the corresponding VRF.
- Since the OSPF route advertised through a sham link has no VPN label, the route cannot be used for forwarding. Actually packets are still forwarded through the BGP VPNv4 route. Therefore, in the actual configuration, you must ensure that the route advertised through the sham link will also be learned through MP-BGP.
- The source address for establishing the sham link must participate in the BGP VPNv4 route advertisement, but cannot join the calculation of VRF OSPF process.

3. Prerequisites

Before configuring the VPN feature, complete the VRF configuration.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *ospf-id* **vrf-instance** *vrf-name*

(4) Configure a sham link.

area *area-id* **sham-link** *source-ipv4-address destination-ipv4-address* [[**auth-algorithm** [**keychain** *kechain-name* | **message-digest** | **null**] | **metric** *number* | **dead-period** *dead-period* | **hello-period**

hello-interval | **retransmit-period** *retransmit-interval* | **transmit-lsu-delay** *transmit-delay*] * | **auth-algorithm-key** [**0** | **7**] *key* | **auth-secret-key** *key-id* **md5** [**0** | **7**] *key*]

No sham link is configured by default.

1.17.7 Disabling Loop Detection for a VRF OSPF Process

1. Overview

Loop detection of OSPF processes aims to prevent possible loop of VPN routes in propagation.

2. Restrictions and Guidelines

- By default, the VRF OSPF process supports the PE-CE OSPF features (that is, conversion from domain ID to LSA, to DN bits, and to VPN route tag). If you do not want a VRF OSPF process to support the PE-CE OSPF features, disable the PE-CE OSPF features by running the **capability vrf-lite** command.
- This command takes effect only for the OSPF process associated with VRF.
- In some application scenarios, you may need to disable the loop detection function of VRF OSPF processes. For example, VPN users use MCE devices to exchange VPN routes with PEs, if the OSPF protocol runs between MCEs and PEs to exchange VPN routes, to enable the VPN site to learn the routes of other VPN sites, you need to run the **capability vrf-lite** command to disable the loop detection function of VRF OSPF processes on the MCE device.

3. Prerequisites

Before configuring the VPN feature, complete the VRF configuration.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *ospf-id* **vrf-instance** *vrf-name*

(4) Enable the loop detection function.

capability vrf-lite [**auto**]

By default, the loop detection is enabled for the OSPF processes associated with VRF.

1.17.8 Configuring Extcommunity Attribute for a VPN Route

1. Overview

When an OSPF route of VRF forms a VPN route, the Router-ID and Router-Type of the OSPF process will also be carried in the extcommunity attribute of the VPN route. To ensure compatibility with the implementations of different vendors, you can manually modify the Router-ID and Router-Type.

2. Restrictions and Guidelines

- When OSPF routes are redistributed to BGP to form VPN routes, the extcommunity attributes of OSPF routes, including the extcommunity attributes of **router-id** and **route-type**, will also be attached. By default, the extcommunity attribute type of **router-id** is 0x0107, and the extcommunity attribute of **route-type** is 0x0306.
- This function takes effect only for the OSPF process associated with VRF.

3. Prerequisites

Before configuring the VPN feature, complete the VRF configuration.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *ospf-id* **vrf-instance** *vrf-name*

(4) Configure the extcommunity type of a VPN route.

extcommunity-type { **router-id** { **0107 | 8001** } | **route-type** { **0306** | **8000** } }

By default, the type of **router-id** is 0107, and the type of **route-type** is 0306.

1.17.9 Disabling Loop Detection Using the DN Bit of LSA

1. Overview

In the CE dual-homing scenario of L3VPN, loop is avoided by suppressing the route computation of DN bit between PEs. However, in a specific scenario, PEs may be allowed to learn routes from each other without generating any loops. In this case, check of the DN bit can be cancelled by configuring this command. When a PE device is connected to an MCE device, the MCE device needs to calculate the route advertised by the PE and the DN bit will not be checked. Type-3, Type-5, and Type-7 LSAs of OSPF can all carry a DN bit.

2. Restrictions and Guidelines

This function takes effect only for the OSPF process associated with VRF.

3. Prerequisites

Before configuring the VPN feature, complete the VRF configuration.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *ospf-id* **vrf-instance** *vrf-name*

(4) Disable loop detection using the DN bit of LSA.

disable-dn-bit-check [**summary | ase | nssa**]

The DN bit loop detection function of LSA is enabled by default.

1.17.10 Disabling Loop Detection Using the Route Tag of LSA

1. Overview

In the CE dual-homing scenario of L3VPN, when the LSA route tag received by a PE is the same as its route tag, the route is not calculated, thus avoiding a loop. In a specific scenario, PEs are allowed to learn routes from each other without generating any loops. In this case, you can set different route tags for multiple PEs, or set to disable route tag check. When a PE device is connected to an MCE device, the MCE device needs to calculate the route advertised by the PE and the route tag will not be checked. Type-5 and Type-7 LSAs of OSPF can carry a route tag.

2. Restrictions and Guidelines

This command takes effect only for the OSPF process associated with VRF.

3. Prerequisites

Before configuring the VPN feature, complete the VRF configuration.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *ospf-id* **vrf-instance** *vrf-name*

(4) Disable loop detection using the route tag of LSA.

disable-tag-check

The loop detection function using the route tag of LSA is enabled by default.

1.18 Configuring Network Management Functions

1.18.1 Overview

Use the network management software to manage OSPF parameters and monitor the OSPF running status.

1.18.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- You must enable the MIB function of the SNMP-Server before enabling the OSPF MIB function.
- You must enable the Trap function of the SNMP-Server before enabling the OSPF Trap function.
- You must enable the logging function of the device before outputting the OSPF logs.

1.18.3 Configuration Tasks

The configuration of network management functions includes the following tasks, which are optional. Select tasks for configuration according to actual condition.

- Binding the MIB with an OSPF Process
- Configuring the Trap Function
- Configuring the Logging Function

1.18.4 Binding the MIB with an OSPF Process

1. Overview

- The OSPFv2 MIB does not have the OSPFv2 process information. Therefore, you must perform operations on a single OSPFv2 process through SNMP. By default, the OSPFv2 MIB is bound with the OSPFv2 process with the smallest process ID, and all user operations take effect on this process. If you want to manage the parameters of a specified OSPFv2 process through network management software, you must bind the MIB to the OSPFv2 process.

2. Restrictions and Guidelines

- If you want to perform operations on a specified OSPFv2 process through SNMP, run this command to bind the MIB to the process.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Bind the MIB with an OSPF process.

enable mib-binding

By default, the MIB is bound to the OSPFv2 process with the minimum process ID.

1.18.5 Configuring the Trap Function

1. Overview

This configuration is required if you want to use the network management software to monitor the OSPF running status.

2. Restrictions and Guidelines

- The function configured by this command is restricted by the **snmp-server** command. You need to configure the **snmp-server enable traps ospf** command and then the **enable traps** command before the corresponding OSPF trap can be correctly sent out.
- This command is not restricted by the MIB bound to the process. The trap switch can be enabled concurrently for different processes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the trap function.

enable traps [**error** [**ifauthfailure** | **ifconfigerror** | **ifrxbadpacket** | **virtifauthfailure** | **virtifconfigerror** | **virtifrxbadpacket**] * | **lsa** [**lsdbapproachoverflow** | **lsdboverflow** | **maxagelsa** | **originatelsa**] * | **retransmit** [**iftxretransmit** | **virtiftxretransmit**] * | **status-change** [**ifstatechange** | **nbrrestarthelperstatuschange** | **nbrstatechange** | **nssatranslatorstatuschange** | **restartstatuschange** | **virtifstatechange** | **virtnbrrestarthelperstatuschange** | **virtnbrstatechange**] *]

The trap message sending function is disabled by default.

1.18.6 Configuring the Logging Function

1. Overview

The logging function can be enabled to record the change log of the adjacent state to facilitate the analysis and search of relevant information. By default, the logging function is enabled. The log records the log information of the following four types of events only: the adjacency reaches the full state; the adjacency leaves the full state; the adjacency reaches the down state; the adjacency leaves the down state.

2. Restrictions and Guidelines

You are advised to retain the default configuration. If you want to reduce the log output, disable this function.

- If the **detail** parameter is configured, the change log of all adjacent states will be recorded.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the logging function.

log-adj-changes [**detail**]

The logging function is enabled by default, without the **detail** parameter.

1.19 Enabling NSR

1.19.1 Overview

If the NSR function is enabled, the device can maintain a stable adjacency state during the active/standby switchover, so as to ensure that data forwarding continues and is not interrupted.

1.19.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- For the same OSPF process, either NSR or GR is enabled because they are mutually exclusive. Nevertheless, when NSR is enabled, the GR helper capability is supported.
- The active/standby switchover takes a period of time. If the OSPF neighbor keepalive duration is shorter than the switchover duration, the OSPF neighbor relationship with the neighbor device is removed, and services are interrupted during the switchover. Therefore, it is recommended that the OSPF neighbor keepalive duration be no smaller than the default value when the NSR function is. If the fast hello function is enabled, the OSPF

neighbor keepalive duration is less than 1s. therefore, you are advised to disable the fast hello function when the NSR function is enabled.

1.19.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Enable the NSR function.

nsr

The NSR function is disabled by default.

1.20 Enabling the iSPF Feature

1.20.1 Overview

OSPF adopts the iSPF algorithm to compute a network topology. The iSPF function is generally used on a large-sized network to ease the pressure on router processors.

1.20.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- This configuration is recommended if you hope to accelerate route convergence in a single area with more than 100 routers.
- This configuration is performed on all routers in the area.
- After iSPF is enabled, OSPF will use the iSPF algorithm to compute the network topology. That is, after the network topology changes, OSPF corrects only the nodes affected by the topological change, instead of re-building the entire SPT.

1.20.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure the **i**SPF feature.

ispf enable

The iSPF feature is disabled by default.

1.21 Configuring to Disable MTU Verification

1.21.1 Overview

On receiving the database description packet, OSPF checks whether the MTU of the interface on the neighbor is the same as the MTU of its own interface. If the interface MTU specified in the received database description packet is greater than the MTU of the local interface, the adjacency cannot be set up. To resolve this problem, you can disable MTU verification.

1.21.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- You are advised to retain the default configuration.
- This configuration is performed on two routers with different interface MTUs.

1.21.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*

- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure to disable MTU verification.

ip ospf skip-mtu-verify

The MTU verification function is disabled by default.

1.22 Configuring to Enable Conflict Detection of Type-4 LSAs

1.22.1 Overview

This function enables conflict detection for received Type-4 LSAs.

1.22.2 Restrictions and Guidelines

The OSPF basic functions must be configured before using this function.

1.22.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) Configure conflict detection of Type-4 LSAs.

detect-conflict lsa-4

Conflict detection for Type-4 LSAs is disabled by default.

1.23 Enabling the Function of Securely Promoting a BDR to DR

1.23.1 Overview

On broadcast and NBMA networks, if a DR goes offline, all devices on the network will promote a BDR to DR to ensure the reliable operation of the network. If the configured SPF latency is too short, SPF route calculation is triggered before some devices on the network promote the BDR to DR. As a result, some routes on the network are

unreachable. The function of securely promoting a BDR to DR can solve this problem. After the function is enabled, SPF route calculation is triggered only after all devices promote a BDR to DR.

1.23.2 Restrictions and Guidelines

The OSPF basic functions must be configured before this function is used.

1.23.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

router-protocol ospf *process-id* [**vrf-instance** *vrf-name*]

(4) The following example disables the function of securely promoting a BDR to DR.

unset capability bdr-security-promoted

The function of securely promoting a BDR to DR is enabled by default.

1.24 Configuring a Super VLAN to Enable OSPF

Note

This function is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.24.1 Overview

In normal cases, a super VLAN contains multiple sub VLANs. Multicast packets of a super VLAN are also sent to its sub VLANs. In this case, when OSPF multicast packets are sent over a super VLAN containing multiple sub VLANs, the OSPF multicast packets are replicated multiple times, and the device processing capability is insufficient. As a result, a large number of packets are discarded, causing protocol flapping. In certain application scenarios, OSPF must run on a super VLAN, and the packets only need to be sent over a sub VLAN of the super VLAN. In this case, you can run the ip ospf subvlan command to specify a sub VLAN ID for sending packets, so as to avoid the neighbor flapping caused by a device processing bottleneck.

1.24.2 Restrictions and Guidelines

- The OSPF basic functions must be configured.
- The designated sub VLAN must be connected to neighbors.

1.24.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI configuration mode.

interface vlan *interface-number*

(4) Configure the Super VLAN feature.

ip ospf subvlan [**all** | *vlan-id*]

The OSPF function is not configured for a super VLAN by default.

1.25 Monitoring

Run the display command to check the running status of a configured function to verify the configuration effect.

You can run the cleanup commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Run the debugging command to output debugging information.

Caution

When the **debugging** command is used to output debugging information, system resources are occupied. Therefore, disable the debugging function immediately after use.

Table 1-10 OSPF Monitoring

Command	Purpose
display ip ospf-instance [*process-id* [*area-id*]]	Displays the OSPF process configuration information.
display ip ospf-instance [*process-id*] **border-routers**	Displays the OSPF internal routing table (the routes to ABRs and ASBRs).
display ip ospf-instance [*process-id* [*area-id*]] **database-info** [[**advertising-router** *ipv4-address* | **database-summary** | **detail** | **max-age** | **self-originate**] | { **asbr-summary** | external**-type** | **network** | **nssa-external-type** | **opaque-area** | **opaque-as** | **opaque-link** | **router** | **summary** } [*link-state-id*] [**advertising-router** *ipv4-address* | **self-originate** | **brief**]]	Displays the information of the OSPF LSDB.

Command	Purpose
display ip ospf-instance [*process-id*] **interface** [*interface-type interface-number* | **brief**]	Displays the OSPF-enabled interface.
display ip ospf-instance [*process-id*] **ispf**	Displays the number of times a route is computed in the OSPF area through iSPF.
display ip ospf-instance [*process-id*] **peer** [[*interface-type interface-number*] [*neighbor-id*] [**detail**] | **statistics**]	Displays the neighbor list of an OSPF process.
display ip ospf-instance [*process-id*] **route** [**count** | *ipv4-address mask*]	Displays the OSPF routing table.
display ip ospf-instance [*process-id*] **sham-links** [**area** *area-id*]	Displays the sham link information of an OSPF process.
display ip ospf-instance [*process-id*] **spf**	Displays the number of times SPT is computed in the OSPF area.
display ip ospf-instance [*process-id* [*area-id*]] **summary-address**	Displays the summarized route of OSPF redistributed routes.
display ip ospf-instance [*process-id* [*area-id*]] **topology** [**advertising-router** *adv-router-id* [*router-id*] | **self-originate** [*router-id*]]	Displays the OSPF network topology information.
display ip ospf-instance [*process-id*] **vlinks** [*ipv4-address*]	Displays OSPF virtual links.
display current-config router-protocol ospf	Displays the OSPF configuration.
cleanup ip ospf [*process-id*] **instance**	Clears and resets an OSPF process.
debugging ip ospf	Debugs the OSPF protocol globally.
debugging ip ospf events-info [**abr** | **asbr** | **lsa** | **mom** | **nssa** | **os** | **restart** | **router** | **slink** | **vlink**] *	Debugs OSPF events.
debugging ip ospf frr [ase | inter-area | intra-area | node | time] *	Debugs the OSPF fast reroute.
debugging ip ospf ifsm [**events-info** | **status** | **timers**]	Debugs OSPF interfaces.
debugging ip ospf nfsm [**events-info** | **status** | **timers**]	Debugs OSPF neighbors.

Command	Purpose
debugging ip ospf nsm [**bfd** | **interface** | **redistribute** | **route** [**add** | **delete** | **receive**]]	Debugs the OSPF NSM.
debugging ip ospf lsa [**flooding** | **generate** | **install** | **maxage** | **refresh**] *	Debugs OSPF LSAs.
debugging ip ospf packet [**dd | deeply** | **detail** | **hello** | **interface** | **ls-ack** | **ls-request** | **ls-update** | **recv** | **send**] *	Debugs OSPF packets.
debugging ip ospf route [**ase** | **ia** | **install** | **ispf** | **spf** | **time**] *	Debugs OSPF routes.

1.26 Configuration Examples

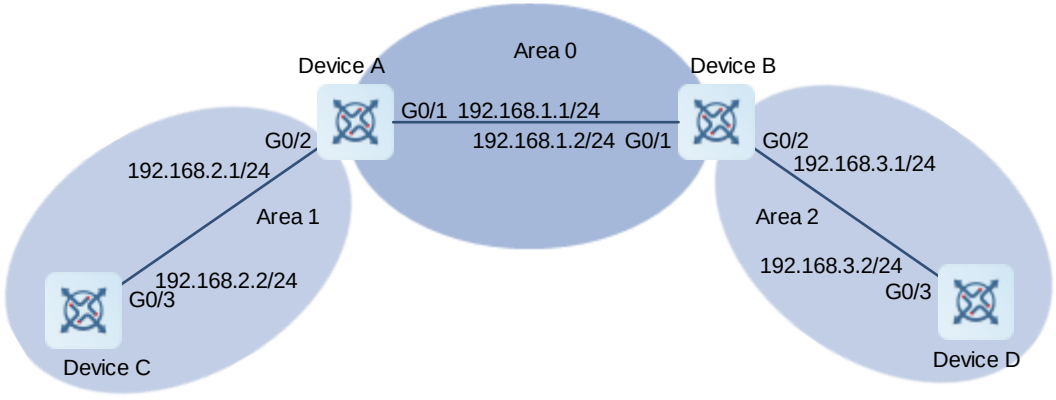
1.26.1 Configuring Basic Functions of OSPF

1. Requirements

OSPF is enabled on all the devices, and three areas are defined in total. Device A and device B are used as ABR forwarding inter-area routes so as to implement the interworking between all networks. After configuration, all devices can learn the routing of all network segments in the AS, and the neighbor relationships are correct.

2. Topology

Figure 1-14 Configuring Basic Functions of OSPF



3. Notes

- Configure interface IP addresses on all the devices.
- Enable the IPv4 unicast routing function on all the devices. (This function is enabled by default.)
- Configure the OSPF processes and router IDs on all the devices.
- Configure interface configuration OSPF on all the devices.

4. Procedure

(1) Configure interface IP addresses on all the devices.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface GigabitEthernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# ip address 192.168.2.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/2)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# ip address 192.168.3.1 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/2)# exit
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# interface gigabitethernet 0/3
Device C(cfg-if-GigabitEthernet 0/3)# ip address 192.168.2.2 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/3)# exit
```

Configure Device D.

```
Device D> privilege
Device D# config-system terminal
Device D(cfg)# interface gigabitethernet 0/3
Device D(cfg-if-GigabitEthernet 0/3)# ip address 192.168.3.2 255.255.255.0
Device D(cfg-if-GigabitEthernet 0/3)# exit
```

(2) Configure the OSPF process, router ID, and add the interface IP address to the OSPF process.

Configure Device A.

```
Device A(cfg)# router-protocol ospf 1
Device A(cfg-router)# router-id 192.168.1.1
Device A(cfg-router)# network-advertise 192.168.1.0 0.0.0.255 area 0
Device A(cfg-router)# network-advertise 192.168.2.0 0.0.0.255 area 1
```

Configure Device B.

```
Device B(cfg)# router-protocol ospf 1
Device B(cfg-router)# router-id 192.168.1.2
Device B(cfg-router)# network-advertise 192.168.1.0 0.0.0.255 area 0
Device B(cfg-router)# network-advertise 192.168.3.0 0.0.0.255 area 2
```

Configure Device C.

```
Device C(cfg)# router-protocol ospf 1
Device C(cfg-router)# router-id 192.168.2.2
Device C(cfg-router)# network-advertise 192.168.2.0 0.0.0.255 area 1
```

Configure Device D.

```
Device D(cfg)# router-protocol ospf 1
Device D(cfg-router)# router-id 192.168.3.2
Device D(cfg-router)# network-advertise 192.168.3.0 0.0.0.255 area 2
```

5. Verification

(1) Check the OSPF neighbor and route of Device A.

Check the OSPF route of Device A.

```
Device A# display ip route-config ospf
O IA 192.168.3.0/24 [110/2] via 192.168.1.2, 00:18:03, GigabitEthernet 0/1
```

Check the neighbor information of Device A.

```
Device A# display ip ospf-instance peer
OSPF instance 1, 2 Neighbors, 2 is Full:
Neighbor ID     Pri     State          Dead Time     Address          Interface
192.168.1.2     1       Full/DR        00:00:40      192.168.1.2      GigabitEthernet 0/1
192.168.2.2     1       Full/BDR       00:00:34      192.168.2.2      GigabitEthernet 0/2
```

(2) Check the OSPF neighbor and route of Device B.

Check the OSPF route of Device B.

```
Device B# display ip route-config ospf
O IA 192.168.2.0/24 [110/2] via 192.168.1.2, 00:18:03, GigabitEthernet 0/1
```

Check the neighbor information of Device B.

```
Device B# display ip ospf-instance peer
OSPF instance 1, 2 Neighbors, 2 is Full:
```

```
Neighbor ID     Pri   State           Dead Time   Address         Interface
192.168.1.1     1       Full/BDR      00:00:32    192.168.1.1     GigabitEthernet 0/1
192.168.3.2     1       Full/BDR      00:00:30    192.168.3.2     GigabitEthernet 0/2
```

(3) Check the OSPF neighbor and route of Device C.

Check the OSPF route of Device C.

```
Device C# display ip route-config ospf
O IA 192.168.1.0/24 [110/2] via 192.168.2.1, 00:19:05, GigabitEthernet 0/3
O IA 192.168.3.0/24 [110/3] via 192.168.2.1, 00:19:05, GigabitEthernet 0/3
```

Check the neighbor information of Device C.

```
Device C# display ip ospf-instance peer
OSPF instance 1, 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           Dead Time   Address         Interface
192.168.1.1     1       Full/BDR      00:00:32    192.168.2.1     GigabitEthernet 0/3
```

(4) Check the OSPF neighbor and route of Device D.

Check the OSPF route of Device D.

```
Device D# display ip route-config ospf
O IA 192.168.1.0/24 [110/2] via 192.168.3.1, 00:19:05, GigabitEthernet 0/3
O IA 192.168.2.0/24 [110/3] via 192.168.3.1, 00:19:05, GigabitEthernet 0/3
```

Check the neighbor information of Device D.

```
Device D# display ip ospf-instance peer
OSPF instance 1, 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           Dead Time   Address         Interface
192.168.1.2     1       Full/BDR      00:00:30    192.168.3.1     GigabitEthernet 0/3
```

(5) Ping 192.168.2.2 on Device D. The network is reachable.

```
Device D# ping 192.168.2.2
Sending 5, 100-byte ICMP Echoes to 192.168.2.2, timeout is 2 seconds:
< press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
 ip address 192.168.1.1 255.255.255.0
!
interface gigabitethernet 0/2
 ip address 192.168.2.1 255.255.255.0
!
```

```
router-protocol ospf 1
 router-id 192.168.1.1
 network-advertise 192.168.1.0 0.0.0.255 area 0
 network-advertise 192.168.2.0 0.0.0.255 area 1
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
 ip address 192.168.1.2 255.255.255.0
!
interface gigabitethernet 0/2
ip address 192.168.3.1 255.255.255.0
!
router-protocol ospf 1
 router-id 192.168.1.2
 network-advertise 192.168.1.0 0.0.0.255 area 0
network-advertise 192.168.3.0 0.0.0.255 area 2
!
```

- Device C configuration file

```
!
interface gigabitethernet 0/3
 ip address 192.168.2.2 255.255.255.0
!
router-protocol ospf 1
 router-id 192.168.2.2
 network-advertise 192.168.2.0 0.0.0.255 area 1
!
```

- Device D configuration file

```
!
interface gigabitethernet 0/3
 ip address 192.168.3.2 255.255.255.0
!
router-protocol ospf 1
 router-id 192.168.3.2
 network-advertise 192.168.3.0 0.0.0.255 area 2
!
```

7. Common Errors

- OSPF cannot be enabled because the IP unicast routing function is disabled.
- The network segment configured by the **network-advertise** command does not include the interface IP

addresses.

- The area IDs configured for adjacent interfaces are inconsistent.
- The same router ID is configured on multiple devices, resulting in a router ID conflict.
- The same interface IP address is configured on multiple devices, resulting in a running error of the OSPF network.

1.26.2 Configuring a Stub Area

1. Requirements

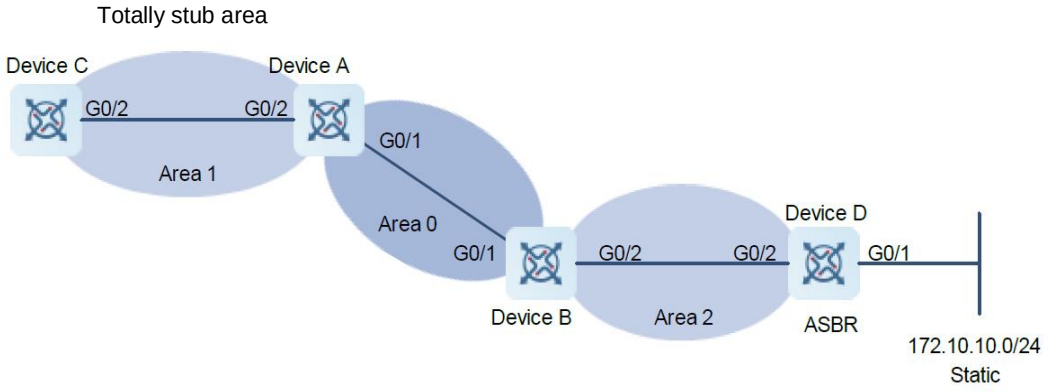
Device A, Device B, Device C, and Device D are interconnected through the OSPF routing protocol.

As ABRs, Device A and Device B are responsible for the routing information transfer between OSPF areas, and Device D functions as an ASBR to introduce the external static route.

To reduce the number of LSAs in Area 1 and save the device performance, Area 1 is configured as a totally stub area.

2. Topology

Figure 1-15 Configuring a Stub Area



3. Notes

- Configure interface IP addresses for all devices (omitted).
- Configure OSPF basic functions on all devices (omitted).
- Configure Area 1 as a stub area on Device A and Device C.

4. Procedure

(1) Introduce an external static route to Device D.

```
Device D> privilege
Device D# config-system terminal
```

```
Device D(cfg)# router-protocol ospf 1
Device D(cfg-router)# redistribute-route static subnets
```

(2) Configure Area 1 as a stub area on Device A and Device C.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# router-protocol ospf 1
Device A(cfg-router)# area 1 stub unset-summary
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# router-protocol ospf 1
Device C(cfg-router)# area 1 stub
```

5. Verification

On Device C, run the display ip route-config ospf command to display the routing table. Verify that there is only one default inter-area route, and no external static route is introduced from Device D.

```
Device C# display ip route-config ospf
O*IA 0.0.0.0/0 [110/2] via 192.168.2.1, 00:30:53, GigabitEthernet 0/2
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol ospf 1
 area 1 stub unset-summary
!
```

- Device C configuration file

```
!
router-protocol ospf 1
 area 1 stub
!
```

- Device D configuration file

```
!
router-protocol ospf 1
 redistribute-route static subnets
!
```

1.26.3 Configuring an NSSA

1. Requirements

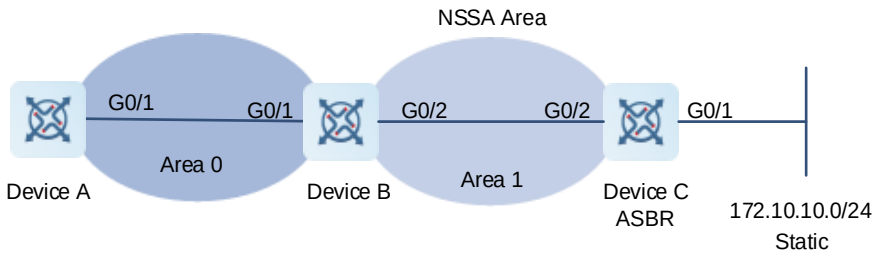
Device A, Device B, and Device C are interconnected through the OSPF routing protocol.

Device B, as an ABR, is responsible for the routing information transfer between OSPF areas, and Device C functions as an ASBR to import the external static route.

To reduce the number of LSAs in Area 1 and save the device performance, Area 1 is configured as an NSSA.

2. Topology

Figure 1-16 Configuring an NSSA



3. Notes

- Configure interface IP addresses for all devices (omitted).
- Configure OSPF basic functions on all devices (omitted).
- Device C needs to import an external static route to the OSPF process.
- Configure Area 1 as NSSA on Device B and Device C.

4. Procedure

(1) Configure a static route for Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# ip route-config 172.10.10.0 255.255.255.0 192.168.6.2
```

(2) Introduce an external static route to Device C.

```
Device C(cfg)# router-protocol ospf 1
Device C(cfg-router)# redistribute-route static subnets
```

(3) Configure Area 1 as NSSA on Device B and Device C.

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# router-protocol ospf 1
Device B(cfg-router)# area 1 nssa
```

Configure Device C.

```
Device C(cfg-router)# area 2 nssa
```

5. Verification

On Device C, check the Type-7 LSAs generated from 172.10.10.0/24.

```
Device D# display ip ospf-instance database-info nssa-external-type
                OSPF Router with ID (192.168.6.2) (Process ID 1)
                    NSSA-external Link States (Area 0.0.0.1 [NSSA])
  LS age: 61
  Options: 0x8 (-|-|-|-|N/P|-|-|-)
  LS Type: AS-NSSA-LSA
  Link State ID: 172.10.10.0 (External Network Number For NSSA)
  Advertising Router: 192.168.6.2
  LS Seq Number: 80000001
  Checksum: 0xc8f8
  Length: 36
  Network Mask: /24
        Metric Type: 2 (Larger than any link state path)
        TOS: 0
        Metric: 20
        NSSA: Forward Address: 192.168.6.2
        External Route Tag: 0
```

On Device B, verify that Type-5 and Type-7 LSAs are generated from 172.10.10.0/24.

```
Device B# display ip ospf-instance database-info nssa-external-type
                OSPF Router with ID (192.168.3.1) (Process ID 1)
                    NSSA-external Link States (Area 0.0.0.1 [NSSA])
  LS age: 314
  Options: 0x8 (-|-|-|-|N/P|-|-|-)
  LS Type: AS-NSSA-LSA
  Link State ID: 172.10.10.0 (External Network Number For NSSA)
  Advertising Router: 192.168.6.2
  LS Seq Number: 80000001
  Checksum: 0xc8f8
  Length: 36
  Network Mask: /24
        Metric Type: 2 (Larger than any link state path)
```

```
        TOS: 0
        Metric: 20
        NSSA: Forward Address: 192.168.6.2
        External Route Tag: 0
```

On Device B, check the Type-N2 routes generated from 172.10.10.0/24.

```
Device B# display ip ospf-instance database-info external-routes
            OSPF Router with ID (192.168.3.1) (Process ID 1)
                AS External Link States
  LS age: 875
  Options: 0x2 (-|-|-|-|-|-|E|-)
  LS Type: AS-external-LSA
  Link State ID: 172.10.10.0 (External Network Number)
  Advertising Router: 192.168.3.1
  LS Seq Number: 80000001
  Checksum: 0xd0d3
  Length: 36
  Network Mask: /24
        Metric Type: 2 (Larger than any link state path)
        TOS: 0
        Metric: 20
        Forward Address: 192.168.6.2
        External Route Tag: 0
```

6. Configuration Files

- Device B configuration file

```
!
router-protocol ospf 1
 area 1 nssa
!
```

- Device D configuration file

```
!
ip route-config 172.10.10.0 255.255.255.0 192.168.6.2
!
router-protocol ospf 1
 redistribute-route static subnets
 area 1 nssa
!
```

7. Common Errors

- Configurations of the area type are inconsistent on devices in the same area.

- External routes cannot be introduced because route redistribution is configured on a device in the stub area.

1.26.4 Correlating OSPF Process with BFD

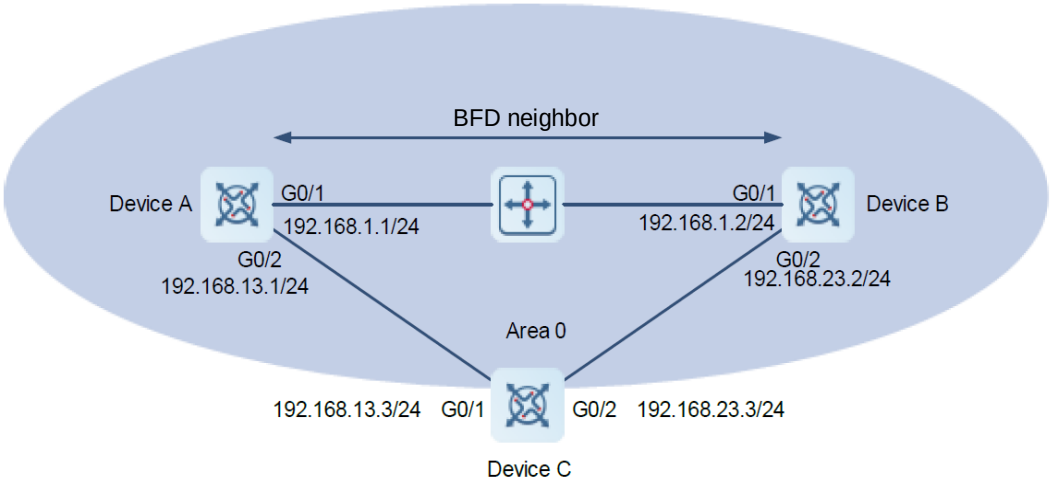
1. Requirements

Device A, Device B, and Device C implement internetworking by running OSPFv2.

Device A and Device B use a switch as the primary link. When the switch or the link connected to it fails, BFD can quickly detect the failure and switch to Device C for communication.

2. Topology

Figure 1-17 Correlating OSPF Process with BFD



3. Notes

- Configure interface IP addresses for all devices (omitted).
- Configure OSPF basic functions on all devices (omitted).
- Configure the BFD parameters for the interfaces of all devices.
- Correlate OSPF with BFD on all the devices.

4. Procedure

(1) Configure BFD parameters for the interfaces of all devices.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# bfd period 200 min_rx 200 multiplier 5
Device A(cfg-if-GigabitEthernet 0/1)# exit
```

```
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# bfd period 200 min_rx 200 multiplier 5
Device A(cfg-if-GigabitEthernet 0/2)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# bfd period 200 min_rx 200 multiplier 5
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# bfd period 200 min_rx 200 multiplier 5
Device B(cfg-if-GigabitEthernet 0/2)# exit
```

(2) Correlate OSPF with BFD on all the devices.

Configure Device A.

```
Device A(cfg)# router-protocol ospf 1
Device A(cfg-router)# bfd whole-interfaces
```

Configure Device B.

```
Device B(cfg)# router-protocol ospf 1
Device B(cfg-router)# bfd whole-interfaces
```

5. Verification

On Device A, verify that correlating OSPF with BFD is up.

```
Device A# display ip ospf-instance peer
OSPF instance 1, 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           BFD State  Dead Time   Address         Interface
192.168.1.2       1   Full/BDR        Up         00:00:40    192.168.1.2     GigabitEthernet 0/1
192.168.13.3      1   Full/BDR        Up         00:00:40    192.168.13.3    GigabitEthernet 0/2
```

On Device B, verify that correlating OSPF with BFD is up.

```
Device B# display ip ospf-instance peer
OSPF instance 1, 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           BFD State  Dead Time   Address         Interface
192.168.1.1       1   Full/BDR        Up         00:00:40    192.168.1.1     GigabitEthernet 0/1
192.168.13.3      1   Full/BDR        Up         00:00:40    192.168.23.3    GigabitEthernet 0/2
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
 bfd period 200 min_rx 200 multiplier 5
```

```
!
interface gigabitethernet 0/2
 bfd period 200 min_rx 200 multiplier 5
!
router-protocol ospf 1
 bfd whole-interfaces
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
 bfd period 200 min_rx 200 multiplier 5
!
interface gigabitethernet 0/2
 bfd period 200 min_rx 200 multiplier 5
!
router-protocol ospf 1
 bfd whole-interfaces
!
```

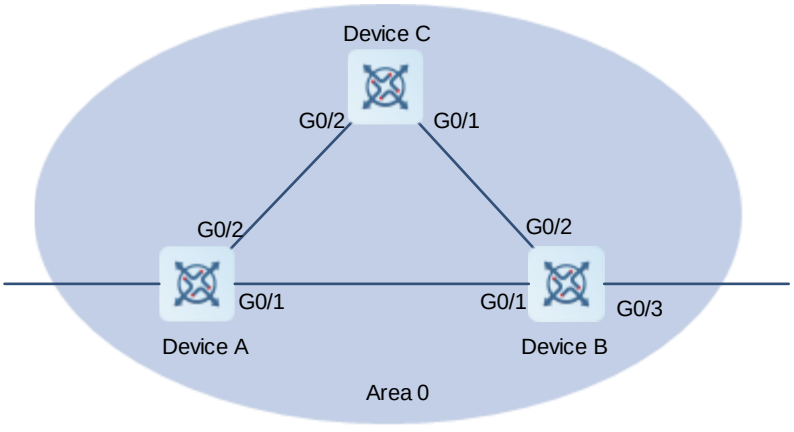
1.26.5 Configuring Fast Reroute

1. Requirements

Devices A, B, and C belong to the same OSPF area, and any two of them implement internetworking through the OSPF protocol. When a link of Device A fails, the service can be quickly switched to the backup link. The handover delay should be reduced as much as possible.

2. Topology

Figure 1-18 Configuring Fast Reroute



3. Notes

- Configure fast reroute on Device A.
- Configure **carrier-delay 0** on the interfaces on Device A.

4. Procedure

(1) Configure interface IP addresses for all devices (omitted).

(2) Configure OSPF basic functions on all devices (omitted).

(3) Configure fast reroute on Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# router-protocol ospf 1
Device A(cfg-router)# frr lfa
Device A(cfg-router)# exit
```

(4) Configure carrier-delay 0 on all the interfaces of Device A.

```
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# carrier-delaytime 0
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# carrier-delaytime 0
```

5. Verification

On Device A, check the routing table to verify that a backup route exists for the entry 192.168.4.0/24.

```
Device A# display ip route-config frr    | begin 192.168.4.0
O    192.168.4.0/24 [ma] via 192.168.1.2, 00:39:28, GigabitEthernet 0/1
                        [b] via 192.168.2.2, 00:39:28, GigabitEthernet 0/2
```

6. Configuration Files

Device A configuration file

```
!
router-protocol ospf 1
frr lfa
!
interface gigabitethernet 0/1
carrier-delaytime 0
!
interface gigabitethernet 0/2
 carrier-delaytime 0
!
```

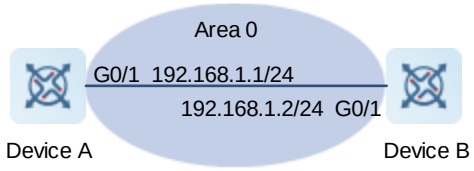
1.26.6 Configuring Network Management Functions

1. Requirements

Configure network management functions on Device A to realize information management interaction from the NMS to the agent through SNMP. The NMS server IP address is 192.168.2.2.

2. Topology

Figure 1-19 Configuring Network Management Functions



3. Notes

- Bind MIB to the OSPF process on Device A.
- Configure the trap function on Device A.

4. Procedure

(1) Configure interface IP addresses for all devices (omitted).

(2) Configure OSPF basic functions on all devices (omitted).

(3) Bind MIB to the OSPF process on Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# snmp-agent host 192.168.2.2 traps version 2c public
Device A(cfg)# snmp-agent community public rw
Device A(cfg)# snmp-agent enable traps
```

(4) Configure the trap function in OSPF process 10.

```
Device A(cfg)# router-protocol ospf 10
Device A(cfg-router)# enable mib-binding
Device A(cfg-router)# enable traps
```

5. Verification

Verify the SNMP information before the trap function is configured for the OSPF process.

```
Device A# display snmp-agent
Chassis: 60FF60
0 SNMP packets input
   0 Bad SNMP version errors
```

```
    0 Unknown community name
    0 Illegal operation for community name supplied
    0 Encoding errors
    0 Number of requested variables
    0 Number of altered variables
    0 Get-request PDUs
    0 Get-next PDUs
    0 Set-request PDUs
    0 Drop PDUs
    0 UDP parse errors
0 SNMP packets output
    0 Too big errors (Maximum packet size 1472)
    0 No such name errors
    0 Bad values errors
    0 General errors
    0 Response PDUs
    0 Trap PDUs
SNMP global trap: disabled
SNMP logging: disabled
SNMP trap logging: disabled
SNMP agent: enabled
SNMP v1:    enabled
SNMP v2c: enabled
SNMP v3:    enabled
```

Verify the SNMP information after the trap function is configured for the OSPF process.

```
Device A# display snmp-agent
Chassis: 60FF60
0 SNMP packets input
    0 Bad SNMP version errors
    0 Unknown community name
    0 Illegal operation for community name supplied
    0 Encoding errors
    8 Number of requested variables
    0 Number of altered variables
    0 Get-request PDUs
    0 Get-next PDUs
    0 Set-request PDUs
    0 Drop PDUs
    0 UDP parse errors
2 SNMP packets output
    0 Too big errors (Maximum packet size 1472)
```

```
0 No such name errors
0 Bad values errors
0 General errors
0 Response PDUs
2 Trap PDUs
```

6. Configuration Files

- Device A configuration file

```
!
snmp-agent host 192.168.2.2 traps version 2c public
snmp-agent community public rw
snmp-agent enable traps
!
router-protocol ospf 10
enable mib-binding
enable traps
!
```

7. Common Errors

- Configurations on the SNMP server are incorrect. For example, the MIB or trap function is not enabled.

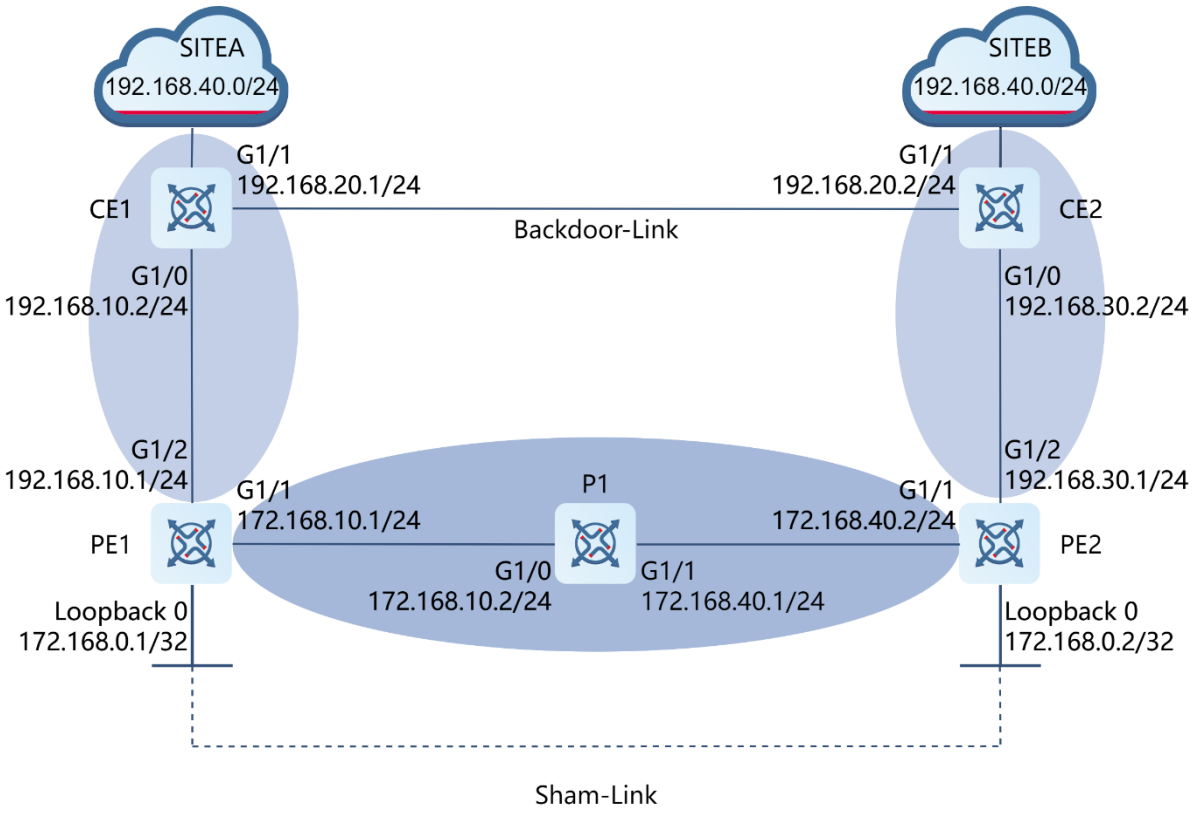
1.26.7 Configuring the Sham Link Function

1. Requirements

Two different sites of the client interact VPN routes through the MPLS backbone network. At the same time, a "backdoor link" is assumed between the two sites, and aims to ensure that, when the MPLS backbone network fails, information interaction between the two sites can be still implemented normally through this backup link.

2. Topology

Figure 1-20 Configuring a Sham Link Function



3. Notes

- Configure on CE1 to run the OSPF protocol with PE 1 and CE2, where the OSPF protocol runs with CE2 through the backdoor link. Configure the OSPF cost for the interface.
- Configure on CE2 to run the OSPF protocol with PE 2 and CE1, where the OSPF protocol runs with CE1 through the backup link. Configure the OSPF cost for the interface.
- Configure a loopback interface on PE 1, create a VRF: VPN A, define the RD value and RT value, associate VRF with the corresponding interface, and associate the interface connecting to the CE with VRF; configure a loopback interface of VRF to establish a sham link, configure BGP, establish an MP-IBGP session with PE 2, implement route interaction with CE through the OSPF protocol, establish a sham link with the OSPF process on PE 2, configure MPLS signaling of the backbone network, enable the MPLS capability of public network interface, and configure the backbone network routing protocol.
- Configure a loopback interface on PE 2, create a VRF: VPN A, define the RD value and RT value, associate VRF with the corresponding interface, and associate the interface connecting to the CE with VRF; configure a loopback interface of VRF to establish a sham link, configure BGP, establish an MP-IBGP session with PE 1, implement VPN route interaction with CE through the OSPF protocol, and configure to establish a sham link with PE 1; configure MPLS signaling of the backbone network, enable the MPLS capability of public network interface, and configure the backbone network routing protocol.
- Configure backbone network MPLS signaling on P 1, enable the MPLS capability of interface, and configure the backbone network routing protocol.

4. Procedure

(1) Make configuration on CE1.

Configure an IP address for the interface.

```
CE1> privilege
CE1# config-system terminal
CE1(cfg)# interface GigabitEthernet 1/0
CE1(cfg-GigabitEthernet 1/0)# ip address 192.168.10.2 255.255.255.0
CE1(cfg-GigabitEthernet 1/0)# ip ospf metric 1
CE1(cfg-GigabitEthernet 1/0)# exit
CE1(cfg)# interface GigabitEthernet 1/1
CE1(cfg-GigabitEthernet 1/1)# ip address 192.168.20.1 255.255.255.0
CE1(cfg-GigabitEthernet 1/1)# ip ospf metric 200
CE1(cfg-GigabitEthernet 1/1)# exit
```

Advertise the interface network to the OSPF process.

```
CE1(cfg)# router-protocol ospf 10
CE1(cfg-router)# network-advertise 192.168.10.0 255.255.255.0 area 0
CE1(cfg-router)# network-advertise 192.168.20.0 255.255.255.0 area 0
```

(2) Make configuration on CE2.

Configure an IP address for the interface.

```
CE2> privilege
CE2# config-system terminal
CE2(cfg)# interface gigabitethernet 1/0
CE2(cfg-GigabitEthernet 1/0)# ip address 192.168.30.2 255.255.255.0
CE2(cfg-GigabitEthernet 1/0)# ip ospf metric 1
CE2(cfg-GigabitEthernet 1/0)# exit
CE2(cfg)# interface gigabitethernet 1/1
CE2(cfg-GigabitEthernet 1/1)# ip address 192.168.20.2 255.255.255.0
CE2(cfg-GigabitEthernet 1/1)# ip ospf metric 200
CE2(cfg-GigabitEthernet 1/1)# exit
```

Advertise the interface network to the OSPF process.

```
CE2(cfg)# router-protocol ospf 10
CE2(cfg-router)# network-advertise 192.168.30.0 255.255.255.0 area 0
CE2(cfg-router)# network-advertise 192.168.20.0 255.255.255.0 area 0
```

(3) Make configuration on PE 1.

Configure loopback interface 0.

```
PE1> privilege
PE1# config-system terminal
PE1(cfg)# interface loopback 0
```

```
PE1(cfg-Loopback 0)# ip address 172.168.0.1 255.255.255.255
PE1(cfg-Loopback 0)# exit
```

Create a VRF VPNA and define its RD and RT attribute values.

```
PE1(cfg)# ip vrf VPNA
PE1(cfg-vrf)# rd 1:100
PE1(cfg-vrf)# route-target both 1:100
PE1(cfg-vrf)# exit
```

Configure an interface IP address and associate VRF with the corresponding interface.

```
PE1(cfg)# interface gigabitethernet 1/2
PE1(cfg-GigabitEthernet 1/2)# ip vrf binding VPNA
PE1(cfg-GigabitEthernet 1/2)# ip address 192.168.10.1 255.255.255.0
PE1(cfg-GigabitEthernet 1/2)# exit
PE1(cfg)# interface loopback 10
PE1(cfg-Loopback 10)# ip vrf binding VPNA
PE1(cfg-Loopback 10)# ip address 192.168.0.1 255.255.255.255
PE1(cfg-Loopback 10)# exit
```

Configure BGP and establish an MP-IBGP session with PE 2.

```
PE1(cfg)# router-protocol bgp 1
PE1(cfg-router)# peer 172.168.0.2 remote-as 1
PE1(cfg-router)# peer 172.168.0.2 update-source loopback 0
PE1(cfg-router)# address-family-mode vpnv4
PE1(cfg-router-af)# peer 172.168.0.2 activate
PE1(cfg-router-af)# exit
PE1(cfg-router)# exit
```

Configure to establish a sham link with the OSPF process on PE 2.

```
PE1(cfg)# router-protocol ospf 10 vrf VPNA
PE1(cfg-router)# network-advertise 192.168.10.0 255.255.255.0 area 0
PE1(cfg-router)# redistribute-route bgp subnets
PE1(cfg-router)# area 0 sham-link 192.168.0.1 192.168.0.2
PE1(cfg-router)# exit
```

Configure OSPF and direct route redistribution by BGP.

```
PE1(cfg)# router-protocol bgp 1
PE1(cfg-router)# address-family-mode ipv4 vrf VPNA
PE1(cfg-router-af)# redistribute-route ospf 10
PE1(cfg-router-af)# redistribute-route connected
PE1(cfg-router-af)# exit
PE1(cfg-router)# exit
```

Configure backbone network MPLS signaling, and enable the MPLS capability of public network interface.

```
PE1(cfg)# mpls ip
```

```
PE1(cfg)# mpls router-protocol ldp
PE1(cfg-mpls-router)# ldp router-id interface loopback 0 force
PE1(cfg-mpls-router)# exit
PE1(cfg)# interface gigabitethernet 1/1
PE1(cfg-GigabitEthernet 1/1)# ip address 172.168.10.1 255.255.255.0
PE1(cfg-GigabitEthernet 1/1)# label-switching
PE1(cfg-GigabitEthernet 1/1)# mpls ip
PE1(cfg-GigabitEthernet 1/1)# exit
```

Advertise the interface network to the OSPF process.

```
PE1(cfg)# router-protocol ospf 1
PE1(cfg-router)# network-advertise 172.168.10.0 0.0.0.255 area 0
PE1(cfg-router)# network-advertise 172.168.0.1 0.0.0.0 area 0
```

(4) Make configuration on PE 2

Configure loopback interface 0.

```
PE2> privilege
PE2# config-system terminal
PE2(cfg)# interface loopback 0
PE2(cfg-Loopback 0)# ip address 172.168.0.2 255.255.255.255
PE2(cfg-Loopback 0)# exit
```

Create a VRF VPNA and define its RD and RT attribute values.

```
PE2(cfg)# ip vrf VPNA
PE2(cfg-vrf)# rd 1:100
PE2(cfg-vrf)# route-target both 1:100
PE2(cfg-vrf)# exit
```

Configure an interface IP address and associate VRF with the corresponding interface.

```
PE2(cfg)# interface gigabitethernet 1/2
PE2(cfg-GigabitEthernet 1/2)# ip vrf binding VPNA
PE2(cfg-GigabitEthernet 1/2)# ip address 192.168.30.1 255.255.255.0
PE2(cfg-GigabitEthernet 1/2)# exit
PE2(cfg)# interface loopback 10
PE2(cfg-Loopback 10)# ip vrf-instance binding VPNA
PE2(cfg-Loopback 10)# ip address 192.168.0.2 255.255.255.255
PE2(cfg-Loopback 10)# exit
```

Configure BGP and establish an MP-IBGP session with PE 2.

```
PE2(cfg)# router-protocol bgp 1
PE2(cfg-router)# peer 172.168.0.1 remote-as 1
PE2(cfg-router)# peer 172.168.0.1 update-source loopback 0
PE2(cfg-router)# address-family-mode vpnv4
PE2(cfg-router-af)# peer 172.168.0.1 activate
```

```
PE2(cfg-router-af)# exit
PE2(cfg-router)# exit
```

Configure to establish a sham link with the OSPF process on PE 2.

```
PE2(cfg)# router-protocol ospf 10 vrf VPNA
PE2(cfg-router)# network-advertise 192.168.30.0 255.255.255.0 area 0
PE2(cfg-router)# redistribute-route bgp subnets
PE2(cfg-router)# area 0 sham-link 192.168.0.2 192.168.0.1
PE2(cfg-router)# exit
```

Configure OSPF and direct route redistribution by BGP.

```
PE2(cfg)# router-protocol bgp 1
PE2(cfg-router)# address-family-mode ipv4 vrf VPNA
PE2(cfg-router-af)# redistribute-route ospf 10
PE2(cfg-router-af)# redistribute-route connected
PE2(cfg-router-af)# exit
```

Configure backbone network MPLS signaling, and enable the MPLS capability of public network interface.

```
PE2(cfg)# mpls ip
PE2(cfg)# mpls router-protocol ldp
PE2(cfg-mpls-router)# ldp router-id interface loopback 0 force
PE2(cfg-mpls-router)# exit
PE2(cfg)# interface gigabitethernet 1/1
PE2(cfg-GigabitEthernet 1/1)# ip address 172.168.40.2 255.255.255.0
PE2(cfg-GigabitEthernet 1/1)# label-switching
PE2(cfg-GigabitEthernet 1/1)# mpls ip
PE2(cfg-GigabitEthernet 1/1)# exit
```

Advertise the interface network to the OSPF process.

```
PE2(cfg)# router-protocol ospf 1
PE2(cfg-router)# network-advertise 172.168.40.0 0.0.0.255 area 0
PE2(cfg-router)# network-advertise 172.168.0.2 0.0.0.0 area 0
```

(5) Make configuration on P 1.

Configure loopback interface 0 on P 1.

```
P1> privilege
P1# config-system terminal
P1(cfg)# interface loopback 0
P1(cfg-Loopback 0)# ip address 172.168.0.3 255.255.255.255
P1(cfg-Loopback 0)# exit
```

Make basic configuration of OSPF.

```
P1(cfg)# router-protocol ospf 1
P1(cfg-router)# network-advertise 172.168.40.0 0.0.0.255 area 0
P1(cfg-router)# network-advertise 172.168.10.0 0.0.0.255 area 0
```

```
P1(cfg-router)# network-advertise 172.168.0.3 0.0.0.0 area 0
P1(cfg-router)# exit
```

Configure backbone network MPLS signaling on P 1, and enable the MPLS capability of interface.

```
P1(cfg)# interface gigabitethernet 1/0
P1(cfg-GigabitEthernet 1/0)# ip address 172.168.10.2 255.255.255.0
P1(cfg-GigabitEthernet 1/0)# mpls ip
P1(cfg-GigabitEthernet 1/0)# label-switch
P1(cfg-GigabitEthernet 1/0)# exit
P1(cfg)# interface gigabitethernet 1/1
P1(cfg-GigabitEthernet 1/1)# ip address 172.168.40.1 255.255.255.0
P1(cfg-GigabitEthernet 1/1)# mpls ip
P1(cfg-GigabitEthernet 1/1)# label-switch
P1(cfg-GigabitEthernet 1/1)# exit
P1(cfg)# mpls ip
P1(cfg)# mpls router-protocol ldp
P1(cfg-mpls-router)# ldp router-id interface loopback 0 force
P1(cfg-mpls-router)# exit
```

5. Verification

(1) Verify the configuration on PE 1.

Verify that an OSPF sham link route exists on PE 1.

```
PE1# display ip ospf-instance 10 sham-links
Sham Link SLINK0 to address 192.168.0.2 is up
  Area 0.0.0.0 source address 192.168.0.1, Cost: 1
  Output interface is GigabitEthernet 1/1
  Nexthop address 172.16.40.2
  Transmit Delay is 1 sec, State Point-To-Point,
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:01
    Adjacency state Full
```

Check the OSPF neighbor of PE 1.

```
PE1# display ip ospf-instance 10 peer

OSPF instance 10, 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           BFD State   Dead Time   Address         Interface
192.168.0.2       1   Full/ -         -           00:00:34    192.168.0.2     SLINK0
```

Check the VRF route of PE 1.

```
PE1# display ip route-config vrf VPNA
Routing Table: VPNA
```

```
Codes:    C - connected, S - static, R - RIP, B - BGP
          O - OSPF, IA - OSPF inter area
          N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
          E1 - OSPF external type 1, E2 - OSPF external type 2
          i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
          ia - IS-IS inter area, * - candidate default

Gateway of last resort is no set
C         192.168.10.0/24 is directly connected, GigabitEthernet 1/2
O         192.168.20.0/24 [110/101] via 192.168.1.2, 00:56:23, GigabitEthernet 1/2
O         192.168.30.0/24 [110/2] via 172.168.0.2, 00:00:36
O         192.168.40.0/24 [110/2] via 172.168.0.2, 00:00:36
```

(2) Verify the configuration on PE 2.

Verify that an OSPF sham link route exists on PE 2.

```
PE2# display ip ospf-instance 10 sham-links
Sham Link SLINK0 to address 192.168.0.1 is up
  Area 0.0.0.0 source address 192.168.0.2, Cost: 1
  Output interface is GigabitEthernet 1/1
  Nexthop address 172.16.10.1
  Transmit Delay is 1 sec, State Point-To-Point,
  Timer intervals configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Hello due in 00:00:01
    Adjacency state Full
```

Check the OSPF neighbor of PE 2.

```
PE2# display ip ospf-instance 10 peer

OSPF process 10, 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           BFD State   Dead Time   Address         Interface
192.168.0.1       1   Full/ -         -           00:00:34    192.168.0.1     SLINK0
```

Check the VRF route of PE 2.

```
PE2# display ip route-config vrf VPNA
Routing Table: VPNA

Codes:    C - connected, S - static, R - RIP, B - BGP
          O - OSPF, IA - OSPF inter area
          N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
          E1 - OSPF external type 1, E2 - OSPF external type 2
          i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
          ia - IS-IS inter area, * - candidate default
```

```
Gateway of last resort is no set
O       192.168.10.0/24 [110/2] via 172.168.0.1, 00:00:36
O       192.168.20.0/24 [110/2] via 172.168.0.1, 00:00:36
C       192.168.30.0/24 is directly connected, GigabitEthernet 1/2
O       192.168.40.0/24 [110/101] via 192.168.30.2, 00:56:23, GigabitEthernet 1/2
```

6. Configuration Files

- CE1 configuration file

```
!
interface gigabitethernet 1/0
 ip address 192.168.10.2 255.255.255.0
 ip ospf metric 1
!
interface gigabitethernet 1/1
 ip address 192.168.20.1 255.255.255.0
 ip ospf metric 200
!
router-protocol ospf 10
 network-advertise 192.168.10.0 255.255.255.0 area 0
 network-advertise 192.168.20.0 255.255.255.0 area 0
!
```

- CE2 configuration file

```
!
interface gigabitethernet 1/0
 ip address 192.168.30.2 255.255.255.0
 ip ospf metric 1
!
interface gigabitethernet 1/1
 ip address 192.168.20.2 255.255.255.0
 ip ospf metric 200
!
router-protocol ospf 10
 network-advertise 192.168.30.0 255.255.255.0 area 0
 network-advertise 192.168.20.0 255.255.255.0 area 0
!
```

- PE 1 configuration file

```
!
ip vrf VPNA
 rd 1:100
 route-target both 1:100
```

```
!
interface gigabitethernet 1/1
 ip address 172.168.10.1 255.255.255.0
 label-switching
 mpls ip
!
interface gigabitethernet 1/2
 ip vrf binding VPNA
 ip address 192.168.10.1 255.255.255.0
interface loopback 0
 ip address 172.168.0.1 255.255.255.255
!
interface loopback 10
 ip vrf binding VPNA
 ip address 192.168.0.1 255.255.255.255
!
router-protocol bgp 1
 peer 172.168.0.2 remote-as 1
 peer 172.168.0.2 update-source loopback 0
 address-family-mode vpnv4
  peer 172.168.0.2 activate
 address-family-mode ipv4 vrf VPNA
  redistribute-route ospf 10
  redistribute-route connected
!
router-protocol ospf 1
 network-advertise 172.168.10.0 0.0.0.255 area 0
 network-advertise   172.168.0.1 0.0.0.0 area 0
!
router-protocol ospf 10 vrf VPNA
 network-advertise 192.168.10.0 255.255.255.0 area 0
 redistribute-route bgp subnets
 area 0 sham-link 192.168.0.1 192.168.0.2
!
mpls ip
mpls router-protocol ldp
 ldp router-id interface loopback 0 force
!
```

- PE 2 configuration file

```
!
ip vrf VPNA
```

```
 rd 1:100
 route-target both 1:100
!
interface gigabitethernet 1/1
 ip address 172.168.40.2 255.255.255.0
 label-switching
 mpls ip
!
interface gigabitethernet 1/2
 ip vrf binding VPNA
 ip address 192.168.30.1 255.255.255.0
!
interface loopback 0
 ip address 172.168.0.2 255.255.255.255
!
interface loopback 10
 ip vrf-instance binding VPNA
 ip address 192.168.0.2 255.255.255.255
!
router-protocol bgp 1
 peer 172.168.0.1 remote-as 1
 peer 172.168.0.1 update-source loopback 0
 address-family-mode vpnv4
  peer 172.168.0.1 activate
 address-family-mode ipv4 vrf-instance VPNA
  redistribute-route ospf 10
  redistribute-route connected
!
router-protocol ospf 1
 network-advertise 172.168.40.0 0.0.0.255 area 0
 network-advertise 172.168.0.2 0.0.0.0 area 0
!
router-protocol ospf 10 vrf VPNA
 network-advertise 192.168.30.0 255.255.255.0 area 0
 redistribute-route bgp subnets
 area 0 sham-link 192.168.0.2 192.168.0.1
!
mpls ip
mpls router-protocol ldp
 ldp router-id interface loopback 0 force
!
```

- P 1 configuration file

```
!
interface gigabitethernet 1/0
 ip address 172.168.10.2 255.255.255.0
 mpls ip
 label-switch
!
interface gigabitethernet 1/1
 ip address 172.168.40.1 255.255.255.0
 mpls ip
 label-switch
!
interface loopback 0
 ip address 172.168.0.3 255.255.255.255
!
router-protocol ospf 1
 network-advertise 172.168.40.0 0.0.0.255 area 0
 network-advertise 172.168.10.0 0.0.0.255 area 0
 network-advertise 172.168.0.3 0.0.0.0 area 0
!
mpls ip
mpls router-protocol ldp
 ldp router-id interface loopback 0 force
!
```

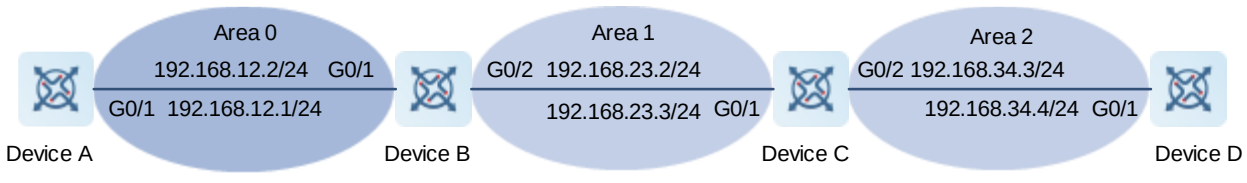
1.26.8 Configuring a Virtual Link

1. Requirements

Area 2 is not directly connected to Area 0. To ensure the normal operation of OSPF on the network, a virtual link is configured between Device B and Device C.

2. Topology

Figure 1-21 Configuring a Virtual Link



3. Notes

- Configure interface IP addresses for all devices (omitted).
- Configure OSPF basic functions on all devices (omitted).
- Configure a virtual link on Devices B and C.

4. Procedure

(1) Configure an IP address for each interface and the OSPF routing protocol (omitted).

(2) Configure a virtual link on Devices B and C.

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# router-protocol ospf 1
Device B(cfg-router)# area 1 vlink 192.168.23.3
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# router-protocol ospf 1
Device C(cfg-router)# area 1 vlink 192.168.12.2
```

5. Verification

Verify that the route of Area 2 can be learned by running the **display ip ospf-instance database-info** command on Device A.

```
Device A# display ip ospf database-info
                OSPF Router with ID (192.168.12.1) (Process ID 1)
                    Router Link States (Area 0)
Link ID         ADV Router      Age         Seq#       Checksum Link count
192.168.12.1    192.168.12.1    1025        0x80000002 0x00292E 1
192.168.12.2    192.168.12.2    3           0x80000004 0x00DA5E 2
192.168.23.3    192.168.23.3    7       (DNA) 0x80000001 0x001328 0
                    Net Link States (Area 0)
Link ID         ADV Router      Age         Seq#       Checksum
192.168.12.2    192.168.12.2    1031        0x80000001 0x00CE8E
                    Summary Net Link States (Area 0)
Link ID         ADV Router      Age         Seq#       Checksum
192.168.23.0    192.168.12.2    820         0x80000001 0x0077CC
192.168.23.0    192.168.23.3    7       (DNA) 0x80000001 0x002414
192.168.34.0    192.168.23.3    7       (DNA) 0x80000001 0x00AA82
```

6. Configuration Files

- Device B configuration file

```
!
router-protocol ospf 1
area 1 vlink 192.168.23.3
!
```

- Device C configuration file

```
!
router-protocol ospf 1
 area 1 vlink 192.168.12.2
!
```

1.26.9 Selecting a DR of OSPF

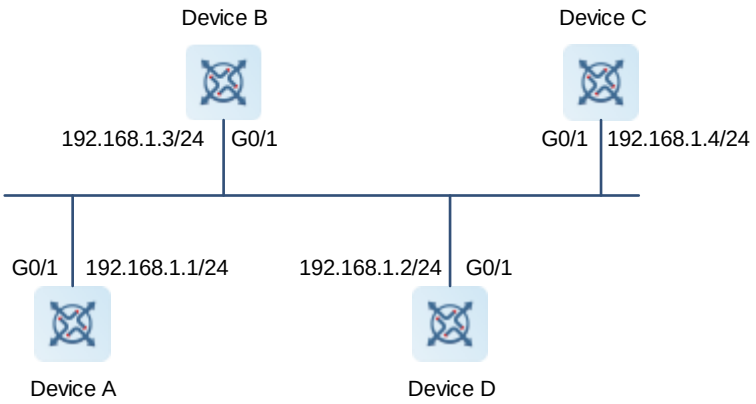
1. Requirements

Devices A, B, C, and D are on the same broadcast network.

Configure the priority of Device A as 10 to turn it into a DR, and the priority of Device B as 5 to turn it into a BDR, set the default priority for Device C, and set the priority of Device D to 0 to avoid participating in DR election.

2. Topology

Figure 1-22 Selecting a DR of OSPF



3. Notes

Set the priorities to 10, 5, and 0 on Devices A, B, and D respectively.

4. Procedure

(1) Configure an IP address for each interface and the OSPF routing protocol (omitted).

(2) Set the priorities on Devices A, B, and D.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip ospf priority-value 10
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip ospf priority-value 5
```

Configure Device D.

```
Device D> privilege
Device D# config-system terminal
Device D(cfg)# interface gigabitethernet 0/1
Device D(cfg-if-GigabitEthernet 0/1)# ip ospf priority-value 0
```

5. Verification

Verify that the displayed neighbor is correct when the display ip ospf-instance peer command is run on Device D.

```
Device D# display ip ospf-instance peer
Neighbor ID     Pri   State           Dead Time   Address         Interface
192.168.1.1     10    FULL/DR         00:00:38    192.168.1.1     GigabitEthernet0/1
192.168.1.2     5     FULL/BDR        00:00:30    192.168.1.2     GigabitEthernet0/1
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
ip ospf priority-value 10
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
ip ospf priority-value 5
!
```

- Device D configuration file

```
!
interface gigabitethernet 0/1
ip ospf priority-value 0
```

```
!
```

Contents

1 Configuring OSPFv3 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Concepts 1
1.1.3 Principles 10
1.1.4 Protocols and Standards 18
1.2 Configuration Task Summary 18
1.3 Configuring Basic OSPFv3 Functions 19
1.3.1 Overview 19
1.3.2 Restrictions and Guidelines 19
1.3.3 Configuration Tasks 19
1.3.4 Configuring OSPFv3 20
1.3.5 Creating a Virtual Link 21
1.4 Configuring OSPFv3 Network Types 22
1.4.1 Overview 22
1.4.2 Restrictions and Guidelines 22
1.4.3 Prerequisites 22
1.4.4 Procedure 22
1.5 Configuring OSPFv3 Route Advertisement 23
1.5.1 Overview 23
1.5.2 Configuration Tasks 23
1.5.3 Configuring External Route Redistribution 24

1.5.4 Generating a Default Route 24
1.5.5 Configuring a Device as ASBR 25
1.5.6 Restrictions and Guidelines 25
1.6 Configuring Stub Area and NSSA 26
1.6.1 Overview 26
1.6.2 Restrictions and Guidelines 26
1.6.3 Prerequisites 26
1.6.4 Procedure 26
1.7 Configuring OSPFv3 Route Summarization 27
1.7.1 Overview 27
1.7.2 Restrictions and Guidelines 27
1.7.3 Prerequisites 27
1.7.4 Procedure 27
1.8 Configuring OSPFv3 Route Filtering 28
1.8.1 Overview 28
1.8.2 Restrictions and Guidelines 28
1.8.3 Prerequisites 28
1.8.4 Procedure 28
1.9 Adjusting OSPFv3 Routing 29
1.9.1 Overview 29
1.9.2 Restrictions and Guidelines 29
1.9.3 Prerequisites 29
1.9.4 Configuration Tasks 29
1.9.5 Configuring Reference Bandwidth 29

1.9.6 Configuring Interface Cost 30
1.9.7 Configuring Default Metric of Redistributed Routes 32
1.9.8 Configuring Administrative Distance 32
1.10 Adjusting OSPFv3 Network Convergence Speed 33
1.10.1 Overview 33
1.10.2 Prerequisites 33
1.10.3 Configuration Tasks 33
1.10.4 Configuring Transmission Time of OSPFv3 Packets 33
1.10.5 Configuring OSPFv3 Route Computation Time 35
1.11 Enabling OSPFv3 Authentication 37
1.11.1 Overview 37
1.11.2 Restrictions and Guidelines 37
1.11.3 Prerequisites 37
1.11.4 Configuration Tasks 37
1.11.5 Configuring Area Authentication and Encryption 37
1.11.6 Configuring Interface Authentication and Encryption 38
1.12 Configuring Fast Reroute 39
1.12.1 Overview 39
1.12.2 Restrictions and Guidelines 39
1.12.3 Prerequisites 40
1.12.4 Configuration Tasks 40
1.12.5 Enabling the FRR Function 40
1.12.6 Configuring Interface LFA Protection 41
1.12.7 Configuring an Interface Not to Serve as a Backup Interface 41

1.13 Enabling Two-Way Maintenance 43
1.13.1 Overview 43
1.13.2 Prerequisites 43
1.13.3 Procedure 43
1.14 Correlating OSPFv3 with BFD 43
1.14.1 Overview 43
1.14.2 Restrictions and Guidelines 43
1.14.3 Configuration Tasks 43
1.14.4 Correlating OSPFv3 and BFD on an Interface 44
1.14.5 Correlating OSPFv3 and BFD Globally 45
1.15 Enabling the GR Function 45
1.15.1 Overview 45
1.15.2 Restrictions and Guidelines 45
1.15.3 Prerequisites 45
1.15.4 Procedure 46
1.16 Enabling NSR 46
1.16.1 Overview 46
1.16.2 Restrictions and Guidelines 46
1.16.3 Prerequisites 46
1.16.4 Procedure 46
1.17 Enabling Conflict Detection of Type 4 LSAs 47
1.17.1 Overview 47
1.17.2 Prerequisites 47
1.17.3 Procedure 47

1.18 Modifying the Maximum Number of Concurrent Neighbors 47
1.18.1 Overview 47
1.18.2 Procedure 47
1.19 Allowing Type-9 LSA to Carry IPv6 Prefixes 48
1.19.1 Overview 48
1.19.2 Procedure 48
1.20 Configuring Network Management Functions 48
1.20.1 Overview 48
1.20.2 Restrictions and Guidelines 48
1.20.3 Prerequisites 48
1.20.4 Procedure 49
1.21 Disabling MTU Verification 49
1.21.1 Overview 49
1.21.2 Prerequisites 49
1.21.3 Procedure 49
1.22 Enabling the Function of Securely Promoting a BDR to DR 50
1.22.1 Overview 50
1.22.2 Restrictions and Guidelines 50
1.22.3 Procedure 50
1.23 Enabling OSPFv3 on a Super VLAN 51
1.23.1 Overview 51
1.23.2 Restrictions and Guidelines 51
1.23.3 Prerequisites 51
1.23.4 Procedure 51

1.24 Monitoring 51

1.25 Configuration Examples 53

1.25.1 Configuring Basic Functions of OSPFv3 53

1.25.2 Configuring OSPFv3 Authentication 59

1.25.3 Configuring a Stub Area 60

1.25.4 Configuring an NSSA 62

1.25.5 Selecting an OSPFv3 DR 64

1.25.6 Configuring BFD for OSPFv3 68

1.25.7 Configuring OSPFv3 Route Summarization 72

1.25.8 Configuring FRR 74

1 Configuring OSPFv3

1.1 Introduction

1.1.1 Overview

The open shortest path first version 3 (OSPFv3) is an Interior Gateway Protocol (IGP) that is used within the autonomous system (AS) to allow routers to obtain a route to a remote network.

Note

OSPFv2 is applicable to IPv4, and OSPFv3 is applicable to IPv6. The protocol running mechanism and most configurations are the same.

OSPFv3 has the following characteristics:

- Wide scope of application: OSPFv3 is applicable to a larger-scale network that supports hundreds of routers.
- Fast convergence: Once the network topology changes, advertisements can be quickly sent between routers to update routes.
- No self-loop: Only the link status information is synchronized between routers. Each router computes routes independently, and a self-loop will not occur.
- Area division: A large routing domain is divided into multiple small areas to save system resources and network bandwidth and ensure stability and reliability of routes.
- Route classification: Routes are classified into several types to support flexible control.
- Equivalent routes: OSPFv3 supports equivalent routes.
- Authentication: OSPFv3 supports packet authentication to ensure security of protocol interaction.
- Multicast transmission: Protocol packets are sent using the multicast address to avoid interfering with irrelevant entities and save system resources.

Note

In this chapter, the term "router" refers to any network device that supports the routing function. These network devices can be layer-3 (L3) switches, routers, or firewalls.

1.1.2 Basic Concepts

1. Routing domain

All routers in an AS must be interconnected and use the same routing protocol. Therefore, an AS is also called a routing domain.

An AS in which OSPFv3 runs is also called OSPFv3 routing domain (short for OSPFv3 domain).

2. OSPFv3 process

OSPFv3 supports multiple instances, and each instance corresponds to an OSPFv3 process.

One or more OSPFv3 processes can be started on a router. Each OSPFv3 process runs OSPFv3 independently, and the processes are mutually isolated.

An OSPFv3 packet header contains the Instance ID field, and multiple OSPFv3 instances can run concurrently on a single link. The process ID is valid only on the local device.

3. Router ID

The router ID uniquely identifies a router in an OSPFv3 domain. Router IDs of any two routers cannot be the same.

If multiple OSPFv3 processes exist on a router, each OSPFv3 process uses one router ID. Router IDs of any two OSPFv3 processes cannot be the same.

4. OSPFv3 areas

OSPFv3 supports multiple areas. An OSPFv3 domain is divided into multiple areas to ease the computing pressure of a large-scale network.

An area is a logical group of routers, and each group is identified by an area ID. The border between areas is a router. A router may belong to one area or multiple areas. One network segment (link) can belong to only one area, or each OSPFv3 interface must belong to a specified area.

- Backbone area and normal area

  Area 0 is a backbone area, and other areas are normal areas. Normal areas must be directly connected to the backbone area.

- Stub area

  Configuring an OSPFv3 stub area can reduce the number of Link-State Advertisements (LSAs) in the area. The area border router (ABR) in the stub area will not transfer the Type 5 LSA (external route), and will advertise the default route to the stub area.

- Totally stub area

  A totally stub area is upgraded from a stub area. The router in the totally stub area will not transfer Type 3 LSA (inter-domain route) and Type 5 LSA (external route), and will advertise the default route to the stub area.

- NSSA

  A not-so-stubby area (NSSA) is similar to a stub area, but allows AS boundary routers (ASBRs). The route redistributed by an ASBR will be transferred in the NSSA in the form of Type 7 LSA, and the Type 7 LSA will be translated to Type 5 LSA on ABR and then transferred to other areas. The ABR also advertises the default route to the NSSA.

- Totally NSSA

  The relationship between the totally NSSA and NSSA is similar to that between the totally stub area and stub area. The totally NSSA forbids transfer of the Type 3 LSA (inter-domain route) on the basis of NSSA.

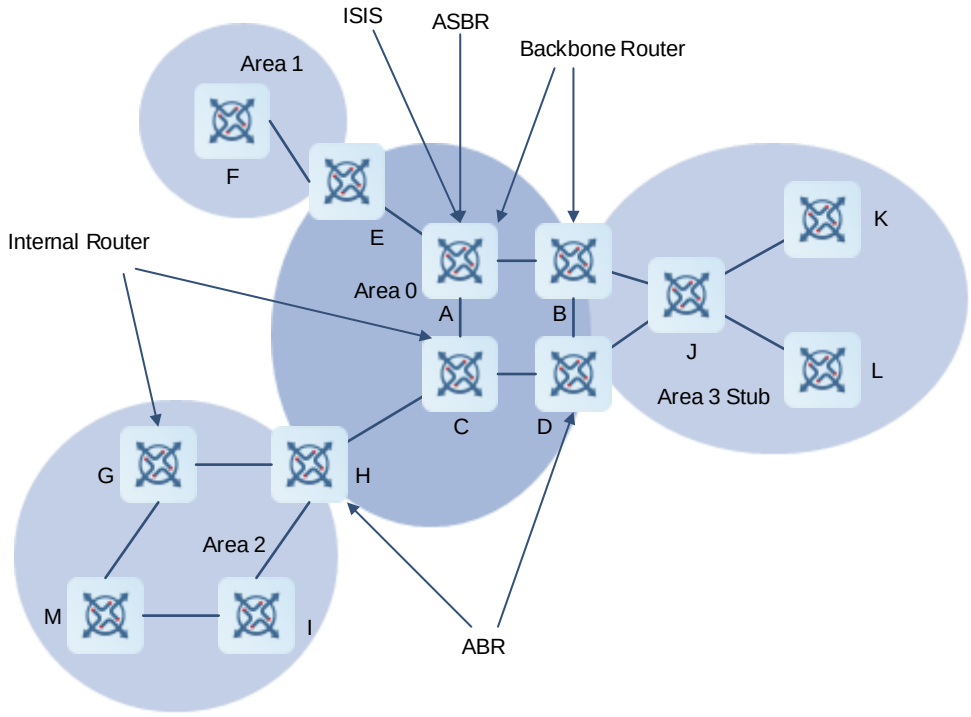
⚠ Caution

- The backbone area cannot be a stub area, totally stub area, NSSA, or totally NSSA.
- The stub area and totally stub area cannot contain any ASBR.

- The stub area, totally stub area, NSSA, and totally NSSA cannot contain any virtual link.

The stub area, totally stub area, NSSA, and totally NSSA are special forms of normal areas and help reduce the load of routers and enhance reliability of OSPFv3 routes.

Figure 1-1 Classification of OSPFv3 Areas



5. OSPFv3 routers

The following types of routers are defined in OSPFv3 and assigned with different responsibilities:

- Internal router

  All interfaces of an internal router belong to the same OSPFv3 area, for example, A, C, F, G, I, M, J, K, and L shown in Figure 1-1.

- ABR

  An ABR is used to connect the backbone area with a normal area. An ABR belongs to two or more areas, and one of the areas must be the backbone area, for example, B, D, E, and H shown in Figure 1-1.

- Backbone router

  A backbone router has at least one interface that belongs to the backbone area. All ABRs and all routers in area 0 are backbone routers, for example, A, B, C, D, E, and H shown in Figure 1-1.

- ASBR

  An ASBR is used to exchange routing information with other ASs. An ASBR is not necessarily located on the border of an AS. It may be a router inside an area or an ABR, for example, A shown in Figure 1-1.

6. OSPFv3 Route Types

Each OSPFv3 route is marked to indicate the type of the route. There are six types of OSPFv3 routes:

- O: Internal route

  This type of route describes how to arrive at a destination network in the local area. The cost of this type of route is equal to the cost of the route from the local router to the destination network.

- OI: Inter-area route

  This type of route describes how to arrive at a destination network in another area. The cost of this type of route is equal to the cost of the route from the local router to the destination network.

- OE1: Type 1 external route

  This type of route describes how to arrive at a destination network outside the AS. The cost of this type of route is equal to the cost of the route from the local router to the ASBR plus the cost of the route from the ASBR to the destination network. This type of route does not exist on routers in the stub or NSSA area.

- OE2: Type 2 external route

  This type of route describes how to arrive at a destination network outside the AS. The cost of this type of route is equal to the cost of the route from the ASBR to the destination network. This type of route does not exist on routers in the stub or NSSA area.

- ON1: Type 1 external route of the NSSA

  This type of route describes how to arrive at a destination network outside the AS through the ASBR in the NSSA. The cost of this type of route is equal to the cost of the route from the local router to the ASBR plus the cost of the route from the ASBR to the destination network. This type of route exists only on routers in the NSSA.

- ON2: Type 2 external route of the NSSA

  This type of route describes how to arrive at a destination network outside the AS through the ASBR in the NSSA. The cost of this type of route is equal to the cost of the route from the ASBR to the destination network. This type of route exists only on routers in the NSSA.

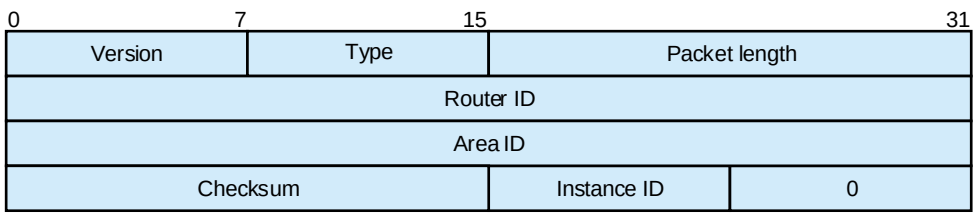
Note

Reliability of OE2 and ON2 routes is poor. OSPFv3 believes that the cost of the route from the ASBR to a destination network outside an AS is far greater than the cost of the route to the ASBR within the AS. Therefore, when the route cost is computed, only the cost of the route from the ASBR to a destination network outside an AS is considered.

7. OSPFv3 packets

The following table lists the OSPFv3 packets. These OSPFv3 packets are encapsulated in IP packets and transmitted in multicast or unicast mode.

Figure 1-2 OSPFv3 packet format



Version: indicates an OSPF version. The value is 3 for OSPFv3.

Type: indicates an OSPFv3 packet type.

Packet length: indicates the length of an OSPFv3 packet plus the packet head, in bytes.

Router ID: indicates the ID of a router that sends OSPFv3 packets.

Area ID: indicates the area of a router that sends OSPFv3 packets.

Checksum: indicates the checksum of an OSPFv3 packet, excluding the authentication field.

Instance ID: indicates an OSPFv3 instance ID.

0: reserved value.

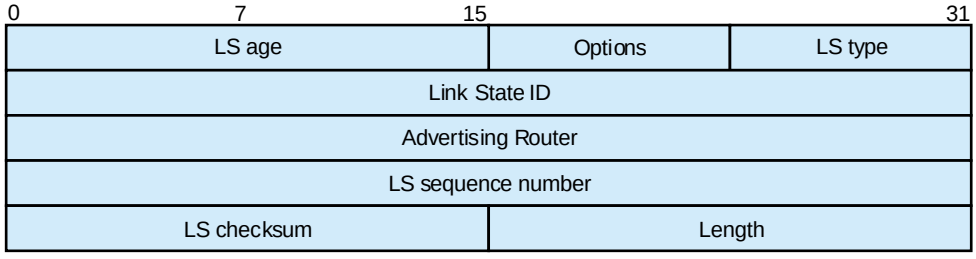
Table 1-1 Types of OSPFv3 Packets

Packet Type	Description
Hello	Hello messages are sent periodically to discover and maintain OSPFv3 neighbor relationships. The value of this OSPFv3 packet type is **1**.
Database description (DD)	DD packets carry brief information about the local link-state database (LSDB) and are used to synchronize databases between OSPFv3 neighbors. The value of this OSPFv3 packet type is **2**.
Link state request (LSR)	LSR packets are sent only after DD packets are exchanged successfully between OSPFv3 neighbors, and are used to request LSAs. The value of this OSPFv3 packet type is **3**.
Link state update (LSU)	LSU packets are used to send the required LSAs to peers. The value of this OSPFv3 packet type is **4**.
Link state acknowledgment (LSAck)	LSAck packets are used to acknowledge the received LSAs. The value of this OSPFv3 packet type is **5**.

8. LSA

OSPFv3 describes the routing information by means of LSAs.

Figure 1-3 LSA Packet Format



LS age: indicates the LS aging time of an LSA after it is generated, in seconds.

Option: is an optional field that decides whether to broadcast AS-external-LSAs, forward IP multicast packets, and process NSSA LSAs and on-demand links.

LS type: indicates different LSAs.

Link State ID: specifies a unique LSA in a routing domain together with the LSA type.

Advertising Router: indicates the ID of a router that generates this LSA.

LS sequence number: indicates the sequence number of an LSA, which is used to describe the generated sequence of LSAs.

LS checksum: indicates checksum of fields except the LS age field.

Length: indicates the total length of an LSA packet plus header, in bytes.

Table 1-2 Types of OSPFv3 LSAs

LSA Type	Description
Router-LSA (Type1)	This LSA is originated by every router. It describes the link state and cost of the router, and is advertised only within the area where the originating router is located.
Network-LSA (Type2)	This LSA is originated by a designated router (DR). It describes the state of the current link, and is advertised only within the area where the DR is located.
Inter-Area-Prefix-LSA (Type3)	This LSA is originated by an ABR. It describes a route to another area, and is advertised to areas except totally stub areas and NSSAs.
Inter-Area-Router-LSA (Type4)	This LSA is originated by an ABR. It describes a route to an ASBR, and is advertised to areas except the area where the ASBR is located.
AS-external-LSA (Type5)	This LSA is originated by an ASBR. It describes a route to a destination outside the AS, and is advertised to all areas except stub areas and NSSAs.
NSSA LSA (Type7)	This LSA is originated by an ASBR. It describes a route to a destination outside the AS, and is advertised only within the NSSA.
Link-LSA (Type8)	This LSA is originated by every router. It describes the link-local address and IPv6 prefix address of each link, and provides the link option that will be set in the Network-LSA. It is advertised only on the current link.
Intra-Area-Prefix-LSA (Type9)	Every router or DR generates one or more this type of LSAs, which are advertised in the area to which the router or DR belongs. The LSAs generated by a router describe the prefix of an IPv6 address associated with the Route-LSA. The LSAs generated by a DR describe the prefix of an IPv6 address associated with the Network-LSA.

9. OSPFv3 neighbor state

In the OSPFv3 network, neighbor devices need to reach the adjacency state before exchanging link information. There are several OSPFv3 neighbor states listed below. When a neighbor enters the full state, the adjacency relationship is established.

Table 1-3 OSPFv3 Neighbor State Machine and Description

State	Description
Down	The state of the first OSPFv3 neighbor, which indicates that the neighbor's hello packet is not received within the neighbor dead interval.
Attempt	The attempt state router sends a hello packet to the manually configured neighbor periodically. The attempt state applies to the interfaces of NBMA type only.
Init	A hello packet has been received from the neighbor, but the packet does not contain the router ID of the neighbor receiving router, that is, the peer end does not receive the hello packet sent by the local end.
Two-Way	Indicates mutual neighbors. When the neighbor router ID field in the received hello packet is each other's router ID, and both ends receive the hello packet sent from the peer end, a neighbor relationship is established. After this phase ends, DR and backup designated router (BDR) will be selected for broadcast and non-broadcast multi-access.
Exstart	Indicates negotiation about the master/slave relationship. The initial sequence number for forming an adjacency is selected, and the master/slave relationship ensures orderly transmission in the subsequent exchange of DD packets.
Exchange	Indicates exchange of DD packets. Checks whether the neighbor can provide new or updated link state information.
Loading	Indicates that, based on the information provided by DBD, the router will send LSR and interactive LSU to implement synchronization with LSDB.
Full	Indicates establishment of an adjacency. Indicates that the LSDBs of the two devices have been synchronized, and the adjacency state has been established between the local device and the neighbor device.

10. OSPFv3 network types

A router does not necessarily need to exchange LSAs with every neighbor or set up an adjacency with every neighbor. To improve efficiency, OSPFv3 classifies networks that use various link layer protocols into five types so that LSAs are exchanged in different ways to set up an adjacency.

By default, Ethernet and FDDI belong to the broadcast type, X.25, frame relay, and ATM belong to the NBMA type, and PPP, HDLC, and LAPB belong to the P2P type.

Table 1-4 Introduction to OSPFv3 Network Types

Network Type	Link Layer Protocol	Neighbor Relationship and DR Election
Broadcast	Ethernet and FDDI belong to the broadcast network type by default.	● Neighbors are automatically discovered, and the DR and BDR are elected. ● The DR or BDR exchanges LSAs with all other routers to set up an adjacency. ● Except the DR and BDR, all other routers do not exchange LSAs with each other, and the adjacency is not set up.
NBMA	X.25, frame relay, and ATM belong to NBMA networks by default.	● Neighbors are manually configured, and the DR and BDR are elected. ● The DR or BDR exchanges LSAs with all other routers to set up an adjacency. ● Except the DR and BDR, all other routers do not exchange LSAs with each other, and the adjacency is not set up.
Point-to-Point (P2P)	PPP, HDLC, and LAPB belong to the P2P network type by default.	● Neighbors are automatically discovered, and the DR or BDR is not elected. ● LSAs are exchanged between routers at both ends of the link, and the adjacency is set up.
Point-to-Multipoint (P2MP)	Networks without any link layer protocol belong to the P2MP network type by default.	● Neighbors are automatically discovered, and the DR or BDR is not elected. ● LSAs are exchanged between any two routers, and the adjacency is set up.
P2MP broadcast	Networks without any link layer protocol belong to the P2MP network type by default.	● Neighbors are manually configured, and the DR or BDR is not elected. ● LSAs are exchanged between any two routers, and the adjacency is set up.

11. DR and BDR

In the broadcast or NBMA network, many unwanted LSAs will be created when an adjacency is set up with related routers. If there are n routers, n × (n-1)/2 adjacency relationships will be set up, and n^2 LSAs will be generated in the network. In the process of setting up the adjacency and exchanging LSAs, many unwanted copies will be produced, which leads to chaos in the flooding of a multi-access network. To avoid these problems on a multi-access network, OSPFv3 defines the concepts of DR and BDR.

After the DR and BDR are elected, all routers will set up an adjacency relationship with the DR and BDR only, send information to the DR/BDR, and the DR will broadcast the network link state. The routers other than DR and BDR are called DROther routers and they will no longer set up an adjacency relationship or exchange any

routing information with each other, thus reducing the number of adjacency relationships between routers on the broadcast and NBMA networks.

Figure 1-4 Number of Adjacency Relationships Before DR Election

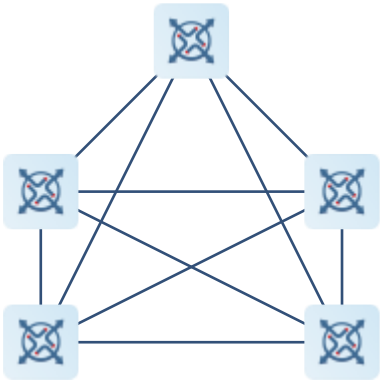
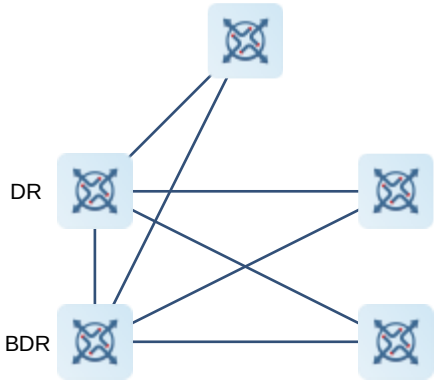


Figure 1-5 Number of Adjacency Relationships After DR Election

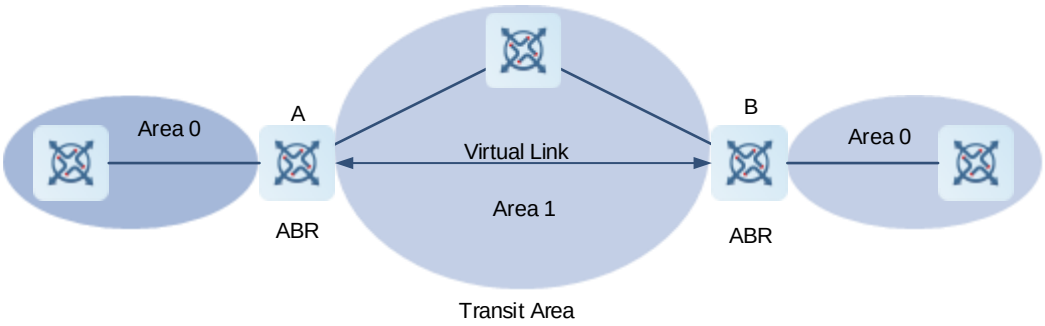


12. Virtual link

OSPFv3 supports virtual links. A virtual link is a logical link that belongs to the backbone area. It is used to resolve the problems such as a discontinuous backbone area or a failure to directly connect a normal area to the backbone area on the physical network. A virtual link supports traversal of only one normal area, and this area is called transit area. Routers on both ends of a virtual link are ABRs.

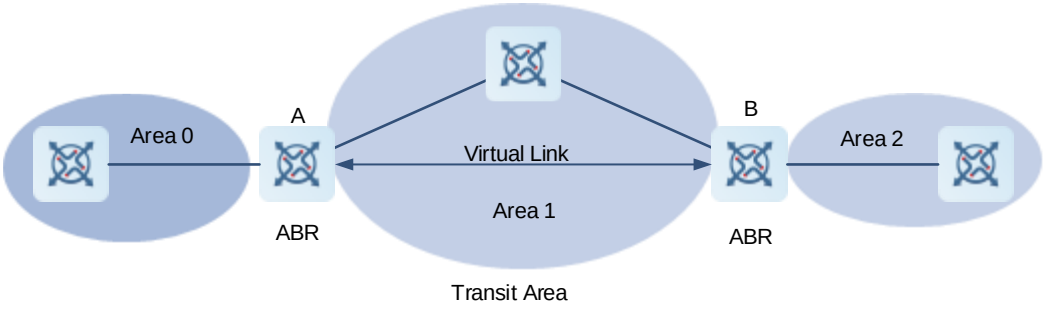
A virtual link relies on an OSPFv3 intra-domain route, making two ABRs adjacent. The OSPFv3 packet information exchanged by ABRs is transparent to the intermediate device.

Figure 1-6 Discontinuous Backbone Area on the Physical Network



As shown in Figure 1-6, a virtual link is set up between A and B to connect to Area 0. Area 1 is a transit area, and A and B are ABRs of Area 1.

Figure 1-7 Failure to Directly Connect a Normal Area to the Backbone Area on the Physical Network



As shown in Figure 1-7, a virtual link is set up between A and B to extend Area 0 to B so that Area 0 can be directly connected to Area 2 on B. Area 1 is a transit area, A is an ABR of Area 1, and B is an ABR of Area 1 and Area 2.

1.1.3 Principles

1. Simple principles

OSPFv3 is a type of link-state routing protocols. Its working process consists of three phases:

(1) Neighbor discovery > Bidirectional communication

An OSPFv3 neighbor relationship is set up between adjacent routers, and bidirectional communication is maintained.

(2) Database synchronization > Full adjacency

A router uses LSAs to advertise all its link states. LSAs are exchanged between neighbors and the LSDB is synchronized to achieve full adjacency.

(3) Shortest Path Tree (SPT) computation > Formation of a routing table

The router computes the shortest path to each destination network based on the LSDB and forms an OSPFv3 routing table.

- Neighbor discovery > Bidirectional communication

  Routers send hello packets through all OSPFv3-enabled interfaces or virtual links. If hello packets can be exchanged between two routers, and parameters carried in the hello packets can be successfully

negotiated, the two routers become neighbors. When routers that are mutual neighbors find their own router IDs from hello packets sent from neighbors, bidirectional communication is set up.

A hello packet includes, but is not limited to, the following information:

- ID of the originating router
- Area ID of the originating router interface or virtual link
- Instance ID of the originating router interface or virtual link
- Interface ID of the originating router interface or virtual link
- Priority of the originating router interface (used for DR/BDR election)
- Hello interval of the originating router interface or virtual link
- Neighbor dead interval of the originating router interface or virtual link
- IDs of the DR and BDR
- Router ID of the neighbor of the originating router

- Database synchronization > Full adjacency

  After bidirectional communication is set up between neighbor routers, the DD, LSR, LSU, and LSAck packets are used to exchange LSAs and set up the adjacency. The brief process is as follows:

  - A router generates an LSA to describe all link states on the router.
  - The LSA is exchanged between neighbors. When a router receives the LSA from its neighbor, it copies the LSA and saves the copy in the local LSDB, and then advertises the LSA to other neighbors.
  - When the router and its neighbors obtain the same LSDB, full adjacency is achieved.

Note

OSPFv3 will be very stable without change in link cost or network addition or deletion. If any change takes place, the changed link states are advertised to quickly synchronize the LSDB.

- SPT computation > Formation of a routing table

  After the complete LSDB is obtained from the router, the Dijkstra algorithm is run to generate an SPT from the local router to each destination network. The SPT records the destination networks, next-hop addresses, and cost. OSPFv3 generates a routing table based on the SPTs.

  If change in link cost or network addition or deletion takes place, the LSDB will be updated. The router again runs the Dijkstra algorithm, generates a new SPT, and updates the routing table.

Note

The Dijkstra algorithm is used to find the shortest path from a vertex to other vertices in a weighted directed graph.

2. OSPFv3 routing

- Routing cost

  If redundancy links or devices exist in the network, multiple paths may exist from the local device to the destination network. OSPFv3 selects the path with the minimum total cost to form an OSPFv3 route. The

total cost of a path is equal to the sum of the costs of individual links along the path. The total cost of a path can be minimized by modifying the costs of individual links along the path. In this way, OSPFv3 selects this path to form a route.

With configuration commands, you can modify the following link costs:

- Cost from the interface to the directly-connected network segment
- Cost from the interface to the neighbor (from local device to a specified neighbor)
- Cost from the ABR to the default network segment (cost of the default route that the ABR automatically advertises to the stub area and NSSA)
- Cost from the ASBR to an external network segment (cost of the default route that the ABR automatically advertises to the stub area and NSSA)
- Cost from the ASBR to the default network segment (cost of the default route that is manually introduced)

- OSPFv3 administrative distance

  The administrative distance (AD) is used to evaluate the reliability of various route sources. Its value is an integer ranging from 0 to 255. A smaller AD value indicates that the route is more trustworthy. If there are multiple routes to the same destination, the router preferentially selects a route with a smaller AD value. The route with a greater AD value becomes a floating route, that is, a standby route of the optimum route.

  By default, the route coming from one source corresponds to an AD value. The AD value is a local concept. Modifying the AD value affects route selection only on the current router.

Table 1-5 Routing Protocols and Default Administrative Distances

Route Source	Directly-connected network	Static route	EBGP route	OSPFv3 route	IS-IS route	RIP route	IBGP route	Unreachable route
Default AD	0	1	20	110	115	120	200	255

3. OSPFv3 route advertisement

- Route redistribution

  Route redistribution refers to the process of introducing routes of other routing protocols, routes of other OSPFv3 processes, static routes, and direct routes that exist on the device to an OSPFv3 process so that these routes can be advertised to neighbors using Type 5 and Type 7 LSAs. A default route cannot be introduced during route redistribution.

  Route redistribution is often used for interworking between ASs. You can configure route redistribution on an ASBR to advertise routes outside an AS to the interior of the AS, or routes inside an AS to the exterior of the AS.

- Default route introduction

  By configuring a command on an ASBR, you can introduce a default route to an OSPFv3 process so that the route can be advertised to neighbors using Type 5 and Type 7 LSAs.

Default route introduction is often used for interworking between ASs. One default route is used to replace all the routes outside an AS.

After configuration of route redistribution and introduction of a default route, the router automatically becomes an ASBR.

4. OSPFv3 route summarization

Route summarization is a process of summarizing routing information with the same prefix into one route, and advertising the summarized route (replacing a large number of individual routes) to neighbors. Route summarization helps reduce the protocol interaction load and the size of the routing table.

By default, the ABR advertises inter-area routing information by using Type 3 LSAs within a network segment, and advertises redistributed routing information by using Type 5 and Type 7 LSAs. If continuous network segments exist, you are advised to configure route summarization.

After route summarization is configured, the ABR sends only Type 3 LSAs that summarize prefix of IPv6 addresses, and the ASBR sends only Type 5 and Type 7 LSAs that summarize prefix of IPv6 addresses and will not send LSAs with the prefix of IPv6 addresses.

5. OSPFv3 route filtering

OSPFv3 allows filtering the learned and exchanged routes and LSAs so as to meet the security requirements in specific scenarios, for example, the routes of some network segments are not intended to be learned by other areas.

With configuration commands, you can configure route filtering for the following items:

- Interface: The interface is prevented from sending routing information (any LSA) or exchanging routing information (any LSA) with neighbors.
- Routing information outside an AS: Only the routing information that meets the filtering conditions can be redistributed to the OSPFv3 process (Type 5 and Type 7 LSAs).
- LSAs received by a router: In the OSPFv3 routing table, only the routes that are computed based on the LSAs meeting the filtering conditions can be advertised.

6. Stub area and NSSA

The stub/totally stub area and NSSA/totally NSSA help reduce the protocol interaction load and the size of the routing table.

- If an appropriate area is configured as a stub/totally stub area or NSSA/totally NSSA, advertisement of a large number of Type 5 and Type 3 LSAs can be avoided within the area.

Table 1-6 LSAs in Stub Area/NSSA

Area	LSA Types 1 and 2	LSA Type 3	LSA Type 4	LSA Type 5	LSA Type 7
Non-stub/totally stub area and non-NSSA/totally NSSA	Allowed	Allowed	Allowed	Allowed	Not allowed

Area	LSA Types 1 and 2	LSA Type 3	LSA Type 4	LSA Type 5	LSA Type 7
Stub area	Allowed	Allowed (containing one default route)	Not allowed	Not allowed	Not allowed
Totally stub area	Allowed	Only one default route is allowed.	Not allowed	Not allowed	Not allowed
NSSA	Allowed	Allowed (containing one default route)	Allowed	Not allowed	Allowed
Totally NSSA	Allowed	Only one default route is allowed.	Allowed	Not allowed	Allowed

Note

- The ABR uses Type 3 LSAs to advertise a default route to the stub/totally stub area or NSSA/totally NSSA.
- The ABR translates Type 7 LSAs in the NSSA and totally NSSA to Type 5 LSAs, and advertises Type 5 LSAs to the backbone area.

- If an appropriate area is configured as a stub/totally stub area or NSSA/totally NSSA, a large number of OE1, OE2, and OI routes will not be added to the routing table of a router in the area.

Table 1-7 Route Types in Stub Areas/NSSAs

Area	Routes Available in the Routing Table of a Router Inside the Area
Non-stub/totally stub area and non-NSSA/totally NSSA	O: a route to a destination network in the local area OI: a route to a destination network segment in another area OE1 or OE2: a route or default route to a destination network segment outside the AS (via any ASBR in the AS)
Stub area	O: a route to a destination network in the local area OI: a route or a default route to a destination network segment in another area
Totally stub area	O: a route to a destination network in the local area OI: a default route

Area	Routes Available in the Routing Table of a Router Inside the Area
NSSA	O: a route to a destination network in the local area OI: a route or a default route to a destination network segment in another area ON1 or ON2: a route or default route to a destination network segment outside the AS (via an ASBR in the local area)
Totally NSSA	O: a route to a destination network in the local area OI: a default route ON1 or ON2: a route or default route to a destination network segment outside the AS (via an ASBR in the local area)

Note

- A backbone area cannot be configured as a stub area or an NSSA.
- A transit area with virtual links going through cannot be configured as a stub area or an NSSA.
- An area containing an ASBR cannot be configured as a stub area.

7. Enhanced security and reliability

The functions such as authentication and bidirectional forwarding detection (BFD) correlation are used to enhance security, stability, and reliability of OSPFv3.

- Authentication

  OSPFv3 uses IPv6 to provide two authentication mechanisms: authentication header (AH) and encapsulating security payload (ESP), to prevent illegal routers from accessing the network and prevent hosts that forge OSPFv3 packets from participating in OSPFv3 route switching. OSPFv3 packets received on the OSPFv3 interface or at both ends of a virtual link are authenticated. If authentication fails, the packets are discarded and no adjacency can be set up. Enabling authentication can avoid learning unauthenticated or invalid routes, thus preventing advertising valid routes to unauthenticated devices. In the broadcast-type network, authentication also prevents unauthenticated devices from becoming designated devices, ensuring stability of the routing system and protecting the routing system against intrusions.

  Authentication is classified into following two types:

  - Area authentication: This area-level authentication authenticates the packets transmitted by all the interfaces in the area. This authentication type is configured in OSPFv3 routing process mode.
  - Interface authentication: This neighbor-level authentication authenticates the packets transmitted on the related interface. This authentication type is configured in interface mode.

- MTU verification

  Upon receiving a DD packet, OSPFv3 checks whether the MTU of the neighbor interface is the same as that of the local interface. If the MTU of the interface specified in the received DD packet is greater than that

of the interface that receives the packet, the adjacency cannot be set up. Disabling MTU verification can avoid this problem.

- Two-way maintenance

  OSPFv3 routers periodically send hello packets to each other to maintain the adjacency. In a large network, a lot of packets may be sent or received, occupying a great proportion of CPU and memory. As a result, some packets are delayed or discarded. If the processing time of hello packets exceeds the dead interval, the adjacency will be destroyed.

  If the two-way maintenance function is enabled, in addition to the hello packets, the DD, LSU, LSR, and LSAck packets can also be used to maintain the bidirectional communication between neighbors, which makes the adjacency more stable.

- Concurrent neighbor interaction restriction

  When a router simultaneously exchanges data with multiple neighbors, its performance may be affected. If the maximum number of neighbors that concurrently initiate or accept interaction with the OSPFv3 process is restricted, the router can interact with neighbors by batches, which ensures data forwarding and other key services.

- Allowing Type-9 LSAs to carry IPv6 prefixes

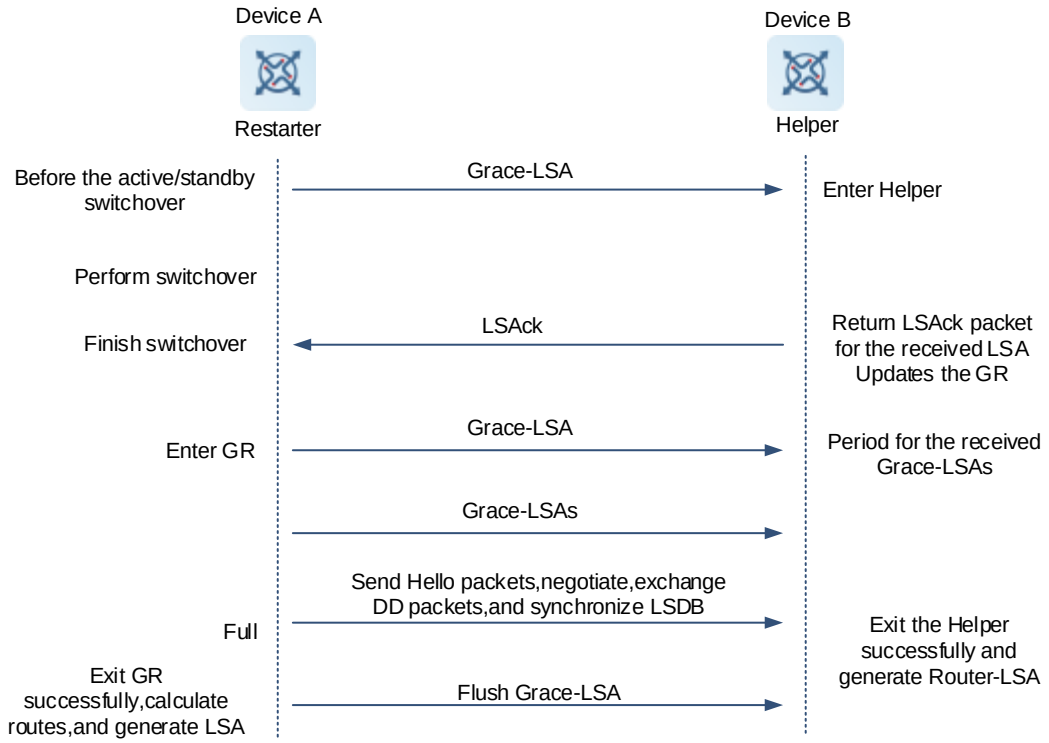
  You can run commands to configure the IPv6 address field in Type-9 LSAs to carry IPv6 addresses or IPv6 prefixes.

- GR

  The control and forwarding separated technology is widely used among routers. On a relatively stable network topology, when a graceful restart (GR) enabled router is restarted on the control plane, data forwarding can continue on the forwarding plane. In addition, actions (such as adjacency re-forming and route computation) performed on the control plane do not affect functions of the forwarding plane. In this way, service interruption caused by route flapping can be avoided, thus enhancing reliability of the entire network.

  Currently, the GR function is used only during active/standby switchover and system upgrade.

Figure 1-8 Normal OSPFv3 GR Process



- The GR process requires collaboration between the restarter and the helper. The restarter is the router where GR occurs. The helper is a neighbor of the restarter.
- When entering or exiting the GR process, the restarter sends a grace-LSA to the neighbor, notifying the neighbor to enter or exit the helper state.
- When the adjacency between the restarter and the helper reaches the full state, the router can exit the GR process successfully.

- **Fast hello**
  - OSPFv3 correlation with BFD

    After a link fails, it takes a period of time (about 40s) before OSPFv3 can sense the failure of the neighbor. Then, OSPFv3 advertises the information and re-computes the SPT. During this period, traffic is interrupted.
  - After the fast hello function is enabled (that is, the neighbor dead interval is set to 1s), OSPFv3 can sense the failure of the neighbor within 1s once a link is faulty. This accelerates route convergence and prevents traffic interruption.
  - BFD is used to test connectivity between devices. A link fault can be detected in as short as 150 ms. After OSPFv3 is correlated with BFD, OSPFv3 can sense the failure of a neighbor in as short as 150 ms once a link is faulty. This accelerates route convergence and prevents traffic interruption.

8. Network management functions

Functions such as management information base (MIB) and Syslog are used to facilitate OSPFv3 management.

- MIB

MIB is the device state information set maintained by a device. You can use the management program to view and set the MIB node.

Multiple OSPFv3 processes can be simultaneously started on a router, but the OSPFv3 MIB can be bound to only one OSPFv3 process.

- Trap

  A trap message is a notification generated when the system detects a fault. This message contains the related fault information.

  If the trap function is enabled, the router can actively send the trap messages to the network management device.

- Syslog

  The syslog records the operations (such as command configuration) performed by users on routers and specific events (such as network connection failures).

  If the syslog is allowed to record the adjacency changes, the network administrator can view the logs to learn the entire process from OSPFv3 adjacency setup to maintenance.

1.1.4 Protocols and Standards

- RFC 2740: This document describes the modifications to OSPFv3 to support version 6 of the Internet Protocol (IPv6).
- draft-ietf-ospf-ospfv3-graceful-restart: This document describes the OSPFv3 graceful restart. The OSPFv3 graceful restart is identical to OSPFv3v2 except for the differences described in this document. These differences include the format of the grace Link State Advertisements (LSA) and other considerations.
- draft-ietf-ospf-ospfv3-mib-11: This memo defines a portion of the Management Information Base (MIB) for use with network management protocols in IPv6-based Internet. In particular, it defines objects for managing the Open Shortest Path First Routing Protocol for IPv6.

1.2 Configuration Task Summary

OSPFv3 configuration includes the following tasks:

(1) Configuring Basic OSPFv3 Functions
   - Configuring OSPFv3
   - (Optional) Creating a Virtual Link

(2) (Optional) Configuring OSPFv3 Network Types

(3) (Optional) Configuring OSPFv3 Route Advertisement. The following configuration tasks are optional. Select tasks for configuration according to the actual condition.
   - Configuring External Route Redistribution
   - Generating a Default Route
   - Configuring a Device as ASBR

(4) (Optional) Configuring Stub Area and NSSA

(5) (Optional) Configuring OSPFv3 Route Summarization

(6) (Optional) Configuring OSPFv3 Route Filtering

(7) (Optional) Adjusting OSPFv3 Routing

(8) (Optional) Adjusting OSPFv3 Network Convergence Speed. The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring Transmission Time of OSPFv3 Packets
- Configuring OSPFv3 Route Computation Time

(9) (Optional) Enabling OSPFv3 Authentication. The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring Area Authentication and Encryption
- Configuring Interface Authentication and Encryption

(10) (Optional) Configuring Fast Reroute

- Enabling the FRR Function
- Configuring Interface LFA Protection
- Configuring an Interface Not to Serve as a Backup Interface

(11) (Optional) Enabling Two-Way Maintenance

(12) (Optional) Correlating OSPFv3 with BFD

(13) (Optional) Enabling the GR Function

(14) (Optional) Enabling NSR

(15) (Optional) Enabling Conflict Detection of Type 4 LSAs

(16) (Optional) Modifying the Maximum Number of Concurrent Neighbors

(17) (Optional) Allowing Type-9 LSA to Carry IPv6 Prefixes

(18) (Optional) Configuring Network Management Functions

(19) (Optional) Disabling MTU Verification

(20) (Optional) Enabling the Function of Securely Promoting a BDR to DR

(21) (Optional) Enabling OSPFv3 on a Super VLAN

1.3 Configuring Basic OSPFv3 Functions

1.3.1 Overview

Set up an OSPFv3 routing domain in the network to provide IPv6 unicast routing service for users in the network.

1.3.2 Restrictions and Guidelines

- Ensure that the IPv6 routing function is enabled, that is, the **ipv6 routing** command is not disabled; otherwise, OSPFv3 cannot be enabled.
- The interface is configured with an IPv6 address.

1.3.3 Configuration Tasks

Basic OSPFv3 function configuration includes the following tasks:

(1) Configuring OSPFv3

(2) (Optional) Creating a Virtual Link

1.3.4 Configuring OSPFv3

1. Overview

The basic OSPFv3 functions are the prerequisites of all OSPFv3 functions. An OSPFv3 process must be created, a router ID can be specified for this process, and OSPFv3 must be configured on the interface.

2. Restrictions and Guidelines

- The configuration is mandatory for every OSPFv3 router.
- You are advised to configure the router ID.
- Run the **ipv6 ospf set-area** command in the interface configuration mode to enable the OSPFv3 function on the interface, and then run the **ipv6 router-protocol ospf** command to configure an OSPFv3 process. After the OSPFv3 process is configured, the interface will automatically participate in the related process. The adjacency can be set up only between devices with the same instance ID.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an OSPFv3 process and enter OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) (Optional) Configure a router ID.

router-id *router-id*

By default, the OSPFv3 routing process elects the largest IPv4 address among all the loopback interfaces as the router ID. If the loopback interfaces configured with IP addresses are not available, the OSPFv3 process elects the largest one among the IP addresses of all its physical interfaces as the router ID.

(5) Return to the global configuration mode.

exit

(6) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregated interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*

- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the Overlay router interface configuration mode.
  - **interface overlayrouter** *interface-number*

(7) Enable OSPFv3 on the interface and specify an area ID.

ipv6 ospf *process-id* **set-area** *area-id* [**instance** *instance-id*]

OSPFv3 is disabled on an interface by default.

1.3.5 Creating a Virtual Link

1. Overview

In an OSPFv3 AS, all areas must be connected to the backbone area to properly learn the routing information of the entire OSPFv3 AS. If an area cannot be directly connected to the backbone area, the virtual link can be used to connect this area to the backbone area.

2. Restrictions and Guidelines

- The area where the virtual link is located cannot be a stub area or NSSA.
- At both ends of neighbors between which the virtual link is set up, settings of **hello-period**, **dead-period**, and **instance** must be consistent; otherwise, the adjacency cannot be set up properly.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Configure a virtual link.

area *area-id* **vlink** *router-id* [**dead-period** *dead-period* | **hello-period** *hello-period* | **instance** *instance-id* | **retransmit-period** *retransmit-period* | **transmit-lsu-delay** *transmit-delay*] * [**auth-algorithm ipsec spi** *spi* { **md5** | **sha1** } [**0** | **7**] *key* | **encipherment ipsec spi** *spi* **esp** [**3des** | **aes-cbc** { **128** | **192** | **256** } [**0** | **7**] *des-key* | **des** [**0** | **7**] *des-key* | **null**] { **md5** | **sha1** } [**0** | **7**] *key*]

No virtual link is configured by default.

1.4 Configuring OSPFv3 Network Types

1.4.1 Overview

If the physical network is X.25, frame relay, or ATM, OSPFv3 can also run to provide the IPv6 unicast routing service.

OSPFv3 classifies networks into five types according to the link layer protocol type so that LSAs are exchanged in different ways to set up an adjacency. You can configure to forcibly change the network type of an interface.

1.4.2 Restrictions and Guidelines

- The broadcast network sends OSPFv3 multicast packets, automatically discovers neighbors, and elects a DR and a BDR.
- The P2P network sends OSPFv3 multicast packets and automatically discovers neighbors.
- The NBMA network sends OSPFv3 unicast packets. Neighbors must be manually specified, and a DR and a BDR must be elected.
- The P2MP network (without carrying the **non-broadcast** parameter) sends OSPFv3 multicast packets and automatically discovers neighbors.
- The P2MP network (carrying the **non-broadcast** parameter) sends OSPFv3 unicast packets. Neighbors must be manually specified.
- On a broadcast network, a DR or BDR must be elected. During the DR/BDR election, the device with a higher priority will be preferentially elected as a DR or BDR. If the priority is the same, the device with a larger router ID will be preferentially elected as a DR or BDR. A device with the priority 0 does not participate in the DR or BDR election.

1.4.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.4.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregated interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*

- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the Overlay router interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure a network type for the interface.

ipv6 ospf network-type { **broadcast** | **non-broadcast** | **point-to-multipoint** [**non-broadcast**] | **point-to-point** } [**instance** *instance-id*]

No interface type of OSPFv3 is configured by default. Routers at both ends of a link must be configured with the same network type.

(5) (Optional) Configure neighbors.

ipv6 ospf peer *ipv6-address* [**metric** *cost* | [**poll-interval** *poll-interval* | **priority-value** *value*] *] [**instance** *instance-id*]

No neighbor is configured by default.

(6) (Optional) Configure the interface priority.

ipv6 ospf priority-value *priority* [**instance** *instance-id*]

The priority value is 1 by default.

1.5 Configuring OSPFv3 Route Advertisement

1.5.1 Overview

The function aims to introduce unicast routes for other AS domains or default routes to other AS domains to the OSPFv3 domain and provides the unicast routing service to other AS domains for users in the OSPFv3 domain. The router that introduces routes will automatically become an ASBR.

OSPFv3 has two types of external routes. The metric of the Type 1 external route changes, but the metric of the Type 2 external route is fixed. If two parallel paths to the same destination network have the same route metric, the priority of Type 1 route is higher than that of Type 2 route. Therefore, running the display ipv6 route-config command displays only the Type 1 route.

1.5.2 Configuration Tasks

The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring External Route Redistribution
- Generating a Default Route
- Configuring a Device as ASBR

1.5.3 Configuring External Route Redistribution

1. Overview

Configure external route redistribution if external routes of the OSPFv3 domain need to be introduced to an ASBR.

2. Restrictions and Guidelines

- Configure **redistribute-route** on an ASBR.
- During redistribution of IS-IS routes, **level-1**, **level-2**, or **level-1-2** parameters can be configured to indicate that IS-IS routes of the specified levels are configured. By default, level-2 IS-IS routes are redistributed.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Configure external route redistribution.

redistribute-route { **bgp** | **connected** | **isis** [*area-tag*] [**level-1** | **level-1-2** | **level-2**] * | **ospf** *process-id* [**filter** { **external-type** [**1** | **2**] | **internal-type** | **nssa-external-type** [**1** | **2**] } *] | **rip** | **static** } [**metric-value** *metric-value* | **metric-type** { **1** | **2** } | **route-map** *route-map-name* | **tag-value** *tag-value*] *

The route redistribution function is not enabled by default. The default metric type is **2**.

Level-2 routes are redistributed by default during IS-IS redistribution.

OSPFv3 routes of all sub-types are redistributed by default during OSPFv3 redistribution. The **route map** is not associated by default.

1.5.4 Generating a Default Route

1. Overview

A default route needs to be introduced to an ASBR so that other routers in the OSPFv3 domain access other AS domains through this ASBR by default.

2. Restrictions and Guidelines

- When the **redistribute-route** or **default-info** command is executed, the OSPFv3 router automatically becomes an ASBR.
- The ASBR, however, does not automatically generate or advertise a default route to all routers in the OSPFv3 routing domain. To enable an ASBR to generate a default route, run the **default-info originate** command.
- If the **constantly** parameter is specified, the OSPFv3 process advertises an external default route to neighbors no matter whether a default route exists in the core routing table. This default route, however, is not displayed on the local router. To confirm whether the default route is generated, run the **display ipv6**

ospf-instance database-info command to display the OSPFv3 link status database. On an OSPFv3 neighbor, you can run the **display ipv6 route-config ospf** command to see the default route.

- The metric of the external default route can be defined only by the **default-info originate** command, instead of the **default-metric** command.
- A router in a stub area cannot generate an external default route.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Generate a default route.

default-info originate [**constantly** | **metric-value** *metric* | **metric-class** *type* | **route-map** *map*] *

No default route is generated by default.

1.5.5 Configuring a Device as ASBR

1. Overview

Configure a device as ASBR in the AS domain or AS boundary.

1.5.6 Restrictions and Guidelines

After the redistribute-route or default-info command is executed, an OSPFv3 router automatically becomes an ASBR. If you want the device to become an ASBR without configuring the above command, run the asbr enable command.

If the asbr enable command is deleted, but the redistribute-route or default-info command configuration remains valid, the device is still an ASBR.

1. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Configure a device as ASBR.

asbr enable

No device is an ASBR by default.

1.6 Configuring Stub Area and NSSA

1.6.1 Overview

The function is used to configure an area located on the stub as a stub area to reduce interaction of routing information and the size of routing table, and enhance stability of routes. Devices in a stub area cannot learn the external routes (Type 5 LSAs) of the AS. In practice, external routes take up a large proportion of the LSDB. Therefore, devices in a stub area can learn only a small amount of routing information, which reduces the amount of system resources required to run the OSPFv3 protocol.

To reduce the size of the routing table on routers in the area and introduce OSPFv3 external routes to the area, configure the stub area as an NSSA.

1.6.2 Restrictions and Guidelines

- A backbone or transit area cannot be configured as a stub area or an NSSA.
- A router in the stub area cannot introduce external routes, but a router in the NSSA can introduce external routes.
- An area located on the stub of a network can be configured as a stub area. You must run the **area stub** command on all routers in a stub area.
- To configure a totally stub area, add the **unset-summary** keyword when you are running the **area stub** command on the ABR.
- To configure a totally NSSA, you can configure the **unset-summary** parameter on the ABR to prevent the ABR from sending the summary LSAs (Type 3 LSAs) to the NSSA.

1.6.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Configure a special OSPFv3 area. Configure one of the following tasks.

- Configure a stub area.

  area *area-id* **stub** [**unset**-**summary**]

  The stub area function is disabled by default.

- Configure an NSSA.

  area *area-id* **nssa** [**default-originate** [**metric-value** *metric* | **metric-class** *metric-type*] | **unset-redistribution** | **unset-summary** | **translator** [**constantly** | **stability**-**interval** *stability-interval*] | **zero**-**address**-**forwarding**]

1.7 Configuring OSPFv3 Route Summarization

1.7.1 Overview

Summarize routes to reduce interaction of routing information and the size of routing table, and enhance stability of routes.

Route summarization is classified into two types:

- Inter-area route summarization means to summarize routes between OSPFv3 areas.
- External route summarization means to summarize external routes of an OSPFv3 routing domain.

1.7.2 Restrictions and Guidelines

- The address range of the routes to be summarized may exceed the actual network range in the routing table. If data is sent to a network beyond the summarization range, a routing loop may be formed and the router processing load may increase. To prevent these problems, a discard route must be added to the routing table to shield or filter routes.
- Inter-area route summarization is configured and only takes effect on an ABR.
- External route summarization is configured on the ASBR that introduces routes.
- When configured on the NSSA ABR translator, **summary-prefix** summarizes redistributed routes and routes obtained based on the LSAs that are translated from Type 7 to Type 5. When configured on the ASBR (not an NSSA ABR translator), **summary-prefix** summarizes only redistributed routes.

1.7.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.7.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) (Optional) Configure inter-area route summarization.

area *area-id* **summary** *ipv6-prefix*/*prefix-length* [**advertise** | **not-advertise**]

Inter-area route summarization is not performed by default.

(5) (Optional) Configure external route summarization.

summary-prefix *ipv6-prefix*/*prefix-length* [[**metric** *cost* | **tag** *tag-value*] * | **not-advertise**]

Route summarization is disabled by default.

1.8 Configuring OSPFv3 Route Filtering

1.8.1 Overview

Route filtering can prevent routes from being loaded to the routing table or advertised to neighbors. Network users cannot access the specified destination network.

Routes are filtered using the following three methods:

- When an interface is configured as a passive interface, it no longer sends or receives hello packets.
- External routes are introduced to the ASBR.
- To prevent users from accessing the specified destination network, run the **route-filtering in** command to filter routes that are computed based on the received LSAs. Only routes meeting the filtering conditions can be forwarded. The command does not affect the LSDB or the routing tables of neighbors.

1.8.2 Restrictions and Guidelines

- The **disable-interface** command takes effect only on OSPFv3 interfaces, not on virtual links.
- The **route-filtering outbound** command is used with the **redistribute-route** command generally. The ACL filtering rules and prefix list filtering rules are mutually exclusive in the configuration. If the ACL is used for filtering routes coming from a source, the prefix list cannot be configured to filter the same routes.
- Filtering routes by using the **route-filtering in** command affects forwarding of local routes only, but does not affect route computation based on LSAs. Therefore, if route filtering is configured on the ABR, Type 3 LSAs will still be generated and advertised to other areas because routes can still be computed based on LSAs. As a result, black hole routes are generated. In this case, you can run the **area filter-rule-list** or **area status** (containing the **not-advertise** parameter) command on the ABR to prevent generation of black-hole routes.

1.8.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.8.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) (Optional) Configure a passive interface.

disable-interface { **default** | *interface-type interface-number* }

The passive mode of interfaces is disabled by default, and all interfaces are allowed to send and receive OSPFv3 packets.

(5) (Optional) Configure redistributed route filtering.

route-filtering { *acl-name* | **prefix-list** *prefix-list-name* | **route-map** } **out** [**bgp** | **connected** | **isis** [*area-tag*] | **ospf** *process-id* | **rip** | **static**]

By default, the filtering function of redistributed routes is disabled, that is, all the redistributed routes pass the filtering rules.

(6) Configure learned route filtering.

route-filtering { *acl-name* | **prefix-list** *prefix-list-name* | **route-map** *route-map-name* } **in** [*interface-type interface-number*]

By default, the function of filtering routes computed based on the received LSAs is disabled, that is, all these routes get passed.

1.9 Adjusting OSPFv3 Routing

1.9.1 Overview

In actual network scenario, the network administrator expects to control OSPFv3 routes so that traffic passes or bypasses some nodes or to adjust priority of a type of OSPFv3 routes.

1.9.2 Restrictions and Guidelines

There are two methods of adjusting OSPFv3 routes:

- Adjust cost values.
  - Configure the reference bandwidth.
  - Configure the interface cost.
  - Configure cost of the default route in a stub area or an NSSA.
  - Configure redistributed route cost.
- Modify the administrative distance.

1.9.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.9.4 Configuration Tasks

The configuration tasks for adjusting OSPF routing are as follows. The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring Reference Bandwidth
- Configuring Interface Cost
- Configure the cost value of the default route in a stub area or an NSSA.
- Configuring Default Metric of Redistributed Routes
- Configuring Administrative Distance

1.9.5 Configuring Reference Bandwidth

1. Overview

The default value of OSPFv3 reference bandwidth is 100 Mbps, and the cost is the reference bandwidth/actual bandwidth. In gigabit and 10-gigabit network environments, the calculated cost is 1. The default bandwidth cannot meet the requirements of today's high-speed transmission networks. The router connects routes of

different bandwidths. You are advised to adjust the configuration according to the actual bandwidth if you want to give priority to lines with large bandwidth.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Configure the reference bandwidth.

auto-gen-cost reference-bandwidth *reference-bandwidth*

The default reference bandwidth value computed based on the metric of an interface is **100** Mbps.

1.9.6 Configuring Interface Cost

1. Overview

You can modify the interface cost of OSPFv3 to optimize routing. The lower the link cost, the higher the priority in routing.

2. Restrictions and Guidelines

- By default, the OSPFv3 interface cost is equal to the automatic cost reference value/interface bandwidth.
- A router connects multiple lines. You are advised to configure this function if you want to manually specify the priority route.
- You can use the **auto-gen-post** command to configure the automatic reference cost. The default value is 100 Mbps.
- You can use the **bandwidth** command to configure the interface bandwidth.
- Default OSPF interface costs of typical lines:
  - For the 64K serial line, the cost is 1562;
  - For the E1 line, the cost is 48;
  - For the 10 Mbps Ethernet line, the cost is 10;
  - For the 100 Mbps Ethernet line, the cost is 1.
- If you have configured the interface cost using the **ipv6 ospf cost** command, the cost calculated automatically will be overwritten by the configured cost.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the Overlay router interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Configure the interface cost.

ipv6 ospf cost *cost* [**instance** *instance-id*]

The default cost value of an OSPFv3 interface is **100** Mbps/Bandwidth. Bandwidth indicates the bandwidth of the interface and it is configured by running the **bandwidth** command in the interface configuration mode.

1.9.7 Configure the cost value of the default route in a stub area or an NSSA.

1. Overview

After an area is configured as a Stub/NSSA area, the ABR will advertise the default route to the Stub/NSSA area to ensure that the route from the Stub/NSSA area to other areas is reachable. The default cost of the default route is 1. If you want to reduce the routing priority of the default route, you can configure a higher cost for the default route.

2. Restrictions and Guidelines

The configuration can only take effect on the ABR of the Stub area or the ABR/ASBR of the NSSA area.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Configure the cost of the default route of the Stub/NSSA area.

area *area-id* **default-metric** *cost*

By default, the cost of the default route is 1.

1.9.8 Configuring Default Metric of Redistributed Routes

1. Overview

In the case of redistributing an external route, the cost of the external route can be flexibly adjusted for routing.

2. Restrictions and Guidelines

- This function must be configured when an external route is imported on the ASBR, in order to specify the cost of the external route.
- The **default-metric** command must be used with the **redistribute-route** routing configuration command to modify the initial metrics of all redistributed routes.
- The **default-metric** command does not take effect on external routes that are injected into the OSPFv3 routing domain using the **default-info originate** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Configure default metric of redistributed routes.

default-metric *metric*

By default, the metric for redistributed routes is 20.

1.9.9 Configuring Administrative Distance

1. Overview

Administrative distance can be used to select the best route when multiple routing protocols are used to reach the same destination. Administrative distance defines the reliability of a routing protocol. The lower the administrative distance, the higher the reliability, and the higher the priority of routing. The default administrative distance of OSPFv3 intra-domain routing, inter-domain routing and external routing is 110. You can set different administrative distances to control routing.

2. Restrictions and Guidelines

- This function must be configured if multiple unicast routing protocols run on the router at the same time, and you want to change the priority of OSPFv3 routing.
- You can use this command to specify different administrative distances for different OSPF routes.
- The lower the value of administrative distance, the higher the routing priority. If the value of administrative distance is 255, the routing table entry is not trusted and will not participate in packet forwarding.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Configure an administrative distance.

distance { *distance* | **ospf** { **external-routes** *distance* | **inter-area-routes** *distance* | **intra-area** *distance* } * }

The default administrative distance of all OSPFv3 routes is **110**.

1.10 Adjusting OSPFv3 Network Convergence Speed

1.10.1 Overview

The OSPFv3 network convergence speed is controlled by adjusting the transmission time and route computation time of OSPFv3 packets.

1.10.2 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.10.3 Configuration Tasks

The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring Transmission Time of OSPFv3 Packets
- Configuring OSPFv3 Route Computation Time

1.10.4 Configuring Transmission Time of OSPFv3 Packets

1. Overview

The OSPFv3 network convergence speed is controlled by adjusting the hello packet sending interval, dead interval and LSU transmission time of an OSPFv3 interface.

2. Restrictions and Guidelines

- Hello packet sending interval
  - A shorter hello interval indicates that OSPFv3 can detect topology changes more quickly, but the network traffic increases.
  - By default, the neighbor dead interval is four times the hello packet sending interval. If the hello packet sending interval is modified, the dead interval is modified automatically.
  - The neighbor dead interval cannot be shorter than the hello packet sending interval.
  - The hello packet sending intervals and the neighbor dead intervals must be the same on all routers in the same network segment.
- Dead interval

- The dead interval cannot be smaller than the hello packet sending interval.
- The dead interval must be the same on all routers in the same network segment.

- The sending delay and line transmission delay of the interface must be considered when the **transmit-lsu-delay** command is configured. For a low-speed line, the transmission delay of the interface must be set to a value greater than the default value.
- If no acknowledgment is received from the neighbor within the time defined by the **retransmit-period** command, the LSU packet is retransmitted. The retransmission delay can be set to a value greater than the default value on a serial line or virtual link to prevent unwanted retransmission.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the Overlay router interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) (Optional) Configure the hello interval.

ipv6 ospf hello-period *hello-interval* [**instance** *instance-id*]

The default hello packet sending interval of the broadcast and P2P networks is **10** seconds. The default hello packet sending interval of the P2MP and NBMA networks is **30** seconds.

(5) (Optional) Configure the dead interval.

ipv6 ospf dead-period { *dead-period* | **minimum hello-multiplier** *multiplier* } [**instance** *instance-id*]

The fast hello function is disabled by default, and the neighbor dead interval is four times the sending interval of hello packets.

(6) (Optional) Configure the LSU transmission time.

ipv6 ospf transmit-lsu-delay *transmit-delay* [**instance** *instance-id*]

The default LSU packet transmission time is **1** second.

(7) (Optional) Configure the LSU retransmission interval.

ipv6 ospf retransmit-period *retransmit-period* [**instance** *instance-id*]

The default LSU retransmission interval is **5** seconds.

1.10.5 Configuring OSPFv3 Route Computation Time

1. Overview

OSPFv3 route convergence speed is controlled by adjusting the LSA generation time, LSA group refresh time, SPF computation time, and route computation time.

2. Restrictions and Guidelines

- LSA generation time
  - If a high requirement is raised for network convergence during link change, you can set *delay-time* to a smaller value. You can also appropriately increase the values of the preceding parameters to reduce the CPU usage.
  - The value of *hold-time* cannot be smaller than the value of *delay-time*, and the value of *max-wait-time* cannot be smaller than the value of *hold-time*.
- LSA group refresh time
  - Every LSA has a time to live (TTL) age. When the LSA age reaches 1800s, a refreshment is needed to prevent LSAs from being cleared because their ages reaching the maximum. If LSA update and aging computation are performed for every LSA, the device will consume a lot of CPU resources. To use CPU resources effectively, you can refresh LSAs by group on the device. The interval of group refreshment is called group pacing interval. The group refreshment operation is to organize the LSAs generated within a group pacing interval into a group and refresh the group as a whole.
  - If the total number of LSAs does not change, a larger group pacing interval indicates that more LSAs need to be processed after timeout. To maintain the CPU stability, the number of LSAs to be processed upon each timeout cannot be too large. If the number of LSAs is large, you are advised to reduce the group pacing interval. For example, if there are 10,000 LSAs in the database, you can reduce the pacing interval; if there are 40 to 100 LSAs, you can set the pacing interval to 10-20 minutes.
- LSA group sending interval

  If the number of LSAs is large and the device load is heavy in an environment, properly configuring *transimit-time* and *transimit-count* can limit the number of LS-UPD packets flooded in the network. If the CPU usage is not high and the network bandwidth load is not heavy, reducing the value of *transimit-time* and increasing the value of *transimit-count* can accelerate the environment convergence.
- SPF computation time
  - The *spf-delay* and *spf-holdtime* can be set to smaller values to accelerate topology convergence, and *spf-max-waittime* can be set to a larger value to reduce SPF computation.

- The value of *spf-holdtime* cannot be smaller than that of *spf-delay*; otherwise, the value of *spf-holdtime* will be automatically set to the value of spf-delay.
- The value of *spf-max-waittime* cannot be smaller than that of *spf-holdtime*; otherwise, *spf-max-waittime* will be automatically set to the value of *spf-holdtime*.
- The configurations of **timers-set throttling spf** and **timers-set spf** are mutually overwritten.
- When neither **timers-set spf** nor **timers-set throttling spf** is configured, the default value of **timers-set throttling spf** prevails.

- Computation delays of inter-area routes and external routes

  If a lot of inter-area or external routes exist in the network and the network is not stable, adjust the corresponding delays and optimize route computation to reduce the load on the device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) (Optional) Configure the LSA generation time.

timers-set throttling lsa all *delay-time hold-time max-wait-time*

The default minimum delay of LSA generation is **0** ms, the minimum interval between the first update and the second update of LSA is **5000** ms, and the maximum interval between consecutive LSA updates is **5000** ms.

(5) (Optional) Configure the LSA group refresh time.

timers-set packet-pacing lsa-batch *update-time*

The default group refresh time of LSAs is **30** seconds.

(6) (Optional) Configure LSA group sending interval.

timers-set packet-pacing lsa-transmit *transmit-time transmit-count*

The default LSA group sending interval is **40** ms, and the number of LS-UPD packets in each group is **1**.

(7) (Optional) Configure the SPF computation delay.

timers-set throttling spf *spf-delay spf-holdtime spf-max-waittime*

The default delay for SPF computation is **50** ms, the minimum interval between two SPF computations is **200** ms, and the maximum interval between two SPF computations is **5000** ms.

(8) Configure the computation delays of inter-area routes and external routes.

timers-set throttling route { **ase** *ase-delay* | **inter-area** *ia-delay* }

The default delay for inter-area route and external route computation is **0** ms.

(9) (Optional) Configure duplicate LSA receiving delay.

timers-set lsa repeat-delay *arrival-time*

The default delay for receiving a duplicate LSA is **1000** ms.

1.11 Enabling OSPFv3 Authentication

1.11.1 Overview

OSPFv3 uses the authentication mechanism, that is, IP authentication header (AH) and IP encapsulating security payload (ESP), provided by IPv6 to prevent unauthorized routers from accessing the network and hosts that forge OSPFv3 packets from participating in OSPFv3 routing.

1.11.2 Restrictions and Guidelines

- If an OSPFv3 instance is associated with SA authentication, SA authentication takes effect on all the interfaces under this OSPFv3 instance.
- If SA authentication is configured for an area, the configuration takes effect on all interfaces that belong to this area.
- Authentication priority of OSPFv3 instances, areas, and interfaces are as follows: OSPFv3 instance authentication < area authentication < interface authentication.
- Area authentication and area associated SA authentication are mutually exclusive, and interface authentication and interface associated SA authentication are mutually exclusive.

1.11.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.11.4 Configuration Tasks

The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring Area Authentication and Encryption
- Configuring Interface Authentication and Encryption

1.11.5 Configuring Area Authentication and Encryption

1. Overview

When area authentication and encryption are enabled, the authentication mode and password of all devices in an area must be consistent to ensure normal setup of neighbor relationship.

2. Restrictions and Guidelines

- The device supports three authentication types:
  - No authentication (when authentication is not configured)
  - MD5
  - SHA1
- The device supports three encryption types:
  - DES
  - 3DES
  - AES-CBC

- After an OSPFv3 area is configured with encryption and authentication, the configuration takes effect on all interfaces (except virtual links) in the area, but the interface encryption and authentication configuration takes precedence over the area configuration.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) (Optional) Configure area authentication.

area *area-id* **auth-algorithm ipsec spi** *spi* { **md5** [**string-key**] | **sha1** } [**0** | **7**] *key*

The OSPFv3 area authentication function is disabled by default.

(5) (Optional) Configure area encryption.

area *area-id* **encipherment ipsec spi** *spi* **esp** { { **3des** | **aes-cbc** { **128** | **192** | **256** } | **des** } [**0** | **7**] *des-key* | **null** } { **md5** | **sha1** } [**0** | **7**] *key*

Encryption is disabled by default.

1.11.6 Configuring Interface Authentication and Encryption

1. Overview

Applied between neighbor devices, interface authentication has higher priority than area authentication. The authentication mode and password for neighbor devices must be the same.

2. Restrictions and Guidelines

- The device supports three authentication types:
  - No authentication (when authentication is not configured)
  - MD5
  - SHA1
- The device supports three encryption types:
  - DES
  - 3DES
  - AES-CBC
- The OSPFv3 encryption and authentication parameters configured on interfaces must be consistent.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the Overlay router interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) (Optional) Configure interface authentication.

ipv6 ospf auth-algorithm { **ipsec spi** *spi* { **md5** [**string-key**] | **sha1** | **sha2-256** } [**0** | **7**] *key* | **null** } [**instance** *instance-id*]

By default, no authentication mode is configured on an interface. In this case, the authentication type of the related area is used on the interface.

(5) (Optional) Configure interface encryption.

ipv6 ospf encipherment { **ipsec spi** *spi* **esp** { { **3des** | **aes-cbc** { **128** | **192** | **256** } | **des** } [**0** | **7**] *des-key* | **null** } { **md5** | **sha1** } [**0** | **7**] *key* | **null** } [**instance** *instance-id*]

Encryption is disabled by default.

1.12 Configuring Fast Reroute

1.12.1 Overview

Once OSPFv3 detects a route failure, the fast reroute (FRR) function can immediately switch the route to the suboptimal path. This configuration helps shorten the traffic interruption time.

1.12.2 Restrictions and Guidelines

- The FRR loop-free alternate (LFA) configuration is mutually exclusive with the virtual link configuration.
- The **carrier-delay** parameter must be set to **0** for an interface.

1.12.3 Prerequisites

The OSPFv3 basic functions must be configured before this function is used.

1.12.4 Configuration Tasks

The configuration for enabling FRR includes the following tasks:

- Enabling the FRR Function
- (Optional) Configuring Interface LFA Protection
- (Optional) Configuring an Interface Not to Serve as a Backup Interface

1.12.5 Enabling the FRR Function

1. Overview

FRR provides a backup route for OSPFv3. Once OSPFv3 perceives the failure of a neighbor, it immediately switches traffic to the backup route, thereby preventing traffic interruption. You can configure the FRR function to shorten the OSPFv3 network convergence time to several milliseconds.

2. Restrictions and Guidelines

- Configure the **lfa** parameter.

If the lfa parameter is configured, LFA computation is enabled. You can also run the command in the interface configuration mode to specify the path protection mode for the interface. You are advised not to enable LFA to automatically compute the backup path if any of the following cases exists on the network:

- Virtual links exist.
- Alternative ABRs exist.
- An ASBR also serves as an ABR.
- Multiple ABSRs advertise the same external route.
- If both **lfa** and **downstream-paths** are configured, downstream path computation is enabled.
- You are advised to also enable BFD.

When using the OSPFv3 FRR function, you are advised to enable BFD so that the device can quickly detect any link failure, thereby shortening the forwarding interruption time. You can configure the carrier-delay 0 command in layer 3 interface configuration mode in interface up/down scenarios to achieve the fastest switchover speed.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

ipv6 router-protocol ospf[*process-id* [**vrf** *vrf-name*]]

(4) Configure FRR.

frr { lfa [downstream-paths] }

The FRR function is disabled by default.

1.12.6 Configuring Interface LFA Protection

1. Overview

A backup route is generated for the primary route based on the LFA protection mode specified in the interface configuration mode. By default, link protection is enabled on each OSPFv3 interface. This protection ensures that a failure of the primary link does not affect data forwarding on the backup route.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the L3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the L3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the L3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the L3 aggregate subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the Overlay router interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure interface LFA protection.

ipv6 ospf frr protection { **disable** | **link**-**node** | **node** }

LFA link protection is disabled by default.

1.12.7 Configuring an Interface Not to Serve as a Backup Interface

1. Overview

To ensure that data traffic is not switched to a specified path after the optimal path fails, it is mandatory to configure an interface not to serve as a backup interface of FRR. After the optimal path fails, traffic is switched

to a suboptimal path, that is, a new optimal path is selected based on the interface costs after OSPFv3 performs convergence.

2. Restrictions and Guidelines

When an interface is unsuitable to serve as a backup interface, you need to enable the backup disabling command in the interface configuration mode. After this command is configured, this interface is excluded from serving a backup interface in the FRR calculation of OSPFv3, and a backup interface is selected from other interfaces. An interface is unsuitable to serve as a backup interface in the following cases:

- The available bandwidth of the interface is low.
- The interface and the master interface may fail at the same time.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

Enter the global configuration mode.

config-system terminal

(2) Enter the interface configuration mode.

- Enter the L3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the L3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the L3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the L3 aggregate subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the Overlay router interface configuration mode.

  interface overlayrouter *interface-number*

(3) Configure an interface not to serve as a backup interface.

ipv6 ospf frr no-eligible-backup

By default, all interfaces can serve as backup interfaces of OSPF FRR.

1.13 Enabling Two-Way Maintenance

1.13.1 Overview

If the two-way maintenance function is enabled, in addition to the hello packets, the DD, LSU, LSR, and LSAck packets from a neighbor can also be used to maintain the bidirectional communication between neighbors. This prevents termination of the adjacency caused by delayed or discarded hello packets.

1.13.2 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.13.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Enable two-way maintenance.

two-way-maintain

The two-way maintenance function of OSPFv3 is enabled by default.

1.14 Correlating OSPFv3 with BFD

1.14.1 Overview

The OSPFv3 protocol dynamically discovers a neighbor by using hello packets. After BFD is enabled, OSPFv3 establishes a BFD session with a neighbor in the full neighbor relationship. The neighbor state is detected using the BFD mechanism. When the BFD neighbor fails, OSFPv3 immediately performs network convergence. This configuration helps shorten the traffic interruption time.

1.14.2 Restrictions and Guidelines

- The BFD parameters must be configured for the interface in advance.
- If BFD is configured for both a process and an interface, the interface-based configuration takes effect preferentially.

1.14.3 Configuration Tasks

Correlating OSPFv3 with BFD includes the following configuration tasks. These configuration tasks are optional and can be configured according to the actual situation.

- Correlating OSPFv3 and BFD on an Interface
- Correlating OSPFv3 and BFD Globally

1.14.4 Correlating OSPFv3 and BFD on an Interface

1. Overview

You can configure to correlate OSPFv3 and BFD on a specified interface to quickly detect errors on the interface and to accelerate OSPFv3 convergence.

2. Restrictions and Guidelines

- The priority of BFD on an interface is higher than that of BFD in OSPFv3 routing process.

You can use the ipv6 ospf set-bfd command to select the specified interface to use BFD for link detection according to the actual environment, or use the bfd whole-interfaces command in the OSPFv3 routing configuration mode to configure all interfaces participating in OSPF operation to use BFD for link detection, and use the ipv6 ospf set-bfd disable command to disable the BFD detection function on the specified interface.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

a Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the Overlay router interface configuration mode.
  - **interface overlayrouter** *interface-number*

b Configure an OSPFv3 interface correlation with BFD.

ipv6 ospf set-bfd [**disable**] [**instance** *instance-id*]

The BFD function is disabled on an interface by default, and the BFD configuration is subject to the configuration in the OSFPv3 process configuration mode.

1.14.5 Correlating OSPFv3 and BFD Globally

1. Overview

OSPFv3 protocol dynamically discovers neighbors through Hello packet. When OSPFv3 enables the BFD detection function, BFD sessions will be established for neighbors that reach Full relationship, and detect neighbor status through BFD mechanism. Once a BFD neighbor fails, OSPFv3 will immediately conduct network convergence.

2. Restrictions and Guidelines

You can use the ipv6 ospf set-bfd [disable] [instance *instance-id*] command in the interface configuration mode to enable or disable the BFD function on the specified interface. The priority of BFD function configured in the interface configuration mode is higher than that of the BFD function configured using bfd whole-interfaces command on all interfaces in the routing configuration mode.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

a Configure an OSPFv3 interface correlation with BFD.

bfd whole-interfaces

The BFD function is disabled on all interfaces by default.

1.15 Enabling the GR Function

1.15.1 Overview

When a GR-enabled router is restarted on the control plane, data forwarding can be still guided on the forwarding plane. In addition, actions such as neighbor relationship re-forming and route computation performed on the control plane do not affect functions of the forwarding plane. In this way, service interruption caused by route flapping can be avoided, thus enhancing reliability of the entire network.

When a router implements GR, GR helper sends a Grace-LSA to notify all neighbor routers. If the GR helper function is enabled on the local router, the local router becomes the GR helper upon receiving the Grace-LSA, and helps the neighbor routers to complete GR.

1.15.2 Restrictions and Guidelines

- The neighbor router must support the GR helper.
- The GR time cannot be shorter than the neighbor relationship maintenance time of the neighbor router.

1.15.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.15.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Enable the GR function.

gr [**grace-period** *grace-period* | **inconsistent-lsa-checking**]

The GR function is enabled by default.

(5) Enable the OSPFv3 GR helper function.

gr helper { **disable** | **internal-lsa-checking** | **check-rigid** }

The GR helper capability is enabled by default. After the GR helper is enabled on the device, LSA changes are not checked.

1.16 Enabling NSR

1.16.1 Overview

During nonstop routing (NSR), OSPFv3-related information is backed up from the active supervisor module of a distributed device to the standby supervisor module, or from the active host of a virtual switching unit (VSU) to the standby host. In this way, the device can automatically recover the link state and re-generate routes without the help of the neighbor devices during the active/standby switchover. Information that should be backed up includes the neighbor relationship and link state.

1.16.2 Restrictions and Guidelines

- This command is used to enable the NSR function. For the same OSPFv3 process, either NSR or GR is enabled because they are mutually exclusive. Nevertheless, when NSR is enabled, the GR helper capability is supported.
- The switchover of devices in distributed or VSU mode takes a period of time. If OSPFv3 neighbor keepalive duration is shorter than the switchover duration, the OSPFv3 neighbor relationship with the neighbor device is removed, and services are interrupted during the switchover. Therefore, it is recommended that the OSPFv3 neighbor keepalive duration be no smaller than the default value when the NSR function is enabled. When the fast hello function is enabled, the OSPFv3 neighbor keepalive duration is less than 1s. As a result, it is recommended that the NSR function be disabled.

1.16.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.16.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Enable the NSR function.

nsr

The NSR function is disabled by default.

1.17 Enabling Conflict Detection of Type 4 LSAs

1.17.1 Overview

This function enables conflict detection for received Type 4 LSAs.

1.17.2 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.17.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Enable conflict detection of Type 4 LSAs.

detect-conflict lsa-4

Conflict detection for Type 4 LSAs is disabled by default.

1.18 Modifying the Maximum Number of Concurrent Neighbors

1.18.1 Overview

When the performance of a router is affected because the router exchanges data with multiple neighbors, you can configure this command to restrict the maximum number of neighbors with which each OSPFv3 instance can concurrently initiate or accept interaction. The maximum number of neighbors can be configured at the process or device level.

1.18.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the maximum number of concurrent neighbors of all processes.

ipv6 router-protocol ospf max-concurrent-dd *number*

The maximum number of neighbors is **10** by default.

(4) Enter the OSPFv3 routing configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(5) (Optional) Configure the maximum number of concurrent neighbors of the current process.

max-concurrent-dd *number*

The default maximum number of concurrent neighbors for a single process is **5**.

1.19 Allowing Type-9 LSA to Carry IPv6 Prefixes

1.19.1 Overview

Type-9 LSAs generated by devices can carry IPv6 addresses or IPv6 prefixes. In specific scenarios, devices of earlier versions need to carry IPv6 addresses. When interworking with the Quagga server, only IPv6 prefixes can be carried; otherwise, the logic of the Quagga server becomes abnormal. You can configure this command as required.

1.19.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Disable the function of allowing Type-9 LSAs to carry IPv6 prefixes on all routers.

unset ipv6 **router-protocol ospf compatible lsa type-9**

The function of allowing Type-9 LSAs to carry IPv6 prefixes is enabled on all routers by default.

1.20 Configuring Network Management Functions

1.20.1 Overview

Use the network management software to manage OSPFv3 parameters and monitor the OSPFv3 running status.

1.20.2 Restrictions and Guidelines

- To enable the MIB function of the OSPFv3, you must enable the MIB function of the SNMP-agent.
- To enable the Trap function of the OSPFv3, you must enable the Trap function of the SNMP-agent.
- To output the OSPFv3 logs, you must enable the logging function of the device.

1.20.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.20.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPFv3 configuration mode.

ipv6 router-protocol ospf [*process-id* [**vrf** *vrf-name*]]

(4) Bind the MIB with an OSPFv3 process.

enable mib-binding

The MIB is bound to the OSPFv3 process with the minimum process ID by default.

(5) Configure the Trap function.

enable traps [**error** [ifc**onfig**e**rror** | ifr**x**b**ad**p**acket** | v**irt**I**f**c**onfig**e**rror** | v**irt**I**f**r**x**b**ad**p**acket**] * | **state**-**change** [if**s**t**ate**c**hange** | n**br**s**tate**c**hange** | n**ssa**t**ranslator**s**tatus**c**hange** | v**irt**I**f**s**tate**c**hange** | v**irt**n**br**s**tate**c**hange** | r**estart**s**tatus**c**hange** | n**br**r**estart**h**elper**s**tatus**c**hange** | v**irt**n**br**r**estart**h**elper**s**tatus**c**hange**] *]

The Trap function is disabled by default.

(6) (Optional) Configure the logging function.

log-adj-changes [**detail**]

The logging function is enabled by default, without the **detail** parameter. The log records the log information of the following four types of events only: the adjacency reaches the full state; the adjacency leaves the full state; the adjacency reaches the down state; the adjacency leaves the down state.

1.21 Disabling MTU Verification

1.21.1 Overview

When receiving a DD packet, OSPFv3 verifies whether the MTU of the neighbor's interface is the same as that of its own interface. If the neighbor interface MTU specified in the local received DD packet is greater than the MTU of the local interface, the neighbor relationship cannot be set up. After this function is enabled, adjacency can be set up and the unicast routing service can be provided even if the MTUs of interfaces on neighbor routers are different.

1.21.2 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.21.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
- **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
- **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
- **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
- **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
- **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
- **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
- **interface loopback** *interface-number*
- Enter the Overlay router interface configuration mode.
- **interface overlayrouter** *interface-number*

(4) Disable MTU verification.

ipv6 ospf skip-mtu-verify

The MTU verification function is disabled by default.

1.22 Enabling the Function of Securely Promoting a BDR to DR

1.22.1 Overview

On broadcast and non-broadcast multiple access (NBMA) networks, if the DR goes offline, all devices on the network promotes a BDR to DR to ensure the reliable operation of the network. If the configured SPF latency is too short, SPF route calculation is triggered before some devices on the network promote a BDR to DR. As a result, some routes on the network are unreachable. If the function of securely promoting a BDR to DR is enabled, SPF route calculation is triggered only after all devices promote a BDR to DR.

1.22.2 Restrictions and Guidelines

The OSPFv3 basic functions must be configured before this function is used.

1.22.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the OSPF configuration mode.

ipv6 router-protocol ospf [*process-id* [vrf *vrf-name*]]

(4) Disable the function of securely promoting a BDR to DR.

unset capability bdr-security-promoted

The function of securely promoting a BDR to DR is enabled by default.

1.23 Enabling OSPFv3 on a Super VLAN

Note

This function is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.23.1 Overview

In normal cases, a super VLAN contains multiple sub VLANs. Multicast packets of a super VLAN are also sent to its sub VLANs. In this case, when OSPFv3 multicast packets are sent over a super VLAN containing multiple sub VLANs, the OSPFv3 multicast packets are replicated multiple times, and the device processing capability is insufficient. As a result, a large number of packets are discarded, causing protocol flapping. In most scenarios, the OSPFv3 function does not need to be enabled in a super VLAN, and it is disabled by default. In some other scenarios, OSPFv3 needs to be run in the super VLAN, but packets need to be sent to only one sub VLAN. In this case, run this command to specify a particular sub VLAN. You must be cautious when configuring packet transmission to all sub VLANs, as the large number of sub VLANs may cause a device processing bottleneck, which will lead to the neighbor flapping.

1.23.2 Restrictions and Guidelines

- The specified sub VLAN must be connected to neighbors.

1.23.3 Prerequisites

The basic OSPFv3 functions must be configured before this function is used.

1.23.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI configuration mode.

interface *vlan interface-number*

(4) Configure OSPFv3 to run on a super VLAN.

ipv6 ospf subvlan { **all** | *vlan-id* } [**instance** *instance*-id]

The OSPFv3 function takes effect in super VLANs only and is disabled by default.

1.24 Monitoring

Run the display command to check the running status of a configured function to verify the configuration effect.

Run the debugging command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the cleanup commands to clear information.

Caution

Running the **cleanup** command may lose vital information and thus interrupt services.

Table 1-8 Monitoring

Command	Purpose
display ipv6 ospf-instance [*process-id*]	Displays the configuration information of the OSPFv3 process.
display ipv6 **ospf-instance** [*process-id*] **database-info** [**database-summary |** *lsa-type* [**advertising-router** *router-id*]]	Displays the OSPFv3 LSDB.
display ipv6 ospf-instance [*process-id* [*area-id*]] **frr-topology** { **advertising-router** *adv-router-id* [*router-id*] | **self-originate** [*router-id*] }	Displays information about the topology calculated by OSPFv3 FRR SPF.
display ipv6 ospf-instance [*process-id*] **interface** [*interface-type interface-number* | **brief**]	Displays the OSPFv3 interface.
display ipv6 ospf-instance [*process-id*] **peer** [*interface-type interface-number* [**detail**] | *neighbor-id* | **statistics**]	Displays the neighbor list of OSPFv3.
display ipv6 ospf-instance [*process-id*] **route** [**count**]	Displays the OSPFv3 routing table.
display ipv6 ospf-instance [*process-id*] **spf**	Displays the number of times that SPT is computed in an OSPF area.
display ipv6 ospf-instance [*process-id*] **summary-prefix**	Displays the summarized routes of OSPFv3 redistributed routes.
display ipv6 ospf-instance [*process-id*] **topology** [**area** *area-id*]	Displays the OSPFv3 network topology information.
display ipv6 ospf-instance [*process-id*] **vlinks**	Displays the OSPFv3 virtual links.
cleanup ipv6 ospf [*process-id*] **instance**	Clears and resets the OSPFv3 process.
debugging ipv6 ospf events-info [**abr** | **asbr** | **ipsec** | **mom** | **nssa** | **os | rdnd-thread | restart | router** | **vlink**]	Debugs the OSPFv3 events.

Command	Purpose
debugging ipv6 ospf frr [ase | inter-area | intra-area | node | time]	Debugs the OSPFv3 FRR.
debugging ipv6 ospf ifsm [**events-info** | **status** | **timers**]	Debugs the OSPFv3 interfaces.
debugging ipv6 ospf nfsm [**events-info** | **status** | **timers**]	Debugs the OSPFv3 neighbors.
debugging ipv6 ospf nsm [**bfd | interface** | **redistribute** | **route** [**add | delete | receive**]]	Debugs the OSPFv3 NSM.
debugging ipv6 ospf lsa [**flooding** | **generate** | **install** | **maxage** | **refresh**]	Debugs the OSPFv3 LSAs.
debugging ipv6 ospf packet-info [**dd | deeply** | **detail** | **hello | interface** | **ls-ack** | **ls-request** | **ls-update** | **recv** | **send**]	Debugs the OSPFv3 packets.
debugging ipv6 ospf route [**ase** | **ia** | **install** | **spf** | **time**]	Debugs the OSPFv3 routes.

1.25 Configuration Examples

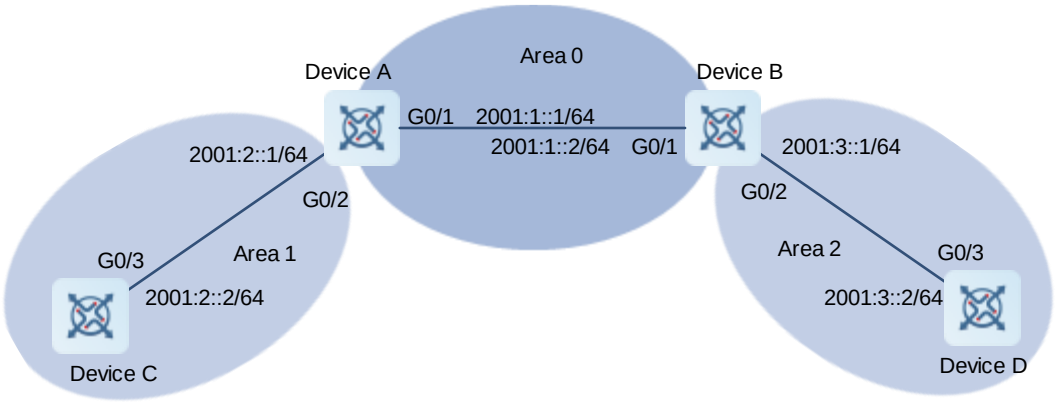
1.25.1 Configuring Basic Functions of OSPFv3

1. Requirements

OSPFv3 is enabled on all the devices, and three areas are defined in total. Device A and device B are used as ABRs forwarding inter-area routes so as to implement the interworking between all networks.

2. Topology

Figure 1-9 Configuring Basic Functions of OSPFv3



3. Notes

- Configure interface IP addresses on all the devices.
- Enable the IPv4 unicast routing function on all the devices. (This function is enabled by default.)
- Configure OSPFv3 instances and router IDs on all the devices.
- Configure OSPFv3 on interfaces of all the devices.

4. Procedure

(1) Perform the following configurations on Device A.

Enable the OSPFv3 process and configure a router ID.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ipv6 router-protocol ospf 1
Device A(cfg-router)# router-id 1.1.1.1
Device A(cfg-router)# exit
```

Configure an IP address for the interface and enable OSPFv3 on the interface.

```
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 enable
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 address 2001:1::1/64
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 1 set-area 0
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/2)# ipv6 enable
Device A(cfg-if-GigabitEthernet 0/2)# ipv6 address 2001:2::1/64
Device A(cfg-if-GigabitEthernet 0/2)# ipv6 ospf 1 set-area 1
Device A(cfg-if-GigabitEthernet 0/2)# exit
```

(2) Perform the following configurations on Device B.

Enable the OSPFv3 process and configure a router ID.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ipv6 router-protocol ospf 1
Device B(cfg-router)# router-id 2.2.2.2
Device B(cfg-router)# exit
```

Configure an IP address for the interface and enable OSPFv3 on the interface.

```
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 enable
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 address 2001:1::2/64
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 1 set-area 0
Device B(cfg-if-GigabitEthernet 0/1)# exit
```

```
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/2)# ipv6 enable
Device B(cfg-if-GigabitEthernet 0/2)# ipv6 address 2001:3::1/64
Device B(cfg-if-GigabitEthernet 0/2)# ipv6 ospf 1 set-area 2
Device B(cfg-if-GigabitEthernet 0/2)# exit
```

(3) Perform the following configurations on Device C.

Enable the OSPFv3 process and configure a router ID.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# ipv6 router-protocol ospf 1
Device C(cfg-router)# router-id 3.3.3.3
Device C(cfg-router)# exit
```

Configure an IP address for the interface and enable OSPFv3 on the interface.

```
Device C(cfg)# interface gigabitethernet 0/3
Device C(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/3)# ipv6 enable
Device C(cfg-if-GigabitEthernet 0/3)# ipv6 address 2001:2::2/64
Device C(cfg-if-GigabitEthernet 0/3)# ipv6 ospf 1 set-area 1
Device C(cfg-if-GigabitEthernet 0/3)# exit
```

(4) Perform the following configurations on Device D.

Enable the OSPFv3 process and configure a router ID.

```
Device D> privilege
Device D# config-system terminal
Device D(cfg)# ipv6 router-protocol ospf 1
Device D(cfg-router)# router-id 4.4.4.4
Device D(cfg-router)# exit
```

Configure an IP address for the interface and enable OSPFv3 on the interface.

```
Device D(cfg)# interface gigabitethernet 0/3
Device D(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device D(cfg-if-GigabitEthernet 0/3)# ipv6 enable
Device D(cfg-if-GigabitEthernet 0/3)# ipv6 address 2001:3::2/64
Device D(cfg-if-GigabitEthernet 0/3)# ipv6 ospf 1 set-area 2
Device D(cfg-if-GigabitEthernet 0/3)# exit
```

5. Verification

(1) Display the following information on Device A.

Display OSPFv3 neighbor information.

```
Device A# display ipv6 ospf-instance peer
OSPFv3 Process (1), 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           Dead Time   Instance ID  Interface
2.2.2.2           1   Full/BDR        00:00:30    0            GigabitEthernet 0/1
3.3.3.3           1   Full/BDR        00:00:35    0            GigabitEthernet 0/2
```

Display OSPFv3 routing table.

```
Device A# display ipv6 route-config ospf

IPv6 routing table name - Default - 0 entries
Codes:  C - Connected, L - Local, S - Static
        R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
        N1 - OSPFv3 NSSA external type 1, N2 - OSPFv3 NSSA external type 2
        E1 - OSPFv3 external type 1, E2 - OSPFv3 external type 2
        SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
        IA - Inter area

O   IA  2001:3::/64 [110/20] via FE80::2D0:F8FF:FE22:4524, GigabitEthernet 0/1
```

(2) Display the following information on Device B.

Display OSPFv3 neighbor information.

```
Device B# display ipv6 ospf-instance peer
OSPFv3 Process (1), 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           Dead Time    Instance ID  Interface
1.1.1.1           1   Full/DR         00:00:30     0            GigabitEthernet 0/1
4.4.4.4           1   Full/BDR        00:00:35     0            GigabitEthernet 0/2
```

Display OSPFv3 routing table.

```
Device B# display ipv6 route-config ospf

IPv6 routing table name - Default - 0 entries
Codes:  C - Connected, L - Local, S - Static
        R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
        N1 - OSPFv3 NSSA external type 1, N2 - OSPFv3 NSSA external type 2
        E1 - OSPFv3 external type 1, E2 - OSPFv3 external type 2
        SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
        IA - Inter area

O   IA  2001:2::/64 [110/20] via FE80::2D0:F8FF:FE22:4536, GigabitEthernet 0/1
```

(3) Display the following information on Device C.

Display OSPFv3 neighbor information.

```
Device C# display ipv6 ospf-instance peer
OSPFv3 Process (1), 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           Dead Time    Instance ID  Interface
1.1.1.1           1   Full/DR         00:00:30     0            GigabitEthernet 0/3
```

Display OSPFv3 routing table.

```
Device C# display ipv6 route-config ospf

IPv6 routing table name - Default - 0 entries
Codes:  C - Connected, L - Local, S - Static
        R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
```

```
          N1 - OSPFv3 NSSA external type 1, N2 - OSPFv3 NSSA external type 2
          E1 - OSPFv3 external type 1, E2 - OSPFv3 external type 2
          SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
          IA - Inter area

O   IA   2001:1::/64 [110/2] via FE80::2D0:F8FF:FE22:4537, GigabitEthernet 0/3
O   IA   2001:3::/64 [110/3] via FE80::2D0:F8FF:FE22:4537, GigabitEthernet 0/3
```

(4) Display the following information on Device D.

Display OSPFv3 neighbor information.

```
Device D# display ipv6 ospf-instance peer
OSPFv3 Process (1), 1 Neighbors, 1 is Full:
Neighbor ID      Pri    State           Dead Time     Instance ID     Interface
2.2.2.2            1    Full/DR         00:00:30      0               GigabitEthernet 0/3
```

Display OSPFv3 routing table.

```
Device D# display ipv6 route-config ospf

IPv6 routing table name - Default - 0 entries
Codes:   C - Connected, L - Local, S - Static
          R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
          N1 - OSPFv3 NSSA external type 1, N2 - OSPFv3 NSSA external type 2
          E1 - OSPFv3 external type 1, E2 - OSPFv3 external type 2
          SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
          IA - Inter area

O   IA   2001:1::/64 [110/2] via FE80::2D0:F8FF:FE22:4547, GigabitEthernet 0/3
O   IA   2001:2::/64 [110/3] via FE80::2D0:F8FF:FE22:4547, GigabitEthernet 0/3
```

Verify that 2001:2::2/64 can be pinged successfully on Device D.

```
Device D# ping 2001:2::2
Sending 5, 100-byte ICMP Echoes to 2001:2::2, timeout is 2 seconds:
< press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 5/9/14 ms.
```

6. Configuration Files

- Device A configuration file

```
!
ipv6 router-protocol ospf 1
 router-id 1.1.1.1
!
interface gigabitethernet 0/1
unset switch-port
 ipv6 enable
 ipv6 address 2001:1::1/64
 ipv6 ospf 1 set-area 0
```

```
!
interface gigabitethernet 0/2
unset switch-port
 ipv6 enable
 ipv6 address 2001:2::1/64
 ipv6 ospf 1 set-area 1
!
```

- Device B configuration file

```
!
ipv6 router-protocol ospf 1
 router-id 2.2.2.2
!
interface gigabitethernet 0/1
unset switch-port
 ipv6 enable
 ipv6 address 2001:1::2/64
 ipv6 ospf 1 set-area 0
!
interface gigabitethernet 0/2
unset switch-port
 ipv6 enable
 ipv6 address 2001:3::1/64
 ipv6 ospf 1 set-area 2
!
```

- Device C configuration file

```
!
ipv6 router-protocol ospf 1
 router-id 3.3.3.3
!
interface gigabitethernet 0/3
unset switch-port
 ipv6 enable
 ipv6 address 2001:2::2/64
 ipv6 ospf 1 set-area 1
!
```

- Device D configuration file

```
!
ipv6 router-protocol ospf 1
 router-id 4.4.4.4
!
interface gigabitethernet 0/3
unset switch-port
 ipv6 enable
 ipv6 address 2001:3::2/64
 ipv6 ospf 1 set-area 2
```

```
!
```

7. Common Errors

- No IPv6 address is configured for the interface.
- OSPFv3 cannot be enabled because the IPv6 unicast routing function is disabled.
- The area IDs configured for adjacent interfaces are inconsistent.
- The same router ID is configured on multiple devices, resulting in a router ID conflict.

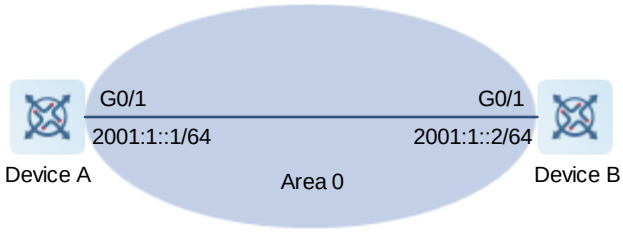
1.25.2 Configuring OSPFv3 Authentication

1. Requirements

OSPFv3 runs between Device A and Device B. MD5 authentication is configured to improve security of an OSPFv3 network.

2. Topology

Figure 1-10 Topology of OSPFv3 Authentication



3. Notes

Configure MD5 authentication on interfaces of all the devices.

4. Procedure

(1) Configure basic OSPFv3 functions (omitted).

(2) Enable OSPFv3 authentication on an interface.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 ospf auth-algorithm ipsec spi 256 md5
01234567890123456789012345678912
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
```

```
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 ospf auth-algorithm ipsec spi 256 md5
01234567890123456789012345678912
```

5. Verification

Verify whether neighbors of Device A and Device B are established.

- Check the neighbor information of Device A.

```
Device A# display ipv6 ospf peer
OSPFv3 Process (1), 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           Dead Time     Instance ID   Interface
2.2.2.2           1   Full/DR         00:00:38      0             GigabitEthernet 0/1
```

- Check the neighbor information of Device B.

```
Device B# display ipv6 ospf peer
OSPFv3 Process (1), 1 Neighbors, 1 is Full:
Neighbor ID     Pri   State           Dead Time     Instance ID   Interface
1.1.1.1           1   Full/BDR        00:00:38      0             GigabitEthernet 0/1
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
unset switch-port
  ipv6 ospf auth-algorithm ipsec spi 256 md5 01234567890123456789012345678912
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
unset switch-port
  ipv6 ospf auth-algorithm ipsec spi 256 md5 01234567890123456789012345678912
!
```

7. Common Errors

- The configured authentication modes are inconsistent.
- The configured authentication keys are inconsistent.

1.25.3 Configuring a Stub Area

1. Requirements

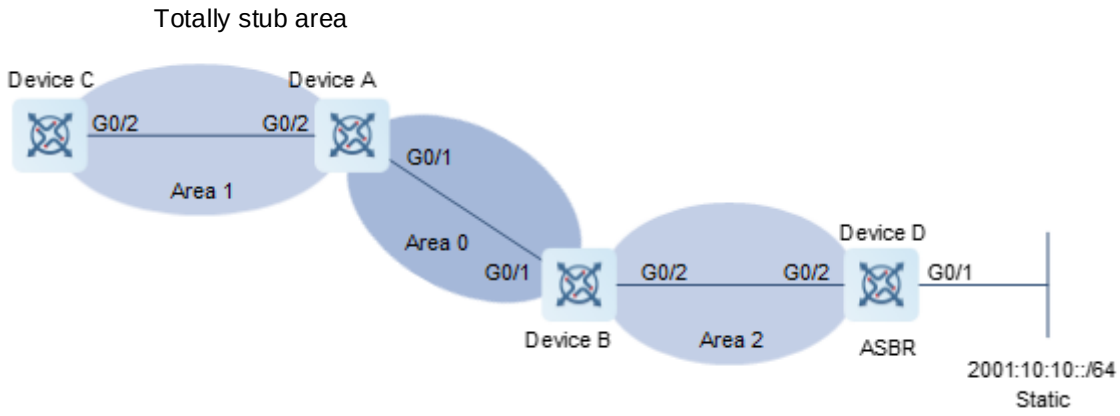
Devices A, B, C, and D are interconnected through the OSPFv3 routing protocol.

As ABRs, Devices A and B are responsible for the routing information transfer between OSPFv3 areas, and Device D functions as an ASBR to introduce the external static routes.

To reduce the number of LSAs in Area 1 and save the device performance, Area 1 is configured as a totally stub area.

2. Topology

Figure 1-11 Topology for a Stub Area



3. Notes

- Configure IPv6 addresses on interfaces of all the devices.
- Configure the basic OSPFv3 functions for all the devices.
- Introduce an external static route to Device D.
- Configure Area 1 as a stub area on Device A and Device C.

4. Procedure

(1) Configure IPv6 addresses on interfaces of all the devices and enable basic OSPFv3 functions (omitted).

(2) Introduce an external static route to Device D.

```
Device D> privilege
Device D# config-system terminal
Device D(cfg)# ipv6 router-protocol ospf 1
Device D(cfg-router)# redistribute-route static
```

(3) Configure Area 1 as a stub area on Device A and filter Type 3 LSAs.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ipv6 router-protocol ospf 1
Device A(cfg-router)# area 1 stub unset-summary
```

(4) Configure Area 1 as a stub area on Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# ipv6 router-protocol ospf 1
Device C(cfg-router)# area 1 stub
```

5. Verification

Run the display ipv6 route-config ospf command on Device C to display the routing table. Verify that there is only one default inter-area route, and no external static route is introduced from Device D.

```
Device C# display ipv6 route-config ospf

IPv6 routing table name - Default - 0 entries
Codes:  C - Connected, L - Local, S - Static
        R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
        N1 - OSPFv3 NSSA external type 1, N2 - OSPFv3 NSSA external type 2
        E1 - OSPFv3 external type 1, E2 - OSPFv3 external type 2
        SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
        IA - Inter area

O   IA::/0 [110/3] via FE80::2D0:F8FF:FE22:4547, GigabitEthernet 0/2
```

6. Configuration Files

- Device A configuration file

```
!
ipv6 router-protocol ospf 1
area 1 stub unset-summary
!
```

- Device C configuration file

```
!
ipv6 router-protocol ospf 1
area 1 stub
!
```

- Device D configuration file

```
!
ipv6 router-protocol ospf 1
redistribute-route static
!
```

7. Common Errors

- Configurations of the area type are inconsistent on devices in the same area.
- External routes cannot be introduced because route redistribution is configured on a device in the stub area.

1.25.4 Configuring an NSSA

1. Requirements

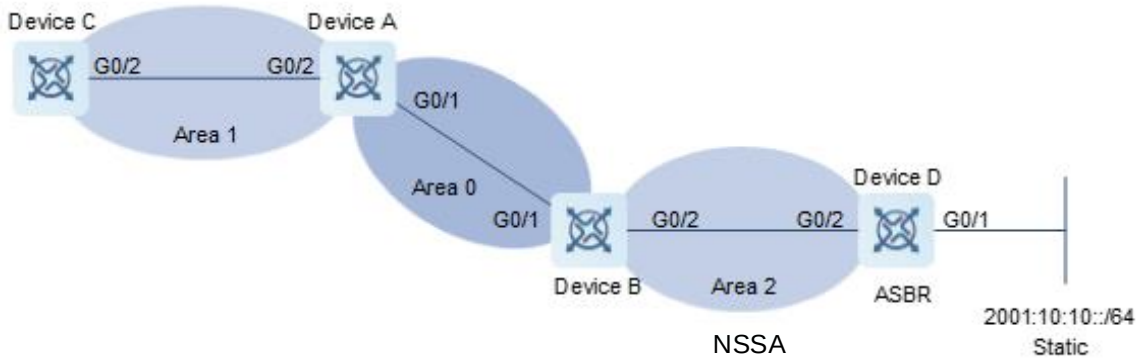
Devices A, B, C, and D are interconnected through the OSPFv3 routing protocol.

As ABRs, Devices A and B are responsible for the routing information transfer between OSPFv3 areas, and Device D functions as an ASBR to introduce the external static routes.

To reduce the number of LSAs in Area 2 and save the device performance, Area 2 is configured as an NSSA.

2. Topology

Figure 1-12 Topology for an NSSA



3. Notes

- Configure IPv6 addresses on interfaces of all the devices.
- Configure the basic OSPFv3 functions for all the devices.
- Introduce an external static route to Device D.
- Configure Area 2 as an NSSA on Device B and Device D.

4. Procedure

(1) Configure IPv6 addresses on interfaces of all the devices and enable basic OSPFv3 functions (omitted).

(2) Introduce an external static route to Device D. Configure Area 2 as an NSSA.

```
Device D> privilege
Device D# config-system terminal
Device D(cfg)# ipv6 router-protocol ospf 1
Device D(cfg-router)# area 2 nssa
Device D(cfg-router)# redistribute-route static
```

(3) Configure Area 2 as an NSSA on Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ipv6 router-protocol ospf 1
Device B(cfg-router)# area 2 nssa
```

5. Verification

(1) Run the **display ipv**6 **ospf-instance database-info** command on Device D to display the database information and verify that Type 7 LSAs are generated.

```
Device D# display ipv6 ospf-instance database-info nssa-external-type
                OSPFv3 Router with ID (1.1.1.1) (Process 1)
                    NSSA-external-LSA (Area 0.0.0.1)
  LS age: 1196
  LS Type: NSSA-external-LSA
  Link State ID: 0.0.0.3
```

```
  Advertising Router: 1.1.1.1
  LS Seq Number: 0x80000004
  Checksum: 0x1F25
  Length: 52
      Metric Type: 2 (Larger than any link state path)
      Metric: 20
Prefix: 2001:10::/64
Prefix Options: 8 (P|-|-|-)
Forwarding Address: 4000::1
```

(2) Run the **display ipv6 route-config ospf** command on Device A to display the routing table and verify that an external static route is introduced by Router D.

```
Device A# display ipv6 route-config ospf

IPv6 routing table name - Default - 0 entries
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
         N1 - OSPFv3 NSSA external type 1, N2 - OSPFv3 NSSA external type 2
         E1 - OSPFv3 external type 1, E2 - OSPFv3 external type 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area

O   N2   2001:10::/64 [110/20] via FE80::2D0:F8FF:FE22:4547, GigabitEthernet 0/1
```

6. Configuration Files

- Device B configuration file

```
!
ipv6 router-protocol ospf 1
area 2 nssa
!
```

- Device D configuration file

```
!
ipv6 router-protocol ospf 1
area 2 nssa
 redistribute-route static
!
```

7. Common Errors

- Configurations of the area type are inconsistent on devices in the same area.

1.25.5 Selecting an OSPFv3 DR

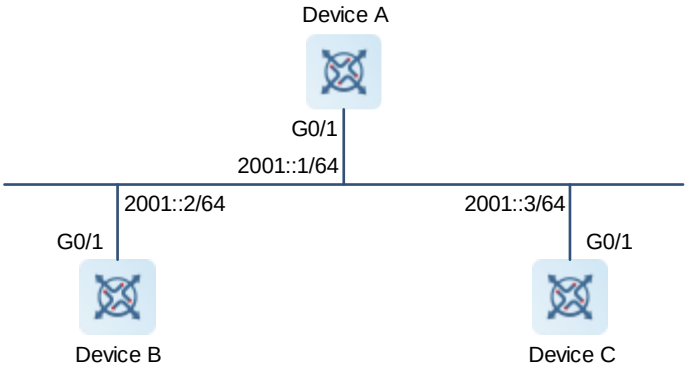
1. Requirements

Devices A, B, and C are in the same broadcast network.

The priority of Device A is set to 10 and Device A is configured as a DR. The priority of Device B is set to 5 and Device B is configured as a BDR. The priority of Device C is set to 0 and Device C does not participate in DR election.

2. Topology

Figure 1-13 Topology of OSPFv3 DR Configuration



3. Notes

- Configure IPv6 addresses on interfaces of all the devices.
- Configure the basic OSPFv3 functions for all the devices.
- Modify the DR election priority of all the devices.

4. Procedure

(1) Configure basic functions of OSPFv3.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ipv6 router-protocol ospf 1
Device A(cfg-router)# router-id 1.1.1.1
Device A(cfg-router)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ipv6 router-protocol ospf 1
Device B(cfg-router)# router-id 2.2.2.2
Device B(cfg-router)# exit
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# ipv6 router-protocol ospf 1
Device C(cfg-router)# router-id 3.3.3.3
Device C(cfg-router)# exit
```

(2) Configure IP addresses of interfaces, configure OSPFv3, and modify the DR election priority of all the devices.

Configure Device A.

```
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 address 2001::1/64
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 1 set-area 0
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 ospf priority-value 10
```

Configure Device B.

```
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 address 2001::2/64
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 1 set-area 0
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 ospf priority-value 5
```

Configure Device C.

```
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 address 2001::3/64
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 1 set-area 0
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 ospf priority-value 0
```

5. Verification

- Run the **display ipv6 ospf-instance interface** command on Device C to display the DR and BDR.

```
Device C# display ipv6 ospf-instance interface
GigabitEthernet 0/1 is up, line protocol is up
  Interface ID 4098
  IPv6 Prefixes
    fe80::250:56ff:feb5:820b/64 (Link-Local Address)
    2001::3/64
  OSPFv3 Process (1), Area 0.0.0.0, Instance ID 0
    Router ID 3.3.3.3, Network Type BROADCAST, Cost: 1
    Transmit Delay is 1 sec, State DROther, Priority 0
    Designated Router (ID) 1.1.1.1
      Interface Address fe80::250:56ff:feb5:694e
    Backup Designated Router (ID) 2.2.2.2
      Interface Address fe80::250:56ff:feb5:fdb4
    Timer period configured, Hello 10, Dead 40, Wait 40, Retransmit 5
    Neighbor Count is 2, Adjacent peer count is 2
  Hello received 73 sent 39, DD received 8 sent 7
  LS-Req received 1 sent 2, LS-Upd received 9 sent 5
  LS-Ack received 5 sent 5, Discarded 6
```

- Run the **display ipv6 ospf-instance peer** command on Device A to display DROther of the broadcast network as Device C.

```
Device A# display ipv6 ospf-instance peer
```

```
OSPFv3 Process (1), 2 Neighbors, 2 is Full:
Neighbor ID     Pri   State           BFD State  Dead Time     Instance ID  Interface
2.2.2.2           5   Full/BDR        -          00:00:36     0            GigabitEthernet 0/1
3.3.3.3           0   Full/DROther    -          00:00:38     0            GigabitEthernet 0/1
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
unset switch-port
ipv6 address 2001::1/64
ipv6 ospf priority-value 10
ipv6 ospf 1 set-area 0
!
ipv6 router-protocol ospf 1
 router-id 1.1.1.1
 gr
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
unset switch-port
 ipv6 address 2001::2/64
ipv6 ospf priority-value 5
 ipv6 ospf 1 set-area 0
!
ipv6 router-protocol ospf 1
 router-id 2.2.2.2
 gr
!
```

- Device C configuration file

```
!
interface gigabitethernet 0/1
unset switch-port
ipv6 address 2001::3/64
 ipv6 ospf priority-value 0
 ipv6 ospf 1 set-area 0
!
ipv6 router-protocol ospf 1
 router-id 3.3.3.3
 gr
!
```

1.25.6 Configuring BFD for OSPFv3

1. Requirements

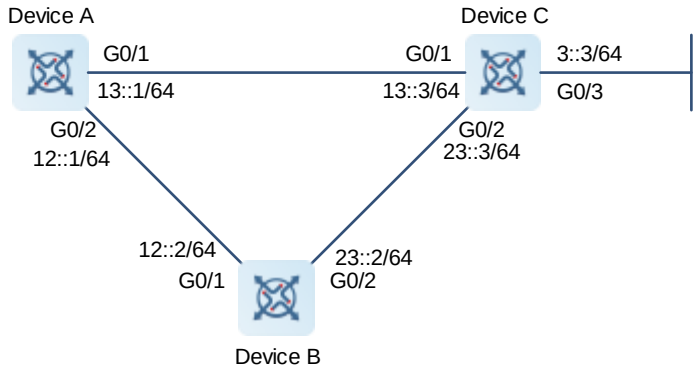
Devices A, B, and C are interconnected through OSPFv3.

The active link from Device A to 3::3 is Device A > Device C, and the standby link is Device A > Device B > Device C.

When the link from Device A to Device C fails, BFD can detect the failure and notify OSPFv3 of the failure, and quickly switch to the standby link.

2. Topology

Figure 1-14 Configuring BFD for OSPFv3



3. Notes

- Configure IPv6 addresses on interfaces of all the devices.
- Configure the basic OSPFv3 functions for all the devices.
- Enable the BFD function in the OSPFv3 routing process mode.

4. Procedure

(1) Configure IP addresses for interfaces and configure basic OSPFv3 functions.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/2)# ipv6 address 12::1/64
Device A(cfg-if-GigabitEthernet 0/2)# ipv6 ospf 1 set-area 0
Device A(cfg-if-GigabitEthernet 0/2)# exit
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 address 13::1/64
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# ipv6 router-protocol ospf 1
```

```
Device A(cfg-router)# router-id 1.1.1.1
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 address 12::2/64
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 1 set-area 0
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/2)# ipv6 address 23::2/64
Device B(cfg)# ipv6 router-protocol ospf 1
Device B(cfg-router)# router-id 2.2.2.2
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 address 13::3/64
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 1 set-area 0
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface gigabitethernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/2)# ipv6 address 23::3/64
Device C(cfg-if-GigabitEthernet 0/2)# ipv6 ospf 1 set-area 0
Device C(cfg-if-GigabitEthernet 0/2)# exit
Device C(cfg)# ipv6 router-protocol ospf 1
Device C(cfg-router)# router-id 3.3.3.3
```

(2) Enable the BFD function in the OSPFv3 routing process mode.

Configure Device A.

```
Device A(cfg-router)# bfd whole-interfaces
```

Configure Device B.

```
Device B(cfg-router)# bfd whole-interfaces
```

Configure Device C.

```
Device C(cfg-router)# bfd whole-interfaces
```

5. Verification

- Normal network environment

Check the routing table of Device A.

```
Device A# display ipv6 route-config ospf

IPv6 routing table name - Default - 11 entries
```

```
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
         N1 - OSPFv3 NSSA external-routes class 1, N2 - OSPFv3 NSSA external-routes class 2
         E1 - OSPFv3 external-routes class 1, E2 - OSPFv3 external-routes class 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area-routes, EV - BGP EVPN, N - Nd to host

O        3::3/128 [110/1] via FE80::2D0:F8FF:FE22:358D, GigabitEthernet 0/1
O        23::/64 [110/2] via FE80::2D0:F8FF:FE22:358D, GigabitEthernet 0/1
                  [110/2] via FE80::274:9CFF:FEEE:53CB, GigabitEthernet 0/2
```

Run **traceroute 3::3** on Device A.

```
Device A# traceroute 3::3
  < press Ctrl+C to break >
Tracing the route to 3::3

 1                                    3::3      5 msec     8 msec     10 msec
```

- Link from Device A to Device C fails

  Check the routing table of Device A.

```
Device A# display ipv6 route-config ospf

IPv6 routing table name - Default - 7 entries
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
         N1 - OSPFv3 NSSA external-routes class 1, N2 - OSPFv3 NSSA external-routes class 2
         E1 - OSPFv3 external-routes class 1, E2 - OSPFv3 external-routes class 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area-routes, EV - BGP EVPN, N - Nd to host

O        3::3/128 [110/2] via FE80::274:9CFF:FEEE:53CB, GigabitEthernet 0/2
O        23::/64 [110/2] via FE80::274:9CFF:FEEE:53CB, GigabitEthernet 0/2
```

Run **traceroute 3::3** on Device A.

```
Device A# traceroute 3::3
  < press Ctrl+C to break >
Tracing the route to 3::3

 1                                   12::2      6 msec     9 msec      9 msec
 2                                    3::3      7 msec     9 msec     10 msec
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/1
unset switch-port
ipv6 address 13::1/64
```

```
ipv6 ospf 1 set-area 0
!
interface gigabitethernet 0/2
unset switch-port
ipv6 address 12::1/64
ipv6 ospf 1 set-area 0
!
ipv6 router-protocol ospf 1
 router-id 1.1.1.1
 gr
 bfd whole-interfaces
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
unset switch-port
ipv6 address 12::2/64
ipv6 ospf 1 set-area 0
!
interface gigabitethernet 0/2
unset switch-port
ipv6 address 23::2/64
ipv6 ospf 1 set-area 0
!
ipv6 router-protocol ospf 1
 router-id 2.2.2.2
 gr
 bfd whole-interfaces
!
```

- Device C configuration file

```
!
interface gigabitethernet 0/1
unset switch-port
ipv6 address 13::3/64
ipv6 ospf 1 set-area 0
!
interface gigabitethernet 0/2unset switch-port
ipv6 address 23::3/64
ipv6 ospf 1 set-area 0
!
ipv6 router-protocol ospf 1
 router-id 3.3.3.3
 gr
 bfd whole-interfaces
!
```

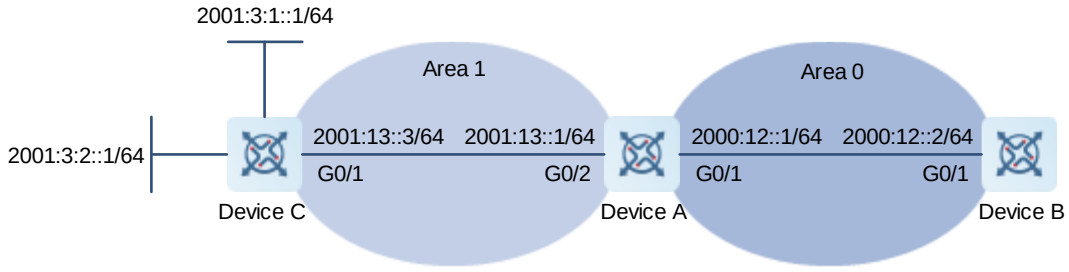
1.25.7 Configuring OSPFv3 Route Summarization

1. Requirements

Devices A and B reside in OSPFv3 Area 0, and Devices A and C reside in OSPFv3 Area 1. Two static routes are introduced to Device C. To reduce the number of routing entries, Device C summarizes introduced static routes as 2001:3::/32 and advertises the routes to the OSPFv3 domain. Device A summarizes routes in Area 1 as 2001:13::/32 and advertises the routes to the OSPFv3 neighbors.

2. Topology

Figure 1-15 Topology of OSPFv3 Route Summarization



3. Notes

- Configure IPv6 addresses on interfaces of all the devices.
- Configure the basic OSPFv3 functions for all the devices.
- Redistribute two static routes on Device C and configure external route summarization.
- Summarize routes in Area 1 on Device A.

4. Procedure

(1) Configure basic functions of OSPFv3 (omitted).

(2) Redistribute two static routes on Device C and summarize the redistributed routes as 2001:3::/32.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# ipv6 router-protocol ospf 1
Device C(cfg-router)# redistribute-route static
Device C(cfg-router)# summary-prefix 2001:3::/32
```

(3) Summarize routes in Area 1 as 2001:13::/32 on Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ipv6 router-protocol ospf 1
Device A(cfg-router)# area 1 summary 2001:13::/32
```

5. Verification

- Display OSPFv3 routes on Device A. Summarized static routes are displayed, and route details cannot be displayed.

```
Device A# display ipv6 route-config ospf
```

```
IPv6 routing table name - Default - 11 entries
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
         N1 - OSPFv3 NSSA external type 1, N2 - OSPFv3 NSSA external type 2
         E1 - OSPFv3 external type 1, E2 - OSPFv3 external type 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area, EV - BGP EVPN, N - Nd to host

O   E2   2001:3::/32 [110/20] via FE80::250:56FF:FEB5:7BF5, GigabitEthernet 0/2
O        2001:13::/32 [110/0] via ::1, Null0
```

- Display OSPFv3 routes on Device B. Area 1 summarized routes are displayed, and route details cannot be displayed.

```
Device B# display ipv6 route-config ospf

IPv6 routing table name - Default - 7 entries
Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPFv3, B - BGP, I - IS-IS, V - Overflow route
         N1 - OSPFv3 NSSA external type 1, N2 - OSPFv3 NSSA external type 2
         E1 - OSPFv3 external type 1, E2 - OSPFv3 external type 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area, EV - BGP EVPN, N - Nd to host

O   E2   2001:3::/32 [110/20] via FE80::250:56FF:FEB5:694E, GigabitEthernet 0/1
O   IA   2001:13::/32 [110/2] via FE80::250:56FF:FEB5:694E, GigabitEthernet 0/1
```

6. Configuration Files

- Device A configuration file

```
!
 router-id 1.1.1.1
 gr
 area 1 summary 2001:13::/32
!
```

- Device C configuration file

```
!
 router-id 3.3.3.3
 gr
 redistribute-route static
 summary-prefix 2001:3::/32
!
```

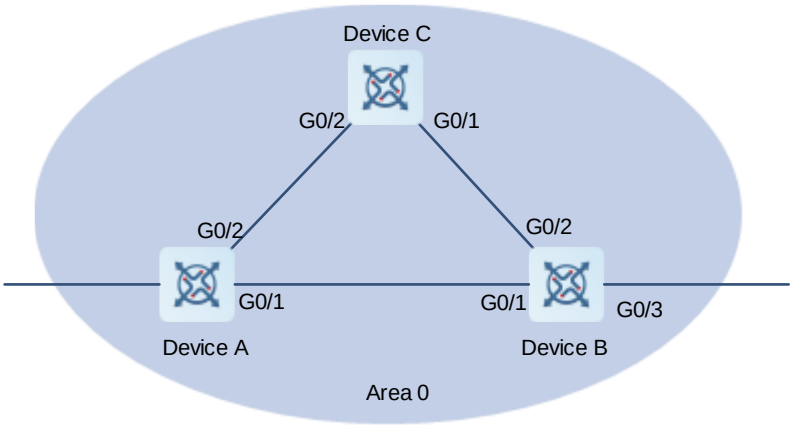
1.25.8 Configuring FRR

1. Requirements

Devices A, B, and C belong to the same OSPFv3 area, and every two of them interwork with each other through the OSPF protocol. When a link of device A fails, services can be quickly switched to the backup link. The switching latency should be as low as possible.

2. Topology

Figure 1-16 FRR Topology



3. Notes

- Configure FRR on device A.
- Configure **carrier-delay 0** on interfaces of device A.

4. Procedure

(1) Configure IPv6 addresses for interfaces of all devices (omitted).

(2) Configure the OSPFv3 basic functions on all the devices (omitted).

(3) Configure FRR on device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ipv6 router-protocol ospf 1
Device A(cfg-router)# frr lfa
Device A(cfg-router)# exit
```

(4) Configure **carrier-delay 0** on all interfaces of device A.

```
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# carrier-delaytime 0
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/2)# carrier-delaytime 0
```

5. Verification

On device A, check the routing table to verify that a backup route exists for the entry 192.168.4.0/24.

```
Device A# display ipv6 route-config frr | begin 12::1/128
O          12::1/128[m]    [110/12] via FE80::A00:27FF:FE47:9335, GigabitEthernet 0/1
                     [b]    [110/12] via FE80::A00:27FF:FE47:9336, GigabitEthernet 0/2
```

6. Configuration Files

Device A configuration file

```
!
ipv6 router-protocol ospf 1
frr lfa
!
interface gigabitethernet 0/1
unset switch-port
carrier-delay 0
!
interface gigabitethernet 0/2

 carrier unset switch-port-delay 0
!
```

Contents

1 Configuring IS-IS 1
1.1 Overview 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 13
1.2 Restrictions and Guidelines 14
1.3 Configuration Task Summary 14
1.4 Configuring Basic Settings of IS-IS 17
1.4.1 Overview 17
1.4.2 Restrictions and Guidelines 17
1.4.3 Procedure 17
1.5 Configuring IS-IS network Levels 19
1.5.1 Overview 19
1.5.2 Restrictions and Guidelines 19
1.5.3 Procedure 19
1.6 Configuring IS-IS Hello Packets 20
1.6.1 Overview 20
1.6.2 Procedure 20
1.7 Configuring IS-IS LSPs 21
1.7.1 Overview 21
1.7.2 Restrictions and Guidelines 21
1.7.3 Procedure 22

1.8 Configuring IS-IS SNPs 24
1.8.1 Overview 24
1.8.2 Restrictions and Guidelines 24
1.8.3 Procedure 24
1.9 Configuring IS-IS Route Advertisement 25
1.9.1 Overview 25
1.9.2 Restrictions and Guidelines 25
1.9.3 Procedure 25
1.10 Configuring IS-IS Route Summarization 26
1.10.1 Overview 26
1.10.2 Procedure 26
1.11 Configuring an IS-IS Passive Interface 26
1.11.1 Overview 26
1.11.2 Restrictions and Guidelines 27
1.11.3 Procedure 27
1.12 Controlling IS-IS Routing Factors 28
1.12.1 Overview 28
1.12.2 Restrictions and Guidelines 28
1.12.3 Procedure 28
1.13 Configuring the Priority of an IS-IS Device 29
1.13.1 Overview 29
1.13.2 Restrictions and Guidelines 29
1.13.3 Procedure 29
1.14 Configuring IS-IS Authentication 30

1.14.1 Overview 30
1.14.2 Restrictions and Guidelines 30
1.14.3 Procedure 31
1.15 Configuring IS-IS GR 33
1.15.1 Overview 33
1.15.2 Restrictions and Guidelines 33
1.15.3 Procedure 33
1.16 Configuring IS-IS NSR 34
1.16.1 Overview 34
1.16.2 Restrictions and Guidelines 34
1.16.3 Procedure 34
1.17 Correlating IS-IS with BFD 35
1.17.1 Overview 37
1.17.2 Restrictions and Guidelines 37
1.17.3 Procedure 37
1.18 Configuring the IS-IS Overload Bit 38
1.18.1 Overview 38
1.18.2 Restrictions and Guidelines 38
1.18.3 Procedure 39
1.19 Configuring IS-IS VRF 40
1.19.1 Overview 40
1.19.2 Restrictions and Guidelines 40
1.19.3 Procedure 40
1.20 Configuring IS-IS MTR 41

1.20.1 Overview 41
1.20.2 Restrictions and Guidelines 41
1.20.3 Procedure 41
1.21 Configuring IS-IS SNMP 41
1.21.1 Overview 41
1.21.2 Restrictions and Guidelines 42
1.21.3 Procedure 42
1.22 Enabling Super VLAN for IS-IS 42
1.22.1 Overview 42
1.22.2 Restrictions and Guidelines 43
1.22.3 Procedure 43
1.23 Configuring IS-IS Two-way Maintenance 43
1.23.1 Overview 43
1.23.2 Restrictions and Guidelines 43
1.23.3 Procedure 43
1.24 Configuring IS-IS BGP-LS 44
1.24.1 Overview 44
1.24.2 Restrictions and Guidelines 44
1.24.3 Procedure 44
1.25 Configuring Extended Feature of IS-IS 44
1.25.1 Overview 46
1.25.2 Configuration Tasks 46
1.25.3 Configuring the Maximum Metric of a Neighbor 47
1.25.4 Configuring the Maximum Number of Area Addresses 47

1.25.5 Configuring SPF Calculation Interval....48
1.25.6 Canceling Three-Way Handshake of a P2P Network....48
1.25.7 Enabling TLV to Be Padded Based on Supported Protocol....49
1.25.8 Configuring the Reachable Time of a Suppressed Neighbor....50
1.25.9 Enabling Adjacency Event Output....50
1.26 Monitoring....51
1.27 Configuration Examples....53
1.27.1 Performing Basic Authentication Configuration of IS-IS....53
1.27.2 Configuring IS-IS MTR....59
1.27.3 Configuring IS-IS Route Summarization....66
1.27.4 Correlating IS-IS with BFD....68

1 Configuring IS-IS

1.1 Overview

1.1.1 Overview

Intermediate System to Intermediate System (IS-IS) is an extensible, robust, and easy-to-use Interior Gateway Protocol (IGP) for route selection and applicable to an IP and ISO (connectionless network service) CLNS dual environment network.

IS-IS was designed for the Connection-Less Network Protocol (CLNP) at the very start. After TCP/IP became mainstream protocols, IETF modified IS-IS so that this protocol could apply to TCP/IP environment. The modified IS-IS is referred to as Integrated IS-IS or Dual IS-IS.

IS-IS has the common characteristics of a link state protocol. It sends Hello packets to discover and maintain neighbor relationships, and sends Link State Protocol Data Units (LSPs) to neighbors to advertise its link state. IS-IS supports Level-1 routing and Level-2 routing. All devices at the same level maintain the same link state database (LSDB), which stores the LSPs generated by the devices to notify each other of the network topology at this level. Each device uses the Dijkstra shortest path first (SPF) algorithm to perform best-route computation, path selection, and fast convergence.

Note

In this chapter, the term "router" refers to any network device that supports the routing function. These network devices can be Layer-3 switches, routers, or firewalls.

1.1.2 Principles

1. Basic Concepts

- IS-IS network hierarchy

  As shown in Figure 1-1, an IS-IS network is divided into Level-1 and Level-2. The nodes on which devices exchange information in the same area form one level (Level-1).

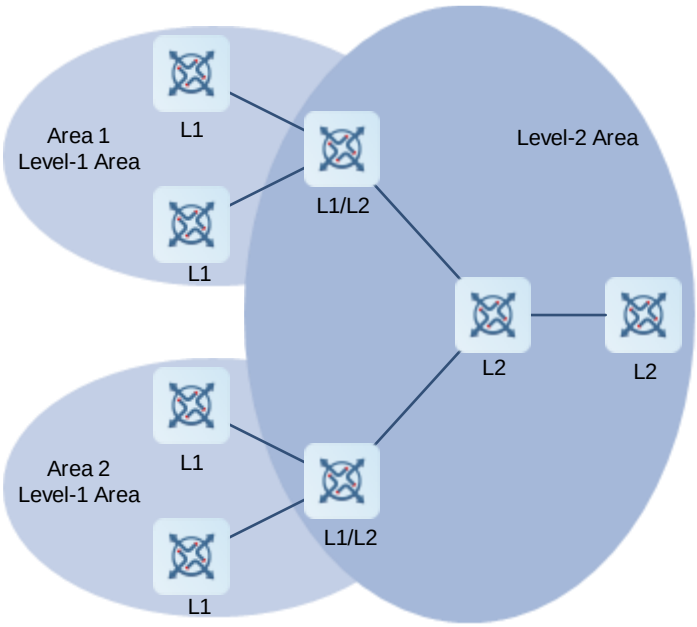
  - Level-1 routers perform routing in the local area and they can form neighbor relationship with local Level-1 routers and Level-1/Level-2 routers and maintain Level-1 LSDB. This LSDB includes routing information of local area and packets destined for another area are forwarded to the closest Level-1/Level-2 router.
  - Level-2 routers perform inter-area routing and they can form neighbor relationship with Level-2 routers in the same area or another area and Level-1/Level-2 routers and maintain Level-2 LSDB. This LSDB includes inter-area routing information. All Level-2 routers and Level-1/Level-2 routers form a backbone network of a routing domain and take charge of communication between different areas. The backbone network must be physically connected. Only Level-2 routers can exchange data packets and routing information with routers outside the routing domain.
  - Level-1/Level-2 routers, belonging to Level-1 and Level-2 areas, can form Level-1 neighbor relationship with Level-1 routers in the same area and Level-1/Level-2 routers and form Level-2 neighbor

relationship with Level-2 routers in the same or another area and Level-1/Level-2 routers. Level-1 routers must be connected to another area through Level-1/Level-2 routers. Level-1/Level-2 routers maintain two LSDBs. Level-1 LSDB is used for internal routing, and Level-2 LSDB is used for inter-area routing. By default, Level-1/Level-2 routers redistribute Level-1 routes to Level-2 routers. If there are many Level-1 routes, the routes must be summarized and then redistributed to Level-2 routers.

All devices in an area know the network topology of the area and exchange data within the area. Level-1/Level-2 devices are border devices that belong to different areas and provide inter-area connections. Areas are connected by Level-2 devices. The border devices in various areas form a Level-2 backbone network for inter-area data exchange.

Level-1 devices are only interested in the local area's topology, including all nodes in the local area and the next-hop devices destined for the nodes. Level-1 devices access other areas through Level-2 devices and forward packets from a target network outside the local area to the closest Level-2 device.

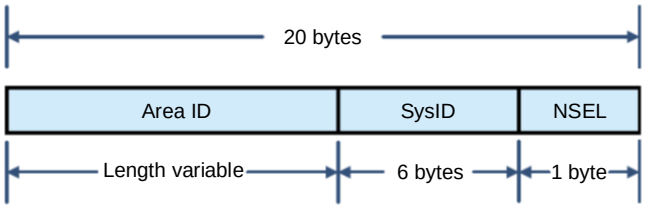
Figure 1-1 IS-IS Network Architecture



- IS-IS address coding mode

  An IS-IS address is called a network entity title (NET) and it consists of an area ID, a system ID, and a network service access point (NSAP) ID. An IS-IS address ranges from 8 to 20 bytes.

Figure 1-2 NET Address Format



  - The area ID identifies the routing domain length in an area and is fixed relative to the routing domain. It

ranges from 1 to 13 bytes.

The area ID consists of an authority and format identifier (AFI), an initial domain identifier (IDI), and a high order DSP (HODSP). The AFI represents an address assignment authority and address format, the IDI is used to identify a domain, and the HODSP is used to distinguish areas.

- The system ID is unique in an autonomous system (AS).

  Generally, a router ID (for example, 192.168.10.1) is translated to a system ID by extending each decimal in an IP address to three decimals. For if the decimals are less than three, 0 is added for supplementation. An extended IP address is 192.168.010.001. After the extended IP address is divided into three parts, the system ID is 1921.6801.0001.

- The NSAP is a network selector and sometimes called SEL. In IS-IS, SEL is typically set to 00 to indicate a device.

- IS-IS packet types

  There are three types of IS-IS packets:

  - LSPs

    LSPs are used to transmit link state records within an area and are classified into Level-1 LSPs and Level-2 LSPs. LSPs are only flooded at the local level.

  - IS-to-IS Hello (IIH) PDUs

    IIH PDUs are used to maintain adjacency. They carry multicast MAC addresses used to determine whether other systems run IS-IS.

  - Sequence number PDUs (SNPs), including complete SNP (CSNP) and partial SNP (PSNP)

    CSNPs are used for LSDB synchronization. By default, a designated IS (DIS) sends a CSNP every 10s in a broadcast network. In a P2P network, a CSNP is sent only after a neighbor relationship is established.

    PSNPs are also used for LSDB synchronization.

- IS-IS supported type-length-value (TLV) types

  IS-IS supports 26 types of TLV.

  Table 1 lists IS-IS supported TLV types.

Table 1-1 IS-IS Supported TLV Types

TLV Code	Description
Code=1	Area address
Code=2	Priority of an IS-IS neighbor
Code =3	ES neighbor
Code=6	MAC address of an IS-S neighbor
Code=8	Padded field
Code=9	LSP entity
Code=10	Verification field

TLV Code	Description
Code=14	Size of the source LSP buffer
Code=22	Extended IS reachability
Code=128	IP internal reachability information
Code=129	Supported protocol
Code =130	IP external reachability information
Code=131	Inter-domain routing protocol information
Code=132	IP interface address
Code=133	Verification information
Code=135	Extended IP reachability TLV
Code=137	Dynamic host name
Code = 211	Graceful Restart
Code = 222	MT IS TLV
Code = 229	Multi Topology TLV
Code=232	IPv6 interface
Code = 235	IPv4 MT IP reachability TLV
Code =236	IPv6 IP reachability TLV
Code = 237	IPv6 MT IP reachability TLV
Code =240	P2P three-way handshake TLV

- IS-IS supported network types

  IS-IS supported network types include P2P links and broadcast links.

2. Working Process of IS-IS

IS-IS working process consists of neighbor establishment, LSDB synchronization, and route computation. In a broadcast network, neighbor establishment involves DIS election.

- Neighbor establishment

  The IS-IS neighbor establishment procedure is similar to the OSPF neighbor establishment procedure. If the received hello packets do not contain any information, the neighbor state is reset to Initial. If the neighbor field in the received packets indicates the local MAC address, the neighbor state is Up. If the neighbor state of the local router and peer routers is Up, the neighbor relationship is successfully established.

- DIS election

  A DIS in a broadcast network works like a DR in OSPF.

A pseudonode is generated by a DIS and sets up a relationship with each device in the local network.

Like DR election in OSPF, a DIS is elected to reduce unnecessary neighbor relationships and route information exchanges. After neighbor establishment, the DIS election starts at respective levels of areas. The election priority also changes with the levels of areas.

A DIS simulates multiple access links as a pseudonode and generates LSPs for the pseudonode. The pseudonode sets up a relationship with each device in the local network and forbids direct communication between the devices. A broadcast subnet and a non-broadcast multiple access (NBMA) network are considered as pseudonodes externally. Non-DIS devices report local link states to the DIS in the same network, and the DIS reports the link states on behalf of all the IS devices in the network.

DIS election in IS-IS is preemptive. The election result can be manually controlled through interface priority configuration. The device with a higher interface priority is more likely to be elected as the DIS. A device with the priority 0 participates in DIS election as well.

- LSDB synchronization

  All devices running link state routing protocols send local link state and directly-connected network link state to neighbors. Upon receiving the link information, the neighbors record and forward the information to their neighbors. After link state update is completed, all routers obtain topological information of the entire network. This is referred to as LSDB flooding synchronization.

  Many neighbor actions such as neighbor or neighbor interface up/down, metric change, redistributed route change and periodic change, trigger LSDB synchronization.

  A router processes a received LSP as follows:

  a The router searches for this LSP in the LSDB. If the LSP does not exist in the LSDB, the router records this packet in the LSDB and broadcasts it for synchronization.

  b If the sequence number of this LSP is greater than that of the local LSP, the router updates the LSDB and broadcasts it for synchronization.

  c If the sequence number of this LSP is smaller than that of the local DPS, the router replies to the local LSP.

- Route computation

  After the LSDB is synchronized, IS-IS computes the SPF and generates IS-IS routes based on the SPF.

  In the LSP sent by the Level-1/Level-2 routers, ATT is set to 1. A Level-1 router gets to know the closest Level-1/Level-2 router based on the ATT and route computation result, and considers the Level-1/Level-2 router a routing transit to other areas.

3. IS-IS Route Control

- Route redistribution

  IS-IS can introduce routes of other routing protocols and routes of other IS-IS processes to the IS-IS process and mutually introduce routes between Level-1 and Level-2 areas.

  In IS-IS, a Level-2 area does not synchronize the known routing information of the backbone areas and other Level-1 areas to a Level-1 area. The Level-2 area cannot obtain external routing information outside Level-1 area. During routing, secondary routes may be generated. During route redistribution, routes of other Level-1 areas and backbone areas that meet conditions can be introduced to the local Level-1 area based on the routing policy.

- Route summarization

  Route summarization is a process of summarizing routing information with the same prefix into one route, and advertising the summarized route (replacing a large number of individual routes) to neighbors. Route summarization helps reduce the protocol interaction load, the size of the routing table, and the impact of route flapping on the routing table.

- Load balancing

  If multiple equal-cost routes exist in a network, IS-IS can distribute load to the equal-cost routes to improve link usage.

4. IS-IS Wide Metric

The default metric type of IS-IS is narrow. In narrow mode, the maximum metric value of an interface is **63**. In a large network, the metric value range in narrow mode is not wide enough. Therefore, wide metric is developed. The metric value can reach 16777215.

5. LSP Fragment Extension

IS-IS advertises link state information by flooding LSPs. The size of an LSP is limited by the MTU size of the link. When the content to be advertised exceeds one LSP, IS-IS will create LSP fragments to carry new link state information. Based on ISO, an LSP fragment is indicated by an LSP number that is one byte long. Therefore, an IS-IS device can generate a maximum of 256 LSP fragments.

The 256 LSP fragments are insufficient in any of the following situations:

(1) New applications (such as traffic engineering) extend new TLV or sub-TLV.

(2) The network is expanded continuously.

(3) Routes with reduced granularity are advertised, or other routes are redistributed to IS-IS.

After LSP fragments are used up, new routing information and neighbor information will be discarded, causing network exceptions such as routing black holes or loops. To ensure normal network operation, LSP fragments must be extended to carry more link state information.

By configuration of an additional system ID and an fragment extension switch, IS-IS can advertise more link state information in extended LSP fragments. Each virtual system can be considered as a virtual device that establishes a neighbor relationship (with the path value 0) with the originating system. Extended LSPs are advertised by the neighbor of the originating system, that is, the virtual system.

The following terms are related to fragment extension:

- Normal system ID

  The system ID defined by ISO is used to establish neighbor relationships and learn routes. It is further defined as the normal system ID in order to be distinguished from the additional system ID introduced to fragment extension.

- Additional system ID

  The additional system ID is configured by an administrator to generate extended LSPs. The additional system ID shares the usage rules of the normal system ID (for example, the additional system ID must be unique in the entire area), except that the additional system ID is not carried in Hello packets for neighbor relationship establishment.

- Originating system

An originating system is an IS-IS-enabled device and maps to a virtual system identified by the additional system ID.

- Virtual system

  A virtual system is identified by the additional system ID and used to generate extended LSPs. The concept of virtual system is proposed by RFC for distinguishing from the concept of originating system. Each virtual system can generate up to 256 LSP fragments. The administrator can configure multiple additional system IDs (virtual systems) to generate more LSP fragments.

- Original LSP

  An original LSP is the LSP whose system ID contained in the LSP ID is a normal system ID. Original LSPs are generated by an originating system.

- Extended LSP

  An extended LSP is the LSP whose system ID contained in the LSP ID is an additional system ID. Extended LSPs are generated by a virtual system.

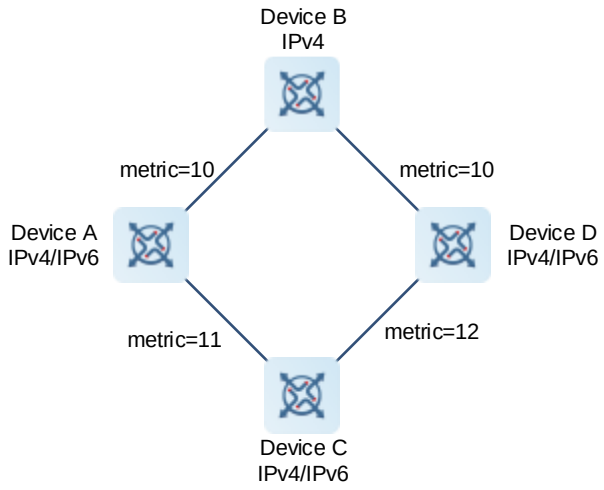
6. IS-IS MTR

IS-IS multi-topology routing (MTR) is an extended feature used to separate IPv4 unicast route computation from IPv6 unicast route computation based on topologies. New TLV types introduced to IIH packets and LSP packets of IS-IS to transmit IPv6 unicast topology information. One physical network can be divided into an IPv4 unicast logical topology and an IPv6 unicast logical topology. The two topologies perform SPF computation separately and maintain respective independent IPv4 and IPv6 unicast routing tables. In this way, IPv4 unicast service traffic and IPv6 unicast service traffic are forwarded by different paths. The IS-IS MTR technique helps users deploy IPv6 unicast networks without the constraint on consistency between IPv4 and IPv6 unicast topology information.

IS-IS MTR is derived from IS-IS MT, which is used to separate IPv4 unicast topology from IPv6 unicast topology, unicast topology from multicast topology, and topologies using different protocol stacks (such as IPv4 and Pv6). IS-IS MTR separates IPv4 unicast topology from IPv6 unicast topology based on IS-IS MT.

Figure 1-3 shows a typical networking application with the following implementation requirements: Deploy an IPv6 unicast topology in incremental mode and upgrade some devices to support IPv4 and IPv6 dual protocol stacks while keeping other IPv4-enabled devices unchanged.

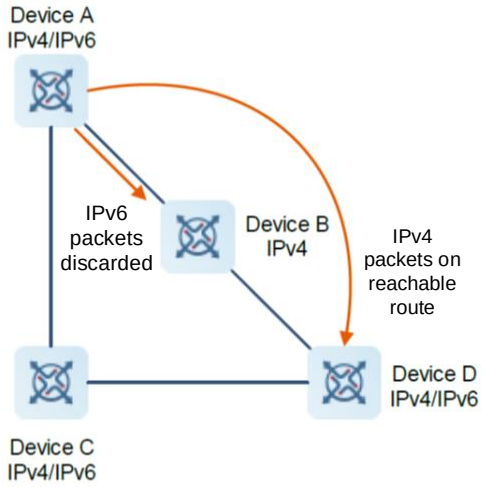
Figure 1-3 Physical Topology for IPv4/IPv6 Hybrid Deployment



As shown in Figure 1-3, each link is marked by a number indicating its metric. Device B only supports the IPv4 protocol stack, whereas other devices support IPv4 and IPv6 dual protocol stacks.

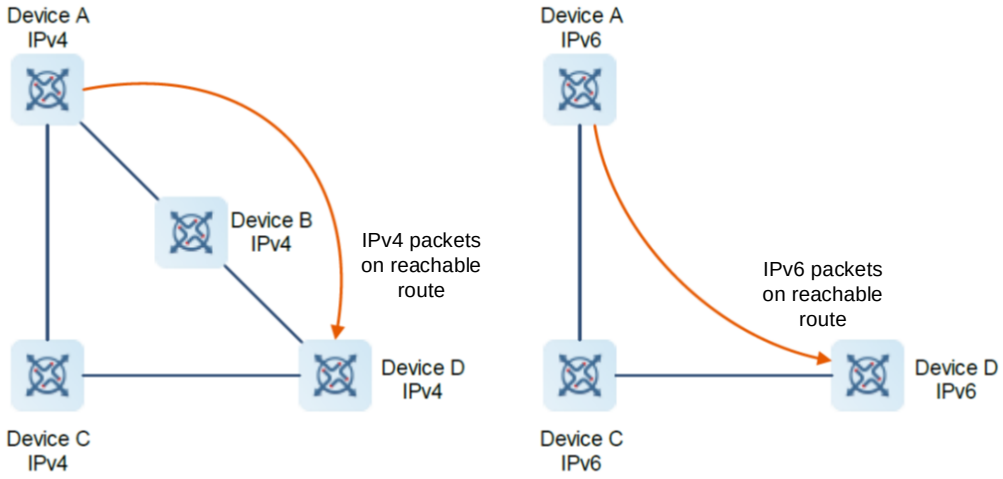
The networking constraint on consistency between IPv4 and IPv6 unicast topologies must be removed to retain the use of Device B (supporting IPv4) so that Device B can establish a neighbor relationship with Device A or D. This, however, may cause new problems.

Figure 1-4 IPv4/IPv6 Hybrid Topology



As shown in Figure 1-4, without IS-IS MTR, SPF computation performed by Devices A, B, C, and D only take into account the single hybrid topology. The calculated shortest path is Device A -> Device B -> Device D, with the cost 20. Device B will discard IPv6 packets because it does not support IPv6.

Figure 1-5 Separation of IPv4 and IPv6 Topologies



As shown in Figure 1-5, after the IS-IS MTR technique is used to separate IPv4 unicast topology from IPv6 unicast topology, Devices A, B, C, and D establish neighbor relationships based on the IPv4 and IPv6 unicast topologies. The left part shows the IPv4 topology formed by IPv4-enabled routers. The calculated IPv4 shortest

path is Device A -> Device B -> Device C, which realizes IPv4 packet forwarding. The right part shows the IPv6 topology formed by IPv6-enabled routers. The calculated IPv6 shortest path is Device A -> Device C -> Device D, which realizes IPv6 packet forwarding.

General IS-IS MTR deployment: IS-IS MTR must be deployed to avoid routing black holes when some devices support only one protocol. IS-IS MTR is not required when all devices support IPv4 and IPv6 dual protocol stacks.

Deployment of a new network: IS-IS MTR is not required when devices only support the IPv4 protocol stack. For devices that only support the IPv6 protocol stack or devices that support IPv4 and IPv6 dual protocol stacks, enable the MT mode of IS-IS MTR. You are not advised to enable Multi-Topology Transition (MTT) because loops may occur.

Reconstruction of an existing network with devices supporting only one protocol stack: Enable the MTT mode of IS-IS MTR on devices that support IPv4 and IPv6 dual protocol stacks in sequence (starting from the device closest to a device supporting only one protocol stack in the network topology). After the MTT mode is enabled on all new devices, switch the MTT mode to the MT mode on these devices in sequence (starting from the device farthest from a device supporting only one protocol stack in the network topology).

7. IS-IS Neighbor

In a broadcast network, neighbors elect a DIS and routers send routing information to the DIS.

On a P2P link, neighbor relationship can be established in two ways:

- Two-way handshake

  As long as an IIH packet is received, neighbor relationship is established. This method comes with low reliability.

- Three-way handshake

  Neighbor relationship can be established after the IIH packet is sent three times. This method boasts high reliability.

  In a link with one down end, if a neighbor relationship is established based on the IIH packet received from the peer device, the established neighbor relationship is abnormal. In a three-way handshake, an acknowledgment packet must be received from the peer device to confirm that the peer device has received the IIH packet before the neighbor relationship is established.

The following conditions must be met for two routing devices to establish a neighbor relationship when IS-IS MTR is not configured:

- The interface addresses on both routing devices are in the same network segment.
- The interface levels on both routing devices match.
- The routing devices are authenticated by each other.
- The routing devices support the same protocol.

The following conditions must be met for two routing devices to establish a neighbor relationship when IS-IS MTR is configured:

- The interface addresses on both routing devices are in the same network segment.
- The interface levels on both routing devices match.
- The routing devices are authenticated by each other.
- The routing devices have at least one consistent MT ID when P2P links are configured.

- There are no constraints on the MT IDs that the routing devices support when LAN links are configured.

8. IS-IS Authentication

Internet requirements for information security increase. To avoid data theft and tampering, operators apply the IS-IS authentication function on specified areas or interfaces to ensure data security.

- IS-IS authentication types
  - Classification based on packet types:
  - Interface authentication: Authenticate Level-1 and Level-2 hello packets.
  - Area authentication: Authenticate Level-1 CSNP, PSNP, and LSP packets.
  - Routing domain authentication: Authenticate Level-2 CSNP, PSNP, and LSP packets.
  - Classification based on authentication mode:
  - Simple authentication: Add a configured password to a packet. This authentication mode comes with low security.
  - MD5 authentication: A password is encrypted using MD5 algorithms and then added to a packet.
  - Keychain authentication: A keychain is configured so that the password changes with time.
- Implementation procedure

  Upon receiving an IS-IS packet from another router, the local router checks whether the authentication password is correct. If not, the packet is discarded. The IS-IS packet carries authentication information by means of TLV. The TLV format is as follows:
  - Type: Indicates an authentication packet type. For TCP/IP packets, this value is 133.
  - Length: Indicates the length of a TLV.
  - Value: Indicates an authentication type and password.

  Interface authentication: Indicates that authentication is configured in interface configuration mode. Interconnected router interfaces must be configured with the same password.

  Area authentication: Indicates that authentication is configured in IS-IS routing configuration mode. If the authentication is performed between routers in the same area, the key string or password must be the same.

  Routing domain authentication: Indicates that authentication is configured in IS-IS routing configuration mode. If the authentication is performed between routers in the same routing domain, the key string or password must be the same.

  SNP packets can be authenticated separately from LSP packets and authentication check can be enabled for the received LSP or SNP packets.

9. IS-IS GR

For GR to be successful, the following two conditions must be met:

- The network topology is stable.
- The device can ensure uninterrupted forwarding when it restarts IS-IS.

Two roles exist during the GR process: Restarter and Helper. Accordingly, the IS-IS GR function is divided into IS-IS GR Restart and IS-IS GR Help capabilities. A device with the IS-IS GR Restart capability can send GR requests and automatically execute GR. A device with the IS-IS GR Help capability can receive GR requests

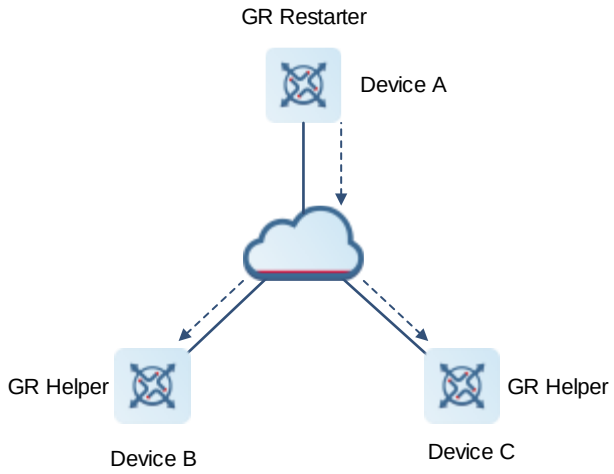
and assist neighbors in executing GR. The GR process starts when the Restarter sends a GR request. After receiving the GR request, the neighboring device enters Help mode to help the Restarter reestablish its LSDB while maintaining the neighbor relationship with the Restarter.

When an IS-IS device performs GR, it notifies its neighbors to maintain their neighbor relationship so that other devices in the network cannot sense the change in the topological relationship and the neighbors will not recalculate the route and update its forwarding table. The IS-IS device synchronizes and restores the LSDB to its pre-GR state with the help of the neighbors to ensure that the routing table and forwarding table remain unchanged before and after GR implementation and data forwarding is not interrupted.

The Restarter performs the following operations during the GR process:

(1) The GR Restarter notifies the GR Helpers that the local device will be restarted.

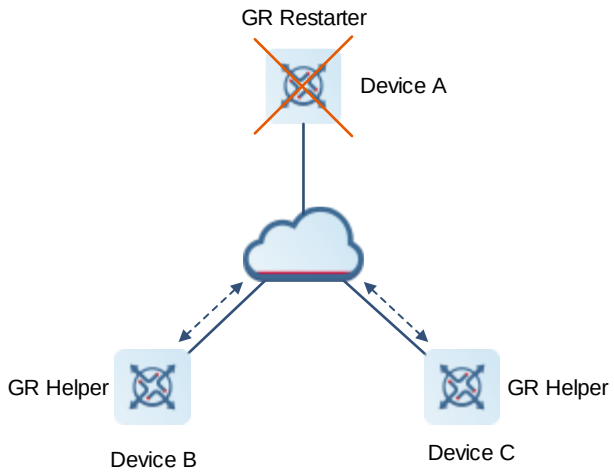
Figure 1-6 Restarter Notifying Helpers of GR



Device A acts as the GR Restarter, and Devices B and C are GR Helpers of Device A. Device A sends a GR request instructing all its neighbors not to delete the neighbor relationships with Device A when it is restarted. After receiving the GR request, the neighbors send GR responses to the GR Restarter, and maintain their neighbor relationships with the GR Restarter during the GR time (specified by GR grace-period) notified by the GR Restarter.

(2) The GR Restarter is restarted.

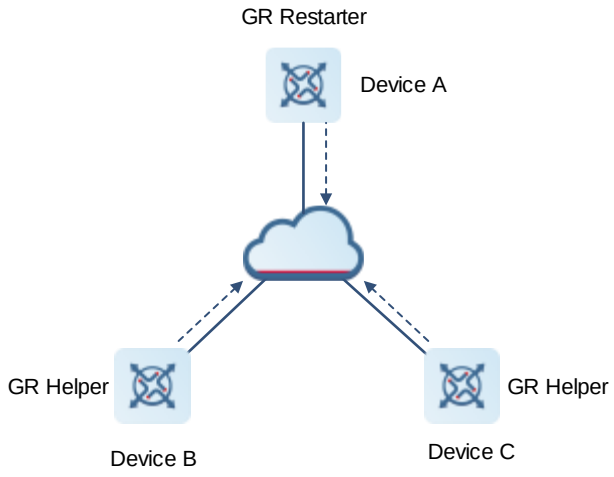
Figure 1-7 Restarter Restart



When the GR Restarter is restarted, its IS-IS interface changes from down state to up state. The GR Helpers know that the GR Restarter is in IS-IS restart state, they maintain their neighbor relationships with the GR Restarter during the GR time and retain the routes from the GR Restarter.

(3) The GR Restarter synchronizes topology and routing information from the GR Helpers.

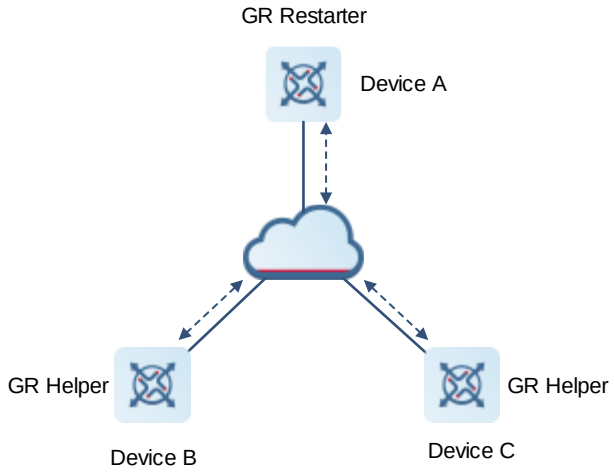
Figure 1-8 LSDB Synchronization



After IS-IS restart, the GR Restarter synchronizes topology or routing information from the GR Helpers and recalculates its routing table. During the process, any change in the routing table is not updated to the forwarding table.

(4) GR is completed when the GR Restarter finishes LSDB synchronization. Then all devices enter IS-IS interaction state.

Figure 1-9 GR Completion



After the GR Restarter synchronizes all required data, all devices enter IS-IS interaction state. The routing table of the GR Restarter is updated to the forwarding table and invalid entries are cleared. Because the GR Restarter is completely restored to the pre-restart state under stable network conditions, its routing table and forwarding table remain unchanged before and after GR.

10. Fast Convergence of IS-IS

IS-IS accelerates convergence by fast advertisement of LSPs and smart timer.

- Fast advertisement of LSPs

  By modification of LSP flooding interval and maximum number of LSPs sent on an interface, LSPs can be advertised to the IS-IS area in the route computation time to reduce route computation and accelerate convergence when the LSPs trigger route computation.

- Smart timer

  LSP packet generation time and SPF algorithm computation time are adjusted based on the smart timer. In a stable IS-IS network, the route computation frequency is low and the first SPF algorithm time is short. In a network with ever-changing topology and network jitter, route computation frequency increases, which consumes CPU resources. In this case, the smart timer will increase SPF computation delay to accelerate convergence and reduce CPU consumption.

1.1.3 Protocols and Standards

- RFC 1142: OSI IS-IS Intra-domain Routing Protocol
- RFC 1195: Use of OSI IS-IS for routing in TCP/IP and dual environments
- RFC 3786: Extending the Number of Intermediate System to Intermediate System (IS-IS) Link State PDU (LSP) Fragments Beyond the 256 Limit
- RFC 3373: Three-Way Handshake for Intermediate System to Intermediate System (IS-IS) Point-to-Point Adjacencies
- RFC 3358: Optional Checksums in Intermediate System to Intermediate System (ISIS)
- RFC 3784: Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)
- RFC 2763: Dynamic Hostname Exchange Mechanism for IS-IS

- RFC 6119(draft-ietf-isis-ipv6-te-00): IPv6 Traffic Engineering in IS-IS
- RFC 2966: Domain-wide Prefix Distribution with Two-Level IS-IS
- RFC 5306: Restart Signaling for IS-IS

1.2 Restrictions and Guidelines

The PIM mode enabled on all interfaces of a device for an instance must be consistent.

1.3 Configuration Task Summary

IS-IS configuration includes the following tasks:

(1) Configuring Basic Settings of IS-ISConfiguring Basic Settings of IS-IS

(2) (Optional) Configuring IS-IS network Levels

(3) (Optional) Configuring IS-IS Hello Packets

(4) (Optional) Configuring IS-IS LSPs

(5) (Optional) Configuring IS-IS SNPs

(6) (Optional) Configuring IS-IS Route Advertisement

(7) (Optional) Configuring IS-IS Route Summarization

(8) (Optional) Configuring an IS-IS Passive Interface

(9) (Optional) Controlling IS-IS Routing Factors

(10) (Optional) Configuring the Priority of an IS-IS Device

(11) (Optional) Configuring IS-IS Authentication

(12) (Optional) Configuring IS-IS GR

(13) (Optional) Configuring IS-IS NSR

(14) (Optional) Configuring IS-IS Fast Reroute

1.3.2 Overview

When a link fault occurs, it takes about 10 to 30 seconds for IS-IS to detect the neighbor disconnection, advertise the link fault, recalculate the Shortest Path Tree (SPT), and update the routing table. During this period, there is a traffic interruption due to the absence of valid routes.

To minimize the traffic interruption caused by network failures, administrators can configure IS-IS Fast Reroute (FRR). When FRR is enabled, IS-IS precomputes primary and backup paths for each route during normal route calculation. These paths are updated to the routing table. In the event of a link fault, the device quickly detects the fault on the primary path and switches the traffic to the backup path, eliminating the need for IS-IS to recalculate the routes. This significantly reduces the traffic interruption time. When used with BFD for path detection, IS-IS FRR allows the link interruption time to be reduced from over 10 seconds to about 50 milliseconds. When the device forwards packets over the backup path, IS-IS recalculates the routes based on the changed network topology and updates the routing table. Subsequently, the device uses the newly calculated optimal routes for packet forwarding.

Administrators can use IS-IS automatic calculation of backup paths or directly specify backup paths for eligible routes based on comprehensive network planning.

1.3.3 Restrictions and Guidelines

- IS-IS basic functions must be configured.
- When FRR is enabled for IS-IS, you are advised to enable BFD at the same time so that the device can quickly detect link faults, thereby shortening the forwarding interruption time. If the interface is up or down, to shorten the forwarding interruption time during IS-IS fast reroute, you can configure carrier-delay 0 in interface configuration mode to achieve the fastest switchover speed.
- If an interface has limited remaining bandwidth or may fail simultaneously with the primary interface, it cannot be a backup interface. In this case, the command needs to be enabled in interface configuration mode to exclude the interface from being used as a backup interface. After this command is run, a backup interface will be selected from other interfaces during IS-IS FRR calculation.

1.3.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable FRR for IS-IS.

frr { **lfa** [**downstream-paths**] **| route-map** *route-map-name* } { **level-1 | level-2** }

By default, FRR is disabled for IS-IS.

(5) Return to the global configuration mode.

exit

(6) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the loopback interface configuration mode.

interface loopback *interface-number*

(7) (Optional) Configure LFA protection mode for an IS-IS interface.

isis frr protection { node | link-node | disable }

By default, LFA link protection is enabled.

This command does not take effect if the **fast-rerotue route-map** command is configured.

(8) (Optional) Configure the function of excluding an interface from being a backup.

isis frr no-eligible-backup

By default, IS-IS is enabled and neighboring interfaces can be used as backup interfaces.

This command does not take effect if **frr route-map** command is configured.

(9) Correlating IS-IS with BFD

(10) (Optional) Configuring the IS-IS Overload Bit

(11) (Optional) Configuring IS-IS VRF

(12) (Optional) Configuring IS-IS MTR

(13) (Optional) Configuring IS-IS SNMP

(14) (Optional) Enabling Super VLAN for IS-IS

(15) (Optional) Configuring IS-IS Two-way Maintenance

(16) (Optional) Configuring IS-IS BGP-LS

(17) (Optional) The following configuration tasks are optional. Select tasks for configuration according to the actual condition.

- Configuring the Maximum Metric of a Neighbor
- Configuring the Maximum Number of Area Addresses
- Configuring SPF Calculation Interval
- Canceling Three-Way Handshake of a P2P Network
- Enabling TLV to Be Padded Based on Supported Protocol
- Configuring the Reachable Time of a Suppressed Neighbor
- Suppressing Advertisement of Directly Connected Routes on an Interface

2. Overview

By default, directly connected routes of an interface are advertised in LSPs for route calculation. If you want to exclude the directly connected routes from route calculation on a specific interface, you can configure this function. After this command is configured, the directly connected routes of a specified protocol will not be advertised on an IS-IS interface and the establishment of IS-IS neighbors is still allowed. This function can suppress the interface prefixes carried in LSPs to keep them invisible to external entities, thereby enhancing security and accelerating route convergence.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable the function of suppressing the advertisement of directly connected routes on an IS-IS interface.

isis suppress-reachability

By default, the advertisement of directly connected routes is not suppressed on an IS-IS interface.

- Enabling Adjacency Event Output

1.4 Configuring Basic Settings of IS-IS

1.4.1 Overview

Basic settings of IS-IS are a precondition for configuration of all IS-IS features and basic internetworking.

1.4.2 Restrictions and Guidelines

- The Level-1 IS-IS devices in an area must be configured with the same area ID.
- The core routing table does not distinguish the routing entries generated by different IS-IS routing processes.
- The IP addresses of interfaces connected between neighbors must be in the same network segment.
- If the IP addresses are in different network segments, a neighbor relationship cannot be established.
- If the parameter *tag* is specified in the creation of IS-IS routing process, the name of the IS-IS instance is the tag name of the process. If the parameter *tag* is not specified, the instance name of the process is **Null**.
- If you need to add an interface to the specified IS-IS routing process and configure the interface with the IPv4 IS-IS function, add the tag parameter after the **ip router-protocol isis** command to indicate the process name.
- If you run the **unset ip routing** command in global configuration mode, IS-IS will disable IPv4 routing on all interfaces. That is, the **unset ip router-protocol isis** *[tag]* command is automatically executed on all interfaces, with other IS-IS settings unchanged.
- If you need to add an interface to the specified IS-IS routing process and configure the interface with the IPv6 IS-IS function, add the tag parameter after the **ipv6 router-protocol isis** command to indicate the process name.
- If you run the **unset ipv6 unicast-routing** command in global configuration mode, IS-IS will disable IPv6 routing on all interfaces. That is, the **unset IPv6 router-protocol isis** *[tag]* command is automatically executed on all interfaces, with other IS-IS settings unchanged.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IS-IS routing process and enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

No IS-IS instance is configured by default.

If you configure the *tag* parameter when you start an IS-IS routing process, you need to add this tag parameter when you stop the IS-IS routing process.

(4) Configure a NET address in IS-IS.

net *net-address*

No NET address is configured in IS-IS by default.

Different NET addresses must have the same system ID.

(5) (Optional) Modify the system ID in IS-IS. Configure one of the following tasks.

- Replace the system ID of an instance with the configured name.

  is-name *name*

  The default system ID is used.

- Replace the system ID of a router with the host name of the destination router.

  hostname dynamic

  The host name replacement function is enabled by default.

(6) Return to the global configuration mode.

exit

(7) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(8) Enable IS-IS on an interface.

- Enable IPv4 IS-IS on an interface.

  ip router-protocol isis [*tag*]

  IPv4 IS-IS is not enabled on an interface by default.

- Enable IPv6 IS-IS on an interface.

ipv6 **router-protocol isis** [*tag*]

IPv6 IS-IS is not enabled on an interface by default.

1.5 Configuring IS-IS network Levels

1.5.1 Overview

IS-IS supports a two-Level system to realize routing management and extensible route selection in a large network. Each level is only concerned about maintaining the topology of the corresponding area.

1.5.2 Restrictions and Guidelines

- If the circuit-type is set to Level-1 or Level-2-only, IS-IS will only send PDUs at the corresponding level.
- If the circuit type of an interface is set to **external**, this interface will work as an external domain interface and IS-IS will not send PDUs at the corresponding level.
- A device can have only one instance running at Level-2 (including Level-1/Level-2).

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure a level of IS-IS.

is-type { **level-1 | level-1-2 | level-2-only** }

IS-IS runs at Level-1/Level-2 by default.

(5) Return to the global configuration mode.

exit

(6) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.

- **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(7) Configure an IS-IS level on an interface.

isis circuit-level { **level-1 | level-1-2 | level-2-only** [**external**] }

The circuit type of an interface is set to Level-1/Level-2 by default.

(8) Configure a broadcast interface as a P2P interface.

isis network-type point-to-point

The broadcast network type is unchanged by default.

1.6 Configuring IS-IS Hello Packets

1.6.1 Overview

Hello packet control parameters, including hello packet sending interval, hello hold time, packet padding, and neighbor detection are configured. Neighbor establishment can be controlled to maintain neighbor interaction.

1.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Configure the Hello packet sending interval of an interface.

isis hello-period { *interval* | **minimal** } [**level-1 | level-2**]

The Level-1 and Level-2 hello packets are sent at an interval of **10** seconds by default.

(5) Configure the Hello packet hold time multiplier on an interface.

isis iih-multiplier *multiplier-number* [**level-1** | **level-2**]

The default hello hold time multiplier is **3**. That is, the hello hold time is three times of the hello interval.

(6) Configure to pad IS-IS hello packets on an interface.

isis hello-packet padding

Padding is enabled by default for hello packets sent on an interface.

(7) Enable neighbor supported protocol detection for hello packets in IPv4 environment.

adjacency-support-check

The neighbor supported protocol detection function for hello packets is enabled by default.

(8) Enter the IS-IS IPv6 address family configuration mode.

address-family-mode ipv6 unicast

(9) Enable neighbor supported protocol detection for hello packets in IPv6 environment.

adjacency-support-check

The neighbor supported protocol detection function for hello packets is enabled by default.

(10) Exit the IS-IS IPv6 address family configuration mode.

exit-address-family-mode

1.7 Configuring IS-IS LSPs

1.7.1 Overview

Adjusting timer and packet length of LSPs will affect LSP update timeliness and LSP frequency.

1.7.2 Restrictions and Guidelines

- The LSP generation time and refresh time can be configured to prevent network topology change from affecting the LSP transmission frequency.
- During LSDB change, IS-IS notifies the change by LSP advertisement. If a great number of IS-IS routes and interfaces exist on a device, a too small sending interval and a too great LSP number may cause LSP bursts. Therefore, the LSP sending interval and LSP number must be set to proper values. LSP retransmission interval is used in P2P network. In such a network, an LSP response must be returned to indicate completion of an LSP retransmission.
- The value of *lsp-refresh-interval* must be smaller than that of *max-lsp-lifetime*.
- The value of *max-lsp-lifetime* must be greater than that of *lsp-refresh-interval*.
- The value of the configured *initial-interval* cannot be greater than that of *maximum-interval*. Otherwise, the value of *initial-interval* is changed to that of *maximum-interval*.
- The value of the configured *hold-interval* cannot be greater than that of *maximum-interval*. Otherwise, the value of *hold-interval* is changed to that of *maximum-interval*.
- The value of the configured *initial-interval* cannot be greater than that of *hold-interval*. Otherwise, the value of *initial-interval* is changed to that of *hold-interval*.

- If no particular level is specified in the command, the default level is Level-1/Level-2, which means that the interval configuration takes effect to Level-1 and Level-2 concurrently.
- When you need to configure **mesh-group** on an IS-IS interface, run the **isis csnp-period** command to configure the non-0 CSNP interval to ensure complete LSP synchronization between neighbors in the network. After that, CNSPs will be periodically sent to synchronize LSPs.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Modify LSP control parameters on the interface. The following configuration tasks are optional. Select tasks for configuration according to actual condition.

- Configure the LSP sending interval.
  - **isis lsp-transmit-period** *interval* [**level-1** | **level-2**]

  LSPs are sent at an interval of **33** ms by default on an IS-IS interface.
- Configure the maximum number of LSPs sent by an IS-IS interface at a time.

  isis lsp-flood *lsp-number* [**level-1** | **level-2**]

  An IS-IS interface sends at most 5 LSPs at a time by default.
- Configure the LSP retransmission interval.

  isis retransmit-interval *retransmit-interval* [**level**-1 | **level**-2]

  LSPs are retransmitted at an interval of **5** seconds by default.

  If *retransmit-interval* is set to **0**, no LSP is retransmitted.

(5) Return to the global configuration mode.

exit

(6) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(7) Modify LSP control parameters in the IS-IS process. The following configuration tasks are optional. Select tasks for configuration according to actual condition.

- Configure the LSP generation interval.

  lsp-gen-period [**level-1** | **level-2**] *maximum-interval* [*initial-interval hold-interval*]

  The default maximum interval between two LSP generation times is **5** seconds, the delay for generating an LSP is **50** ms, and the minimum interval between the first LSP generation time and the second LSP generation time is **200** ms.

- Configure the LSP refresh interval.

  lsp-refresh-period *interval*

  The default LSP refresh interval is **900** seconds.

- Configure the LSP lifetime.

  lsp-max-life *max-lifetime*

  The default maximum LSP lifetime is **1200** seconds.

- Configure the interval for receiving duplicate LSP packets.

  min-lsp-arrival [**level-1** | **level-2**] *minimum-interval initial-interval hold-interval*

  The function of receiving duplicate LSP packets is not supported by default.

- Configure to ignore LSP checksum errors.

  lsp-error-ignore

  LSP checksum errors are processed by default.

- Configure LSP fragment extension.

  lsp-fragments-extend [**level-1** | **level-2**] [**reference rfc3786**]

  The fragment extension function is disabled by default. To enable fragment extension, run this command to configure fragment extension at the specified level and run the **virtual-system** command to configure an additional system ID.

- Configure the additional system ID of the IS-IS routing process.

  virtual-system *system-id*

  No additional system ID is configured by default. To enable fragment extension, run this command to configure an additional system ID and run the **lsp-fragments-extend** command to configure fragment extension.

- Configure the maximum length of received LSPs.

  lsp-length receive *size*

  The default maximum length of received LSPs is **1492** bytes.

- Configure the maximum length of sent LSPs.

  lsp-length originate *size* [**level-1** | **level-2**]

The default maximum length of sent LSPs is **1492** bytes.

- Add an IS-IS interface to a specified mesh group.

  isis fully-connected-group { **blocked** | *mesh-group-id* }

  No interface joins any mesh group by default.

1.8 Configuring IS-IS SNPs

1.8.1 Overview

By default, a DIS in a broadcast network sends a CSNP every 10s for LSDB synchronization. In a P2P network, CSNPs are sent only after a neighbor relationship is established.

1.8.2 Restrictions and Guidelines

- If an interface is added to a mesh group, CSNP sending interval can be configured. If the CSNP interval is set to **0**, no CSNP is sent.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Configure the CSNP broadcast interval on an IS-IS interface.

isis csnp-period *interval* [**level-1** | **level-2**]

CSNPs are sent at an interval of **10** seconds in a broadcast network by default. No CSNPs are sent in a P2P network by default.

When you configure a new CSNP interval without specifying Level-1 or Level-2, the interval configuration takes effect for Level-1 and Level-2 CSNPs by default.

(5) Configure the minimum PSNP sending interval.

isis psnp-period *psnp-interval* [**level-1** | **level-2**]

The function of configuring the PSNP sending interval is disabled by default. In this case, the minimum PSNP sending interval is **2** seconds by default and it takes effect for Level-1 and Level-2 PSNPs.

1.9 Configuring IS-IS Route Advertisement

1.9.1 Overview

IS-IS can introduce routes of other routing protocols and routes of other IS-IS processes to the IS-IS process and mutually introduce routes between Level-1 and Level-2 areas.

1.9.2 Restrictions and Guidelines

- Because Level-2 domains do not generate any default route, the **default-info originate** command is run to allow a default route to enter a Level-2 domain.
- In the old versions of some vendors, if **metric-type** is set to **external-route**, the metric of redistributed routes is added by 64 during route calculation and used to determine routing. This practice does not comply with the related protocol. In the actual application, external routes may be preferred over internal routes. If this happens during interworking with old versions of some vendors, you can modify the related setting (such as **metric-value** or **metric-type**) of each device to ensure that internal routes are preferred over external routes.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Generate a default route.

default-info originate [**route-map** *route-map-name*]

No default route is generated by default.

(5) Redistribute other routes to IS-IS.

redistribute-route { **arp-host | bgp** | **connected** | **nd-route** | **ospf** *process-id* [**filter** { **external-type**[**1** | **2**] | **internal-type** | **nssa-external-type** [**1** | **2**] }] | **rip** | **static** } [[**level-1** | **level-1-2** | **level-2**] | **metric-value** *metric-value* | **metric**-**class** { **external-type** | **internal-type** } | **route**-**map** *route-map-name*]

Redistribution is disabled by default.

(6) Redistribute the Level-1 reachable routing information of the specified IS-IS instance to Level-2 of the current instance.

redistribute-route isis [*tag*] **level-1 to level-2** [**route-filtering** *acl-name* | **route-map** *route-map-name*]

Redistribution is disabled by default.

(7) Redistribute the Level-2 reachable routing information of the specified IS-IS instance to Level-1 of the current instance.

redistribute-route isis [*tag*] **level-2 to level-1** [**route-filtering** *acl-name* | { **ipv6-prefix** *ipv6-address/length* | **prefix** *ipv4-address net-mask* } | **route-map** *route-map-name*]

Redistribution is disabled by default.

1.10 Configuring IS-IS Route Summarization

1.10.1 Overview

Configuring route summarization can effectively reduce the size of a routing table and system resource usage. If the links in the range of the IP addresses to be summarized jitter (up and down), advertisement of the summarized routes is not affected to avoid route flapping. After route summarization, if any routing information about a reachable address or network segment exists, a summarized route, instead of a detailed route, is advertised externally.

1.10.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure an IPv4 summarized route.

summary-address *ipv4-address net-mask* [**level-1** | **level-1-2** | **level-2**] [**metric-value** *metric-value*]

The route summarization function is disabled by default.

(5) Enter the IS-IS IPv6 address family configuration mode.

address-family-mode ipv6 unicast

(6) Enable IPv6 route summarization.

summary-prefix *ipv6-prefix/prefix-length* [**level-1** | **level-1-2** | **level-2**]

The route summarization function is disabled by default.

1.11 Configuring an IS-IS Passive Interface

1.11.1 Overview

This function prevents the specified interface from receiving and sending IS-IS packets, but the IP address of this interface will be flooded by other interfaces.

1.11.2 Restrictions and Guidelines

- If the number of interfaces with IS-IS not enabled exceeds 255, only the first 255 interfaces will be configured as passive interfaces. The remaining interfaces are non-passive interfaces.

1.11.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure a passive interface in the IS-IS routing configuration mode.

disable-interface { **default** | *interface-type interface-number* }

The passive interface function is disabled by default.

(5) Return to the global configuration mode.

exit

(6) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(7) Configure a passive interface.

isis passive

The configured passive interface in the IS-IS routing configuration mode prevails by default.

1.12 Controlling IS-IS Routing Factors

1.12.1 Overview

IS-IS routing is determined by adjusting the number of equal-cost routes, reference bandwidth values, and interface costs.

1.12.2 Restrictions and Guidelines

- The **max-ecmp-paths** command is used by IS-IS to control the number of IS-IS equal-cost paths to be added to a routing table. There is also a routing table command used to control the number of equal-cost paths. The number of effective equal-cost paths is determined by either of the two command values, whichever is smaller.
- The **bandwidth-reference** command is used to calculate the interface metric in an instance based on the bandwidth value configured in the instance. If a metric is configured, the metric prevails.
- The metric, which is used in SPF calculation, is stored in the IP reachability information TLV. A greater metric value indicates a greater routing consumption of this interface and a longer path of SPF calculation. The Narrow-metric value is valid only when Metric-style is set to **Narrow**. The Wide-metric value is valid only when Metric-style is set to **Wide**.

1.12.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure the administrative distance of an IS-IS route.

distance *distance-value*

The default administrative distance of IS-IS is **115**.

(5) Configure the metric for an instance.

metric *metric* [**level-1** | **level-2**]

By default, Level-1 and Level-2 use the computation result of **bandwidth-reference**.

(6) Configure the maximum number of equal-cost paths.

max-ecmp-paths *maximum*

The default maximum number of equal-cost paths is **2**.

(7) Configure the reference bandwidth value for cost computation.

bandwidth-reference *value*

The default reference bandwidth value for cost computation is **100** Mbps.

(8) Configure the metric type.

metric-class { **narrow-metric** [**transition**] | **wide-metric** [**transition**] | **transition** } [**level-1** | **level-2** | **level-1-2**]

The default metric type is Narrow.

(9) Return to the global configuration mode.

exit

(10) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(11) Configure metric of the IS-IS interface.

isis metric-value *metric* [**level-1 | level-2**]

Level-1 and Level-2 use the computation result of bandwidth-reference by default.

(12) Configure the wide metric value of an interface.

isis wide-metric *metric* [**level-1 | level-2**]

The default wide metric value of Level-1 and Level-2 is the computation result of **bandwidth-reference**.

1.13 Configuring the Priority of an IS-IS Device

1.13.1 Overview

Adjusting DIS election priority of a device on an interface can manually affect DIS election result and specify a device to become the DIS.

1.13.2 Restrictions and Guidelines

- A lower DIS priority of an interface indicates a lower priority of being elected as a DIS.
- This function is invalid on a P2P network interface.

1.13.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Configure device priority in DIS election.

isis priority-value *value* [**level-1 | level-2**]

The default priority of a device for Level-1 and Level-2 DIS election is **64**.

1.14 Configuring IS-IS Authentication

1.14.1 Overview

Internet requirements for information security increase. To avoid data theft and tampering, operators apply the IS-IS authentication function on specified areas or interfaces to ensure data security.

- Interface authentication is intended for establishing and maintaining neighbor relationships. A neighbor relationship cannot be established between two IS-IS devices with different interface authentication passwords. This prevents unauthorized or unauthenticated IS-IS devices from joining an IS-IS network that requires authentication. Interface authentication passwords are encapsulated in Hello packets before being sent.
- Area authentication and routing domain authentication in IS-IS are performed to verify LSPs, CSNPs, and PSNPs to prevent unauthorized or unauthenticated routing information from being injected into the LSDB. Authentication passwords are encapsulated in LSPs, CSNPs, and PSNPs before being sent.

1.14.2 Restrictions and Guidelines

- An interface authentication password is encapsulated in a Hello packet before being sent by an interface. When an interface receives a Hello packet, it checks the password in the packet against the existing one.

- Area authentication passwords are encapsulated in Level-1 LSPs, CSNPs, and PSNPs. When an interface receives an LSP, CSNP, or PSNP, it checks the password in the packet against the existing one.
- Routing domain authentication passwords are encapsulated in Level-2 LSPs, CSNPs, and PSNPs. When an interface receives an LSP, CSNP, or PSNP, it checks the password in the packet against the existing one.
- If no particular Level is specified in the command, the default level is Level-1/Level-2, which means that the authentication configuration takes effect to Level-1 and Level-2 concurrently.
- A key chain may contain multiple passwords. A password with a smaller SN is preferentially used for sending a packet. When the packet arrives at the peer device, the device will receive the packet if the password in the packet is consistent with a password in the key chain.
- For plaintext authentication, the *string-text* in the key chain cannot exceed 80 characters; otherwise, the key chain will be invalid.
- Only one key chain can be used at a time. After you configure a new key chain, it will replace the original one.
- Before you deploy IS-IS interfaces and area authentication on all devices in the network or before you change the authentication password or authentication mode, run the **transmit-only** command. The devices will not authenticate received Hello packets to avoid network flapping when IS-IS interface authentication is deployed.

1.14.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a key chain, and enter the key chain configuration mode.

key chain *key-chain-name*

(4) Create a key chain, and enter the key chain configuration mode.

key *key-id*

(5) Configure the key.

key-string [**0** | **7**] *string-text*

(6) Exit the key chain configuration mode.

exit

(7) Return to the global configuration mode.

exit

(8) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.

- **interface** *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(9) Configure Hello packet authentication on an interface.

- Specify the authentication mode as plaintext or MD5.

  isis auth method { **md5** | **text** } [**level-1** | **level-2**]

  The interface authentication mode is disabled by default.

  The priority of the password is higher than that of the password configured using the **isis auth-password** command.

- Configure the password for interface authentication.

  isis auth keychain *key-chain-name* [**level-1** | **level-2**]

  Authentication is not performed if no key chain is configured using the **key chain** command. In addition to the key chain command, you also need to run the **isis auth method** command to make IS-IS key chain authentication take effect.

- (Optional) Configure the password for plaintext authentication of hello packets on an interface.

  isis auth-password [**0** | **7**] *password-string* [**transmit-only**] [**level-1** | **level-2**]

  Plaintext authentication is not enabled by default.

  This command does not take effect if the **isis auth method** command has been run to configure authentication before.

- (Optional) Configure an authentication password in the Hello packets sent. Received Hello packets are not authenticated.

  isis auth transmit-only [**level-1** | **level-2**]

  Packets sent and received on an interface are authenticated by default.

(10) Return to the global configuration mode.

exit

(11) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(12) Configure LSP, CSNP, and PSNP packet authentication in an IS-IS area.

- (Optional) Configure a password for plaintext authentication of a Level-1 area.

  area-auth [**0** | **7**] *password-string* [**transmit-only**]

  The plaintext authentication function of a Level-1 area is disabled by default.

All IS-IS devices in an area must be configured with the same authentication password. This command does not take effect if the **auth method** command has been run to configure authentication before.

- (Optional) Configure a password for plaintext authentication of a Level-2 routing domain.

  domain-auth [**0** | **7**] *password-string* [**transmit-only**]

  The Level-2 area authentication password function is disabled by default.

  All IS-IS devices in Level-2 domain must be configured with the same authentication password. This command does not take effect if the **auth method** command has been run to configure authentication before.

- Specify the IS-IS routing domain authentication mode.

  auth method { **md5** | **text** } [**level-1** | **level-2**]

  The authentication function is disabled by default.

- Specify the key chain for IS-IS authentication.

  auth keychain *name-of-chain* [**level-1** | **level-2**]

  No authentication key chain is configured by default.

  Authentication is not performed if no key chain is configured using the **key chain** command. In addition to the key chain command, you also need to run the **auth method** command to make IS-IS key chain authentication take effect.

- (Optional) Apply IS-IS authentication only to send packets.

  auth transmit-only [**level-1** | **level-2**]

  Packets sent and received are authenticated by default.

1.15 Configuring IS-IS GR

1.15.1 Overview

As long as the network conditions remain unchanged, GR enables IS-IS to be restarted and restored to the pre-restart state without impact on data forwarding. This improves system reliability.

Two roles exist during the GR process: Restarter and Helper. Accordingly, the IS-IS GR function is divided into IS-IS GR Restart and IS-IS GR Help capabilities. A device with the GR Restart capability can send a GR request and execute GR. A device with the GR Help capability can receive a GR request and help its neighbor with GR implementation. The GR process starts when the Restarter sends a GR request. After receiving the GR request, the neighboring device enters Help mode to help the Restarter reestablish its LSDB while maintaining the neighbor relationship with the Restarter. The main GR working mechanism is as follows:

1.15.2 Restrictions and Guidelines

Specification
All products support the IS-IS GR Helper capability.

1.15.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable the IS-IS GR Restart capability.

gr

The IS-IS GR Restart capability is enabled by default.

(5) (Optional) Configure the maximum GR interval of a device.

gr grace-period *max-interval*

The maximum GR interval is **300** seconds by default.

(6) Enable the IS-IS GR Help capability.

gr helper cancel

The IS-IS GR Helper function is enabled by default.

This command enables IS-IS to ignore the GR request sent by the device to be restarted.

1.16 Configuring IS-IS NSR

1.16.1 Overview

IS-IS nonstop routing (IS-IS NSR) helps improve system reliability. During the active/standby switchover of devices in distributed or Virtual Switch Unit (VSU) mode, user data traffic keeps being forwarded without interruption. Information that can be backed up includes command configuration, neighbor backup, and LSP DB.

NSR backs up relevant IS-IS information from the master Supervisor Engine to the slave Supervisor Engine of the distributed device, or from the master device to the salve device in VSU mode, so that the device can automatically recover the link state and regenerate a route upon active/standby switchover, without requiring help from neighbor devices during the recovery.

Compared with GR that needs the help of a neighbor, NSR does not need any help from neighbors.

1.16.2 Restrictions and Guidelines

- For the same IS-IS process, either NSP or GR is enabled, because they are exclusive.
- The switchover of distributed devices and VSU devices takes a period of time. If the IS-IS neighbor keepalive duration is less than the switchover duration, IS-IS neighbor relationship with the neighbor device is removed, and the services are interrupted during the switchover. Therefore, it is recommended that the IS-IS neighbor keepalive duration be set not less than the default value. When Fast Hello is enabled, the IS-IS neighbor keepalive duration is less than 1 second and the IS-IS neighbor relationship times out during the switchover, causing NSR failures. Therefore, it is recommended that Fast Hello be disabled when NSR is enabled.

1.16.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable the NSR function.

nsr

The NSR function is disabled by default.

1.17 Configuring IS-IS Fast Reroute

1.17.1 Overview

When a link fault occurs, it takes about 10 to 30 seconds for IS-IS to detect the neighbor disconnection, advertise the link fault, recalculate the Shortest Path Tree (SPT), and update the routing table. During this period, there is a traffic interruption due to the absence of valid routes.

To minimize the traffic interruption caused by network failures, administrators can configure IS-IS Fast Reroute (FRR). When FRR is enabled, IS-IS precomputes primary and backup paths for each route during normal route calculation. These paths are updated to the routing table. In the event of a link fault, the device quickly detects the fault on the primary path and switches the traffic to the backup path, eliminating the need for IS-IS to recalculate the routes. This significantly reduces the traffic interruption time. When used with BFD for path detection, IS-IS FRR allows the link interruption time to be reduced from over 10 seconds to about 50 milliseconds. When the device forwards packets over the backup path, IS-IS recalculates the routes based on the changed network topology and updates the routing table. Subsequently, the device uses the newly calculated optimal routes for packet forwarding.

Administrators can use IS-IS automatic calculation of backup paths or directly specify backup paths for eligible routes based on comprehensive network planning.

1.17.2 Restrictions and Guidelines

- IS-IS basic functions must be configured.
- When FRR is enabled for IS-IS, you are advised to enable BFD at the same time so that the device can quickly detect link faults, thereby shortening the forwarding interruption time. If the interface is up or down, to shorten the forwarding interruption time during IS-IS fast reroute, you can configure carrier-delay 0 in interface configuration mode to achieve the fastest switchover speed.
- If an interface has limited remaining bandwidth or may fail simultaneously with the primary interface, it cannot be a backup interface. In this case, the command needs to be enabled in interface configuration mode to exclude the interface from being used as a backup interface. After this command is run, a backup interface will be selected from other interfaces during IS-IS FRR calculation.

1.17.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable FRR for IS-IS.

frr { **lfa** [**downstream-paths**] **| route-map** *route-map-name* } { **level-1 | level-2** }

By default, FRR is disabled for IS-IS.

(5) Return to the global configuration mode.

exit

(6) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet sub-interface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregate sub-interface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the tunnel interface configuration mode.

  interface tunnel *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

(7) (Optional) Configure LFA protection mode for an IS-IS interface.

isis frr protection { node | link-node | disable }

By default, LFA link protection is enabled.

This command does not take effect if the **fast-rerotue route-map** command is configured.

(8) (Optional) Configure the function of excluding an interface from being a backup.

isis frr no-eligible-backup

By default, IS-IS is enabled and neighboring interfaces can be used as backup interfaces.

This command does not take effect if **frr route-map** command is configured.

1.18 Correlating IS-IS with BFD

1.18.1 Overview

Bidirectional Forwarding Detection (BFD) is used to test connectivity between devices. A link fault can be detected in as short as 150 ms. After IS-IS is correlated with BFD, IS-IS can sense the death of a neighbor in as short as 150 ms once a link is faulty. This accelerates route convergence and prevents traffic interruption.

Normally, BFD sends detection packets at millisecond intervals to detect the link state. When a link exception (such as a disconnected link) occurs, BFD can quickly detect it and instruct IS-IS to delete the neighbor relationship and the neighbor reachability information in LSPs. Then IS-IS recalculates and generates a new route to bypass the abnormal link, thus realizing fast convergence. With the introduction of new techniques such as the Multi-Service Transport Platform (MSTP), link congestion tends to occur during peak hours of data communication. BFD quickly detects the link exception and instructs IS-IS to delete the neighbor relationship and the neighbor reachability information in LSPs. Link switch is performed to bypass the congested link. A Hello packet for IS-IS neighbor detection is sent every 10s and its expiration time is 30s. The Hello packet can still be received normally when BFD detects an exception, and therefore an IS-IS neighbor relationship is reestablished quickly, causing the route to be restored to the congested link. Then BFD detects the abnormal link and link switch is performed again. This process is repeated, which makes the route be switched between the congested link and other links, causing repetitive flapping.

The anti-congestion option is used to avoid routing flapping in case of link congestion. After the option is configured, the IS-IS neighbor state is still kept alive when link congestion occurs, but the neighbor reachability information in LSPs is deleted. The route is switched to a normal link. When the congested link is restored, the neighbor reachability information in LSPs is recovered and the route is switched back, which avoids route flapping.

1.18.2 Restrictions and Guidelines

- You must configure BFD session parameters before you enable IS-IS correlation with BFD.
- When you run the **bfd up-dampening** command on an interface with IS-IS correlation with BFD, you must run the **bfd whole-interfaces [anti-congestion]** command.
- When you run the **bfd whole-interfaces [anti-congestion]** command, you must run the **bfd up-dampening** command on the interface. The two commands must be used together. If you run only one command, the anti-congestion feature may not take effect or other network exceptions may occur.
- The IS-IS BFD TLV enable option takes effect only after it is configured at both ends and BFD-enable TLV information at both ends is consistent.
- IP routing may cause a neighbor's interface for BFD session setup to be inconsistent with the interface for outgoing BFD packets. If this happens, the BFD session cannot be set up.
- If a neighbor's interface for BFD session setup is inconsistent with the interface for outgoing BFD packets, the BFD session cannot be set up.

1.18.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable BFD correlation in IS-IS process mode.

bfd whole-interfaces [**anti-congestion | tlv-enable**]

The IS-IS correlation with BFD function is disabled on all interfaces by default.

(5) Return to the global configuration mode.

exit

(6) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(7) Enable IS-IS correlation with BFD on an interface.

isis bfd-enable [**cancel** | **anti-congestion | tlv-enable**]

The IS-IS correlation with BFD function is enabled on an interface by default if the **bfd whole-interfaces** command is run. The IS-IS correlation with BFD function is not enabled on an interface if the **bfd whole-interfaces** command is not run. The anti-congestion option is disabled by default.

1.19 Configuring the IS-IS Overload Bit

1.19.1 Overview

An IS-IS node is forced to configure the overload bit in a non-virtual LSP to instruct its IS-IS neighbors to avoid using the local node as a forwarding device.

1.19.2 Restrictions and Guidelines

- If the **startup** keyword is selected, the IS-IS node automatically enters overload state after restart. If the **startup** keyword is not selected, the IS-IS node enters overload state immediately after restart.

- The overload bit is used in the following three situations:
  - Device overload

    The local IS-IS node has overload issues, such as insufficient memory or full CPU load; as a result, its routing table has incomplete routes or does not have resource forwarding data. You can configure the overload bit in an LSP to instruct the neighbor to avoid using the local node as a forwarding device.

    To configure the overload bit, run the **remain-overload** command without the **startup** keyword. The overload bit can be configured or removed manually. When the local IS-IS node is restored, manually remove the command configuration; otherwise, the node is always in overload state.
  - Instantaneous black hole

    In the scenario described by RFC 3277, the IS-IS convergence speed is faster than the BGP speed; as a result, after an IS-IS node is restarted, a route may be instantaneously unreachable, which is called an instantaneous black hole. You can set the overload bit in an LSP to instruct the neighbor to avoid using the local node as a forwarding device until the specified time has elapsed.

    To set the overload bit, run the **remain-overload** command with the **startup** keyword. The overload bit can be configured or removed automatically by the IS-IS node based on the configuration. After the **startup** keyword is selected, the IS-IS node automatically enters instantaneous black hole state after restart. When a neighbor relationship is established, the IS-IS node sends an LSP with the overload bit to notify the neighbor that the local node enters instantaneous black hole (or overload) state and instruct the neighbor to avoid using the local node as a forwarding device. After the specified time has elapsed, the IS-IS node immediately sends an LSP with the overload bit canceled to notify the neighbor that the local node has exited instantaneous black hole (or overload) state and can work as a forwarding device.
  - Disabling real data forwarding on the local IS-IS node

    If you only need to connect the local IS-IS node to a production network for testing or to meet other functional requirements, but does not require the node to forward real data in the network, you can set the overload bit in an LSP to instruct the neighbor to avoid using the local node as a forwarding device.

    To set the overload bit, run the **remain-overload** command without the **startup** keyword. The overload bit can be configured or removed manually. You can set the **restrain** keyword based on requirements to limit the routing information carried in an LSP in case of overload. For example, internal and external routes can be suppressed, and only the local direct route is advertised. For example, the advertised direct route and the metric value to reach a neighbor are set to the maximum values.

1.19.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure the IS-IS overload bit.

remain-overload [**startup** { *overload-time* | **wait-bgp-converge** [*bgp-convergence-time*] }] [**restrain** { **interlevel** | **external** | **max-advertise-metric** } *] [**level-1** | **level-2**]

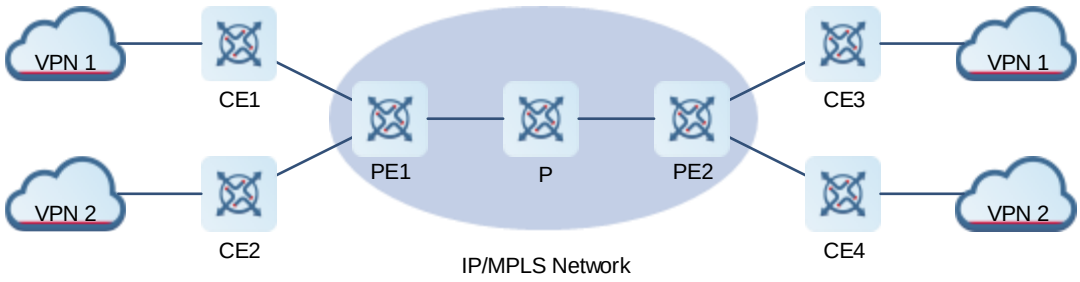
A neighbor considers the local IS-IS node as a forwarding device to forward data by default.

1.20 Configuring IS-IS VRF

1.20.1 Overview

VPN routing and forwarding (VRF) is mainly used for local routing and packet. It avoids route conflict caused by use of the same prefix by multiple VPNs. IPv4 VPN and IPv6 VPN combine Multiprotocol Label Switching (MPLS) advantages in terms of Quality of Service (QoS) and security assurance, and are the primary solutions for interconnecting the geographically different office branches of an enterprise or industry user.

Figure 1-10 Separation of Different VPNs by VRF Tables Configured on Provider Edge (PE) Devices



As shown in Figure 1-10, configure the two sites (CE1 and CE3) in VPN1 to access each other and the two sites (CE2 and CE4) in VPN2 to access each other, and forbid access between the sites in VPN1 and those in VPN2 because VPN1 and VPN2 belong to different customers or departments and may have identical IP addresses.

The customer edge (CE) devices connect the customer network to the PEs to exchange VPN routing information with the PEs, that is, advertise local routes to the PEs and learn remote routes from the PEs.

Each PE learns routes from directly connected CEs and exchanges the learned VPN routes with the other PEs through the Border Gateway Protocol (BGP). The PEs provide access to the VPN service.

The Provider (P) device in the Service Provider (SP) network-advertise is not directly connected to the CEs. The P device only needs the MPLS forwarding capability and does not maintain VPN information.

The IS-IS protocol running between the PEs and CEs requires the VRF capability to separate routing information between VPN1 and VPN2. That is, IS-IS only learns routes through VRF.

1.20.2 Restrictions and Guidelines

- The IS-IS instances bound with the same VRF table must be configured with different system IDs. The IS-IS instances bound with different VRF tables can be configured with the same system ID.
- One IS-IS instance can be bound with only one VRF table, but one VRF table can be bound to multiple IS-IS instances.
- When the VRF table bound to an IS-IS instance is changed, all IS-IS interfaces associated with the instance will be deleted. That is, the **ip router-protocol isis [tag] interface** configuration and the redistribution configuration in routing configuration mode will be deleted.

1.20.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure IS-IS VRF.

vrf *vrf-name*

VRF binding of an IS-IS instance is disabled by default.

1.21 Configuring IS-IS MTR

1.21.1 Overview

The IS-IS MTR technique helps users deploy IPv6 unicast networks without the constraint on consistency between IPv4 and IPv6 unicast topology information.

1.21.2 Restrictions and Guidelines

- Metric-style must be set to **Wide** or **Transition** before you enable this function.
- The MTR feature will be disabled if Metric-style is set to **Narrow** or only one Level is configured to support Wide or Transition mode.

1.21.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enter the IS-IS IPv6 address family configuration mode.

address-family-mode ipv6 unicast

(5) Configure IS-IS MTR.

ipv6-multitopology [*transition*]

IS-IS is not configured with MTR by default. That is, IS-IS does not support IPv6 unicast topologies. Configure IS-IS SNMP.

1.22 Configuring IS-IS SNMP

1.22.1 Overview

Use the network management software to manage OSPF parameters and monitor the OSPF running status.

1.22.2 Restrictions and Guidelines

- The latest standards stipulate that the MIB operation can be performed on a single instance. By default, the MIB operation is performed on the first displayed IS-IS instance. Because multiple IS-IS instances can be configured, the administrator can use this command to specify the instances on which the MIB operation will be performed.

1.22.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IS-IS trap globally.

snmp-agent enable traps [*isis*]

The SNMP agent is forbidden to send Trap messages to the NMS by default.

This command must be used with the **snmp-agent host** command in global configuration mode so that trap messages can be sent.

(4) Configure an SNMP host globally.

snmp-agent host { *ipv4-addrress* | **ipv6** *ipv6-address* } [**vrf** *vrf-name*] [**traps**] [**version** { **1** | **2c** | **3** { **auth** | **noauth** | **priv** }] *community-string* [**udp-port** *port-num*] [*notification-type*]

No SNMP host address is configured by default.

(5) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(6) Bind the instances on which the IS-IS MIB operation will be performed.

enable mib-binding

By default, the SNMP operation is performed on the first displayed IS-IS instance.

(7) Allow sending trap messages.

enable traps { **all** | *traps set* }

The IS-IS trap message sending function is disabled by default.

1.23 Enabling Super VLAN for IS-IS

Note

This function is supported by only the S6920-4C, S6520-64CQ, S6510-4C, S6510-32CQ, or S6510-48VS8CQ.

1.23.1 Overview

In normal cases, a super VLAN contains multiple sub VLANs. Multicast packets of a super VLAN are also sent to its sub VLANs. In this case, when IS-IS multicast packets are sent over a super VLAN containing multiple sub VLANs, the IS-IS multicast packets are replicated multiple times, which exceeds the processing capability of the device. As a result, a large number of packets are discarded, causing protocol flapping. In most scenarios, the IS-IS function does not need to be enabled on a super VLAN, and it is disabled by default. In some application

scenarios, the IS-IS function must be run on the super VLAN, but packets only need to be sent to one sub VLAN. In this case, you can decide to send multicast packets to a certain sub VLAN or to all sub VLANs as actually needed using this command. You must be cautious when configuring packet transmission to all sub VLANs, as the large number of sub VLANs may cause a device processing bottleneck, which will lead to the neighbor flapping.

1.23.2 Restrictions and Guidelines

- The IS-IS basic functions must be configured.
- The specified sub VLAN must be connected to neighbors.

1.23.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI configuration mode.

- **interface vlan** *interface-number*

(4) Configure IS-IS to run on a super VLAN.

isis subvlan { **all |** *vlan-id* }

The IS-IS function takes effect in super VLAN only and is disabled by default.

1.24 Configuring IS-IS Two-way Maintenance

1.24.1 Overview

In a large-scale network, a large number of packets are sent and received, which occupies lots of CPU and memory resources, causing some IS-IS packets to be delayed or discarded. If the time required for processing hello packets exceeds the neighbor relationship maintenance time, the corresponding neighbor relationship times out and is removed. If the two-way maintenance function is enabled, in addition to the hello packets, the LSP, CSNP, and PSNP packets from a neighbor can also be used to maintain the bidirectional communication between neighbors when a large number of packets exist in the network. This prevents termination of the adjacency caused by delayed or discarded hello packets.

1.24.2 Restrictions and Guidelines

- The IS-IS basic functions must be configured.
- The neighbor relationship is successfully established.

1.24.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable the IS-IS two-way maintenance function.

two-way-maintain

The two-way maintenance function is enabled by default.

1.25 Configuring IS-IS BGP-LS

1.25.1 Overview

BGP-link state (BGP LS) is a new way to collect the network topology. It makes topology collection more simple and efficient. For the IS-IS protocol, network topology information includes information about nodes, links, and route prefixes, which are all carried in LSP packets. When the BGP-LS function is enabled on an IS-IS instance, IS-IS parses out information about nodes, links, and route prefixes from LSP packets, encapsulates the information according to the RFC-7752 standard, and advertises the information to BGP. BGP summarizes information and reports it to the upper level controller, which performs path selection and calculation.

1.25.2 Restrictions and Guidelines

- The IS-IS basic functions must be configured.
- LSP packets must exist in the LSPDB of the corresponding level in the local IS-IS instance. For the LSP configuration, see "Configuring IS-IS LSPs"

1.25.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing process configuration mode.

router-protocol isis [*tag*]

(4) Configure the BGP-LS function on the IS-IS instance.

distribute bgp-ls [**instance-id** *instance-id*] [**level-1** | **level-2**] [**throttle** *delay-time*]

By default, the BGP-LS function is disabled on an IS-IS instance.

1.26 Configuring IS-IS ATT Bit

1.26.1 Overview

An IS-IS Attached (ATT) bit generated by a Level-1-2 router identifies whether a Level-1 area is attached to other areas. The ATT bit is related to Level-1 LSPs only. After a Level-1 router receives the LSPs carrying the ATT bit 1 from the Level-1-2 router, the Level-1 router generates a route with the same destination address as the default route of the Level-1-2 router. Traffic can be forwarded along this route, achieving mutual communication with other areas. To modify the default route generation of routers in a Level-1 area, you can

configure this function. After IS-IS setting ATT bit is configured, IS-IS will modify the ATT bit in Level-1 LSPs based on the configuration, and the default route generation will also change accordingly.

1.26.2 Restrictions and Guidelines

This command takes effect for Level-1-2 routers only.

The ATT bit in Level-1 LSPs identifies whether a Level-1 area is attached to other areas. The default rule is as follows:

- If a Level-1-2 router is attached to a Level-2 backbone area, the Level-1-2 router sets the ATT bit in its generated Level-1 LSP to 1. After a Level-1 router receives the LSPs carrying the ATT bit 1 from the Level-1-2 router, the Level-1 router generates a route with the same destination address as the default route of the Level-1-2 router. Traffic can be forwarded along this route, achieving mutual communication with other areas. To prevent the Level-1 router connected to the Level-1-2 router from advertising the default route when the ATT bit is set to 1, use either of the following methods:
  - Run the **att-bit-ignore** command on the Level-1 router that is connected to the Level-1-2 router.
  - Run the **att-bit-set never** command on the Level-1-2 router to disable the router from sending Level-1 LSPs with the ATT bit 1.
- If the Level-1-2 router is connected to only one area, the Level-1-2 router sends the Level-1 LSPs carrying the ATT bit 0. To enable the ATT bit in Level-1 LSPs remain at 1, run the **att-bit-set always** command.

If multiple topology configurations exist, the ATT bit in the IPv6 unicast topology is set in the MT ID 2 of TLV 229 in Level-1 LSPs. The default rule and the rule for changing the default settings are the same as the ATT bit rule in LSPs.

When this command is configured in IS-IS routing process configuration mode, it applies to the ATT bit in Level-1 LSPs, that is, controlling the calculation of default IPv4 and IPv6 routes in the default topology.

When this command is configured in IS-IS address-family ipv6 configuration mode, it applies to the ATT bit in MT ID 2 of TLV 229 in Level-1 LSPs, that is, controlling the calculation of default routes for IPv6 single-topology in multiple topologies.

1.26.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure the **att-bit-set** function on an IS-IS instance.

att-bit-set { **always** | **never** }

By default, an IS-IS Level-1 router sets the ATT bit in LSPs using the default rule.

(5) Configure the **att-bit-ignore** function on an IS-IS instance.

att-bit-ignore

By default, the IS-IS-enabled router advertises the default route when the ATT bit is set to 1.

1.27 Configuring Extended Feature of IS-IS

1.27.1 Overview

IS-IS provides rich extended features that can be configured based on actual needs.

1.27.2 Configuration Tasks

The IS-IS extended feature configuration tasks are as follows and optional. Select tasks for configuration according to the actual condition.

- Configuring the Maximum Metric of a Neighbor
- Configuring the Maximum Number of Area Addresses
- Configuring SPF Calculation Interval
- Canceling Three-Way Handshake of a P2P Network
- Enabling TLV to Be Padded Based on Supported Protocol
- Configuring the Reachable Time of a Suppressed Neighbor
- Suppressing Advertisement of Directly Connected Routes on an Interface

1. Overview

By default, directly connected routes of an interface are advertised in LSPs for route calculation. If you want to exclude the directly connected routes from route calculation on a specific interface, you can configure this function. After this command is configured, the directly connected routes of a specified protocol will not be advertised on an IS-IS interface and the establishment of IS-IS neighbors is still allowed. This function can suppress the interface prefixes carried in LSPs to keep them invisible to external entities, thereby enhancing security and accelerating route convergence.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable the function of suppressing the advertisement of directly connected routes on an IS-IS interface.

isis suppress-reachability

By default, the advertisement of directly connected routes is not suppressed on an IS-IS interface.

- Enabling Adjacency Event Output

1.27.3 Configuring the Maximum Metric of a Neighbor

1. Overview

In the Overlay scene where IS-IS is applied to Underlay, the Overlay tunnel may rely on Underlay routing. After the IS-IS neighbor is up, the Underlay route is reachable but the Overlay tunnel may not be created, which may lead to traffic interruption. This function can prevent traffic interruption.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure the maximum metric for the directly-connected route after the first neighbor is up.

max-advertise-metric on-neighbor-up *delay-time*

The metric of the directly-connected route is not set to the maximum value by default.

1.27.4 Configuring the Maximum Number of Area Addresses

1. Overview

Generally, an IS-IS process is configured with a NET address. During area reallocation, an IS-IS process can be configured with multiple NET addresses to ensure routing correctness. The system ID of multiple NET addresses must be the same. For Level-1 IS-IS routers, neighbor relationship can be created between the routers only when the maximum numbers of area addresses are the same.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure the maximum number of area addresses.

maximum-area-addr *max-area-number*

The maximum number of area addresses is **3** by default.

1.27.5 Configuring SPF Calculation Interval

1. Overview

Increasing the maximum interval for performing SPF calculations can avoid frequent SPF calculations and waste of CPU resources. However, a larger interval also leads to slower responses to route changes.

2. Restrictions and Guidelines

- The first wait time for performing the SPF calculation is *initial-interval*. If the network is unstable, the SPF calculation trigger interval is smaller than *maximum-interval* and the second wait time for performing the SPF calculation is changed to *hold-interval*. A corresponding penalty is added to this interval: The next interval for the SPF calculation doubles the previous interval for the same SPF calculation, until the SPF calculation interval reaches the *maximum interval*. Subsequent SPF calculations are performed at the *maximum interval*. When the network becomes stable, the interval for performing the SPF calculation becomes greater than the *maximum interval*, and the wait time for performing the SPF calculation is restored to the *initial interval*.
- The value of the configured *initial-interval* cannot be greater than that of *maximum-interval*. Otherwise, the value of *initial-interval* is changed to that of *maximum-interval*.
- The value of the configured *hold-interval* cannot be greater than that of *maximum-interval*. Otherwise, the value of *hold-interval* is changed to that of *maximum-interval*.
- The value of the configured *initial-interval* cannot be greater than that of *hold-interval*. Otherwise, the value of *initial-interval* is changed to that of *hold-interval*.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure the SPF calculation interval.

spf-throttling-interval [**level-1** | **level-2**] *maximum-interval* [*initial-interval hold-interval*]

By default, the maximum calculation interval of two SPF calculations is **10** seconds, the delay of the first SPF calculation is **50** ms, and the maximum interval for the first and second SPF calculations is **200** ms. Exponential backoff algorithm uses Level-1/Level-2. That is, it takes effect for Level-1 and Level-2 concurrently.

1.27.6 Canceling Three-Way Handshake of a P2P Network

1. Overview

Neighbor establishment in a P2P network requires three-way handshake. The neighbor relationship can be established only after the three-way handshake succeeds. If you want to accelerate neighbor establishment or

there is device that does not support three-way handshake, you can run this command to cancel three-way handshake.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregated sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Cancel three-way handshake of a P2P network.

isis three-way-handshake disable

Three-way handshake is performed by default.

1.27.7 Enabling TLV to Be Padded Based on Supported Protocol

1. Overview

When network devices supporting IPv4 IS-IS, IPv6 IS-IS, and IPv4 IS-IS and IPv6 are configured to establish a neighbor relationship, the calculated routes may be unreachable, resulting in routing black holes. To prevent routing black holes, the TLV (#129) field of device IS-IS protocol is padded based on protocols supported by an instance, that is, IS-IS instances that establish a neighbor relationship need to support the same protocol.

2. Restrictions and Guidelines

The TLV (#129) field of the IS-IS protocol supported by devices of some vendors is padded based on protocols supported by an interface. In single topology mode, a device connects to a device of another vendor, a loopback interface is configured on both devices, both IS-IS IPv4 IS-IS and IPv6 IS-IS are enabled, and only a single protocol stack (IPv4 IS-IS or IPv6 IS-IS) is configured on the interconnected interfaces of the two devices. The

device of the other vendor sends TLV#129 (supporting only a single protocol stack) based on interfaces whereas device sends TLV#129 (supporting dual protocol stacks) based on instances. As a result, the displayed neighbor status on the device of the other vendor is "Init". Devices failed to establish a neighbor relationship with the device of the other vendor. For this, the **interfaces-protocol-compatible** command can be configured to enable devices to establish a neighbor relationship with devices of other vendors.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Configure the TLV field of the IS-IS protocol to be padded based on protocols supported by an interface.

interfaces-protocol-compatible

The TLV field of the IS-IS protocol is padded based on protocols supported by an interface by default.

1.27.8 Configuring the Reachable Time of a Suppressed Neighbor

1. Overview

After the interface neighbor is up, this command prevents the neighbor reachability information from being added to LSP so as to delay the routing calculation. When the timer expires, the neighbor reachability information is added to LSP to start the routing calculation. This function prevents the route calculation from using the old LSP, which may lead to route flapping.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Suppress route calculation after the IS-IS neighbor is up.

isis suppress on-neighbor-up *delay-time*

The route calculation suppression function is disabled by default.

1.27.9 Suppressing Advertisement of Directly Connected Routes on an Interface

1. Overview

By default, directly connected routes of an interface are advertised in LSPs for route calculation. If you want to exclude the directly connected routes from route calculation on a specific interface, you can configure this function. After this command is configured, the directly connected routes of a specified protocol will not be

advertised on an IS-IS interface and the establishment of IS-IS neighbors is still allowed. This function can suppress the interface prefixes carried in LSPs to keep them invisible to external entities, thereby enhancing security and accelerating route convergence.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable the function of suppressing the advertisement of directly connected routes on an IS-IS interface.

isis suppress-reachability

By default, the advertisement of directly connected routes is not suppressed on an IS-IS interface.

1.27.10 Enabling Adjacency Event Output

1. Overview

You can run a debugging command to record neighbor state changes of IS-IS, but the command consumes many system resources. You can enable adjacency event output to avoid this issue.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IS-IS routing configuration mode.

router-protocol isis [*tag*]

(4) Enable neighbor relationship event output.

log-adj-changes

The function of recording neighbor state changes of IS-IS is enabled by default without the debugging command.

1.28 Monitoring

Run the display command to check the running status of a configured function to verify the configuration effect.

Run the cleanup command to clear information.

⚠ Caution

Running the **cleanup** command during operation of the device may lose vital information and interrupt services.

Run the debugging command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-2 Monitoring

Command	Purpose
cleanup clns neighbors-table	Clears all IS-IS neighbor relationship tables.
cleanup isis *	Clears all IS-IS data structures.
cleanup isis [*tag*] **counter**	Clears different statistical information of IS-IS.
cleanup isis [*tag*] **bgp-ls**	Clears BGP-LS information of IS-IS.
display clns [*tag*] **is-neighbors-info** [*interface-type interface-number*] [**detail**]	Displays all IS-IS neighbors and provides device adjacency relationships.
display clns [*tag*] **neighbors-info** [*interface-type interface-number*] [**detail**]	Displays all IS neighbors and provides device information and adjacency relationship information about terminals.
display isis-instance [*tag*] **counter**	Displays different statistical information of IS-IS.
display isis-instance [*tag*] **database** [**detail** | **verbose**] [**l1** | **l2** | **level-1** | **level-2**] [*lspid*]	Displays LSP DB information.
display isis-instance [*tag*] **gr**	Displays the state information related to IS-IS GR.
display isis-instance [*tag*] **hostname**	Displays the relationship between the device name and system ID.
display isis-instance [*tag*] **interface** [*interface-type interface-number*] [*counter*]	Displays the details of an IS-IS interface.
display isis-instance [*tag*] **fully-connected-groups**	Displays the mesh group configuration of different interfaces.
display isis-instance [*tag*] **neighbors** [**detail**]	Displays IS-IS neighbor information.
display isis-instance [*tag*] **nsr**	Displays IS-IS NSR information.
display isis-instance [*tag*] **virtual-neighbors**	Displays the neighbor information of virtual systems in IS-IS.
display isis-instance [*tag*] **protocol**	Displays the protocol information of IS-IS.
display isis-instance [*tag*] **topology-info** [**l1** | **l2** | **level-1** | **level-2**]	Displays the topology of IS-IS device connection.

Command	Purpose
display isis-instance [*tag*] **ipv6 topology-info** [**l1** | **l2** | **level-1** | **level-2**]	Displays IS-IS IPv6 unicast topology information of IS-IS.
display isis-instance [*tag*] **frr**	Displays IS-IS FRR information.
debugging isis { **all** | **auth | events** | **gr** | **ifsm** | **intf** | **lsp** | **mtr** | **nfsm** | **nsm** | **nsr** | **pdu** | **rdnd | segment-routing** [**mpls** | **prefix-sid** [*sid*] | **adjacency-sid** [*sid*] | **srgb** | **srlb**] | **distribute** | **mom** | **mpls | spf** | **warn** }	Enables IS-IS debugging.

1.29 Configuration Examples

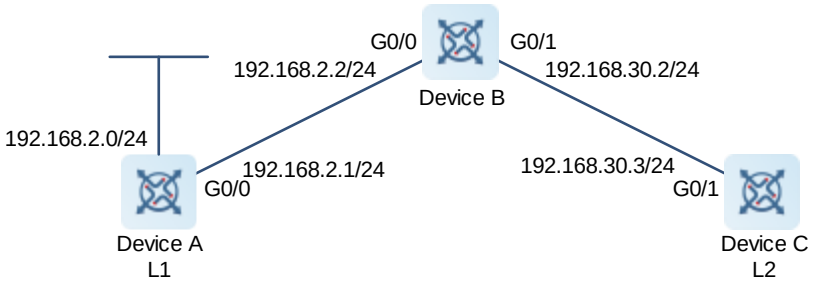
1.29.1 Performing Basic Authentication Configuration of IS-IS

1. Requirements

Devices A, B, and C are connected through the Ethernet and run IS-IS. Device A is a Level-1 device, Device B is a Level-1/Level-2 device, and Device C is a Level-2 device. The following configuration requirements exist: Plaintext authentication is applied to the Hello packets between Device A and Device B, as well as Level-1 LSPs and SNPs. MD5 authentication is applied to the Hello packets between Router B and Router C, as well as Level-2 LSPs and SNPs.

2. Topology

Figure 1-11 Configuring Basic Authentication Configuration of IS-IS



3. Notes

- Configure a key string.
- Configure IS-IS area authentication and interface authentication.

4. Procedure

(1) Configure IP addresses for interfaces.

Configure Device A.

```
Device A> privilege
```

```
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/0
Device A(cfg-if-GigabitEthernet 0/0)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/0)# ip address 192.168.20.1 255.255.255.0
Device A(cfg-if)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/0
Device B(cfg-if-GigabitEthernet 0/0)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/0)# ip address 192.168.20.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/0)# exit
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.30.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# exit
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/1)# ip address 192.168.30.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/1)# exit
```

(2) Perform basic configuration of IS-IS.

Configure Device A.

```
Device A(cfg)# router-protocol isis
Device A(cfg-router)# net 49.0001.0000.0000.0001.00
Device A(cfg-router)# is-type level-1
Device A(cfg-router)# exit
Device A(cfg)# interface gigabitethernet 0/0
Device A(cfg-if-GigabitEthernet 0/0)# ip router-protocol isis
```

Configure Device B.

```
Device B(cfg)# router-protocol isis
Device B(cfg-router)# net 49.0001.0000.0000.0002.00
Device B(cfg-router)# exit
Device B(cfg)# interface gigabitethernet 0/0
Device B(cfg-if-GigabitEthernet 0/0# ip router-protocol isis
Device B(cfg-if-GigabitEthernet 0/0)# exit
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device B(cfg-if-GigabitEthernet 0/1)# exit
```

Configure Device C.

```
Device C(cfg)# router-protocol isis
```

```
Device C(cfg-router)# net 49.0002.0000.0000.0003.00
Device C(cfg-router)# is-type level-2-only
Device C(cfg-router)# exit
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device C(cfg-if-GigabitEthernet 0/1)# exit
```

(3) Enable IS-IS authentication.

Set the area authentication password of Device A to **aa**, and set the interface authentication password of GigabitEthernet 0/0 to **cc**.

```
Device A(cfg)# router-protocol isis
Device A(cfg-router)# area-auth aa
Device A(cfg-router)# exit
Device A(cfg)# interface gigabitethernet 0/0
Device A(cfg-if-GigabitEthernet 0/0)# isis auth-password cc
```

Configure Device B with Level-1 area plaintext authentication and set the key string to **kc1**; configure Level-2 area plaintext authentication and set the key string to kc**2**. Configure Device A with interface plaintext authentication and set the key string to **kc3**. Configure Device C with interface MD5 authentication and set the key string to **kc3**.

```
Device B(cfg)# key chain kc1
Device B(cfg-keychain)# key 1
Device B(cfg-keychain-key)# key-string aa
Device B(cfg-keychain-key)# exit
Device B(cfg-keychain)# exit
Device B(cfg)# key chain kc2
Device B(cfg-keychain)# key 1
Device B(cfg-keychain-key)# key-string bb
Device B(cfg-keychain-key)# exit
Device B(cfg-keychain)# exit
Device B(cfg)# key chain kc3
Device B(cfg-keychain)# key 1
Device B(cfg-keychain-key)# key-string cc
Device B(cfg-keychain-key)# exit
Device B(cfg-keychain)# exit
Device B(cfg)# router-protocol isis
Device B(cfg-router)# auth method text level-1
Device B(cfg-router)# auth keychain kc1 level-1
Device B(cfg-router)# auth method md5 level-2
Device B(cfg-router)# auth keychain kc2 level-2
Device B(cfg-router)# exit
Device B(cfg)# interface gigabitethernet 0/0
Device B(cfg-if-GigabitEthernet 0/0)# isis auth method text
Device B(cfg-if-GigabitEthernet 0/0)# isis auth keychain kc3
Device B(cfg-if-GigabitEthernet 0/0)# exit
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# isis auth method md5
```

```
Device B(cfg-if-GigabitEthernet 0/1)# isis auth keychain kc3
```

Configure Device C with Level-2 area MD5 authentication and set the key string to **kc2**. Configure Device B with interface MD5 authentication and set the key string to **kc3**.

```
Device C(cfg)# key chain kc2
Device C(cfg-keychain)# key 1
Device C(cfg-keychain-key)# key-string bb
Device C(cfg-keychain-key)# exit
Device C(cfg-keychain)# exit
Device C(cfg)# key chain kc3
Device C(cfg-keychain)# key 1
Device C(cfg-keychain-key)# key-string cc
Device C(cfg-keychain-key)# exit
Device C(cfg-keychain)# exit
Device C(cfg)# router-protocol isis
Device C(cfg-router)# auth method md5 level-2
Device C(cfg-router)# auth keychain kc2
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# isis auth method md5
Device C(cfg-if-GigabitEthernet 0/1)# isis auth keychain kc3
```

5. Verification

Check the neighbor information of Device A.

```
Device A# display isis-instance neighbors
Area (null):
System Id          Type   IP Address          State    Holdtime   Circuit              Interface
B                  L1     192.168.20.2        Up       28         A.01                 GigabitEthernet 0/0
```

Check the IS-IS database information of Device A.

```
Device A# display isis-instance database detail
Area (null):
IS-IS Level-1 Link State Database:
LSPID                 LSP Seq Num  LSP Checksum  LSP Holdtime      ATT/P/OL
A.00-00             * 0x00000011   0xBE84        918               0/0/1
  Auth: TEXT(aa) Length: 3
  Area Address: 49.0001
  NLPID:          0xCC
  Hostname:       A
  IP Address:     192.168.20.1
  Metric:    1         IS A.01
  Metric:    1         IP 192.168.20.0 255.255.255.0
A.01-00             * 0x0000000A   0xF80A        918               0/0/0
  Auth: TEXT(aa) Length: 3
  Metric:    0         IS A.00
  Metric:    0         IS B.00
B.00-00               0x00000013   0xF62B        1129              1/0/0
  Auth: TEXT(aa) Length: 3
```

```
  Area Address: 49.0001
  NLPID:          0xCC
  Hostname:       B
  IP Address:     192.168.20.2
  Metric:    1          IS A.01
  Metric:    1          IP 192.168.20.0 255.255.255.0
  Metric:    1          IP 192.168.30.0 255.255.255.0
```

Check the neighbor information of Device B.

```
Device B# display isis-instance neighbors
Area (null):
System Id       Type   IP Address        State   Holdtime   Circuit          Interface
C               L2     192.168.30.3      Up      21         B.02             GigabitEthernet 0/1
A               L1     192.168.20.1      Up      8          A.01             GigabitEthernet 0/0
```

Check the neighbor information of Device C.

```
Device C# display isis-instance neighbors
Area (null):
System Id       Type   IP Address        State   Holdtime   Circuit          Interface
B               L2     192.168.30.2      Up      9          B.02             GigabitEthernet 0/0
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol isis
 net 49.0001.0000.0000.0001.00
 is-type level-1
 area-auth aa
!
interface gigabitethernet 0/0
 unset switch-port
 ip address 192.168.20.1 255.255.255.0
 ip router-protocol isis
 isis auth-password cc
!
```

- Device B configuration file

```
!
key chain kc1
 key 1
  key-string aa
key chain kc2
 key 1
  key-string bb
key chain kc3
 key 1
  key-string cc
```

```
!
router-protocol isis
 net 49.0001.0000.0000.0002.00
 auth method text level-1
 auth keychain kc1 level-1
 auth method md5 level-2
 auth keychain kc2 level-2
!
interface gigabitethernet 0/0
 unset switch-port
 ip address 192.168.20.2 255.255.255.0
 ip router-protocol isis
 isis auth method text
 isis auth keychain kc3
!
interface gigabitethernet 0/1
 unset switch-port
 ip address 192.168.30.2 255.255.255.0
 ip router-protocol isis
 isis auth method md5
 isis auth keychain kc3
!
```

- Device C configuration file

```
!
key chain kc2
 key 1
  key-string bb
!
key chain kc3
 key 1
  key-string cc
!
router-protocol isis
 net 49.0002.0000.0000.0003.00
 is-type level-2
 auth method md5 level-2
 auth keychain kc2
!
interface gigabitethernet 0/1
 unset switch-port
 ip address 192.168.30.3 255.255.255.0
 ip router-protocol isis
 isis auth method md5
 isis auth keychain kc3
!
```

7. Common Errors

- Different authentication passwords are configured between neighbors.
- Different authentication modes are configured between neighbors.

1.29.2 Configuring IS-IS MTR

1. Requirements

The typical application scenario of MTR is to retain devices that only support IPv4 services in a network where IPv6 service extension will be performed.

As shown in Figure 1-12, Device B only supports the IPv4 protocol stack but does not support the MTR feature; therefore, it can only run IPv4 services. The network capacity needs to be scaled to support IPv6 services in order to meet service extension requirements. (Devices A, C, and D that support the MTR feature will be added.) Device B that supports only one protocol stack must be replaced to maintain the stability of the network running IPv4 and IPv6 dual protocol stacks; otherwise, IPv6 routing black holes may occur.

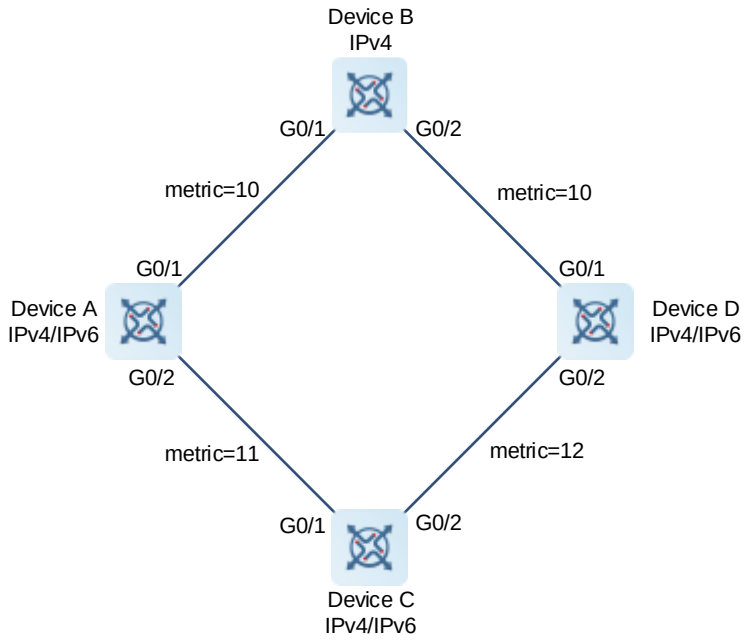
If you need to retain Device B, you can configure the MTR feature on the new devices. The MTR feature enables Device B to continue to run IPv4 services without interference on the IPv4 and IPv6 services on the new devices. The MTR feature improves networking flexibility, indirectly prolongs the service life of old devices, and meets service extension requirements while maximizing the values of old devices.

The configuration requirements are as follows:

- Retain Device B, which only supports IPv4 services.
- Add devices that support IPv4 and IPv6 dual protocol stacks, and separate IPv4 route calculation and IPv6 route calculation.

2. Topology

Figure 1-12 Configuring IS-IS MTR



3. Notes

- Perform basic configuration of IS-IS.
- Enable the MTR function.

4. Procedure

- Configure Device A.

  Configure IS-IS, set the metric type to Wide, and enable the MTR function.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# router-protocol isis
Device A(cfg-router)# net 49.0001. 0000.0000.0001.00
Device A(cfg-router)# is-type level-1
Device A(cfg-router)# metric-class wide-metric
Device A(cfg-router)# address-family-mode ipv6
Device A(cfg-router-af)# ipv6-multitopology
Device A(cfg-router-af)# exit
Device A(cfg-router)# exit
```

  Configure an IP address for the interface and add the interface to IS-IS.

```
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 enable
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 address 1002::1/112
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 router-protocol isis
Device A(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/2)# ipv6 enable
Device A(cfg-if-GigabitEthernet 0/2)# ipv6 address 1003::1/112
Device A(cfg-if-GigabitEthernet 0/2)# ipv6 router-protocol isis
Device A(cfg-if-GigabitEthernet 0/2)# ip address 192.168.2.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/2)# ip router-protocol isis
Device A(cfg-if-GigabitEthernet 0/2)#isis wide-metric 11
```

- Configure Device B.

  Configure IS-IS, set the metric type to Wide, and enable the MTR function.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# router-protocol isis
Device B(cfg-router)# net 49.0001. 0000.0000.0002.00
Device B(cfg-router)# is-type level-1
Device B(cfg-router)# metric-class wide-metric
Device B(cfg-router)# address-family-mode ipv6
Device B(cfg-router-af)# unset adjacency-support-check
```

```
Device B(cfg-router-af)# exit
Device B(cfg-router)# exit
```

Configure an IP address for the interface and add the interface to IS-IS.

```
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/2)# ip address 192.168.3.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/2)# ip router-protocol isis
```

- Configure Device C.

  Configure IS-IS, set the metric type to Wide, and enable the MTR function.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# router-protocol isis
Device C(cfg-router)# net 49.0001. 0000.0000.0003.00
Device C(cfg-router)# is-type level-1
Device C(cfg-router)# metric-class wide-metric
Device C(cfg-router)# address-family-mode ipv6
Device C(cfg-router-af)# ipv6-multitopology
Device C(cfg-router-af)# exit
Device C(cfg-router)# exit
```

Configure an IP address for the interface and add the interface to IS-IS.

```
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 enable
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 address 3001::1/112
Device C(cfg-if-GigabitEthernet 0/1)# ipv6 router-protocol isis
Device C(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device C(cfg-if-GigabitEthernet 0/1)#isis wide-metric 11
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface gigabitethernet 0/2
Device C(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/2)# ipv6 enable
Device C(cfg-if-GigabitEthernet 0/2)# ipv6 address 3004::1/112
Device C(cfg-if-GigabitEthernet 0/2)# ipv6 router-protocol isis
Device C(cfg-if-GigabitEthernet 0/2)# ip address 192.168.4.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/2)# ip router-protocol isis
Device C(cfg-if-GigabitEthernet 0/2)# isis wide-metric 12
```

- Configure Device D.

  Configure IS-IS, set the metric type to Wide, and enable the MTR function.

```
Device D> privilege
Device D# config-system terminal
Device D(cfg)# router-protocol isis
Device D(cfg-router)# net 49.0001.0000.0000.0004.00
Device D(cfg-router)# is-type level-1
Device D(cfg-router)# metric-class wide-metric
Device D(cfg-router)# address-family-mode ipv6
Device D(cfg-router-af)# ipv6-multitopology
Device D(cfg-router-af)# exit
Device D(cfg-router)# exit
```

Configure an IP address for the interface and add the interface to IS-IS.

```
Device D(cfg)# interface gigabitethernet 0/1
Device D(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device D(cfg-if-GigabitEthernet 0/1)# ipv6 enable
Device D(cfg-if-GigabitEthernet 0/1)# ipv6 address 4002::1/112
Device D(cfg-if-GigabitEthernet 0/1)# ipv6 router-protocol isis
Device D(cfg-if-GigabitEthernet 0/1)# ip address 192.168.3.4 255.255.255.0
Device D(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device D(cfg-if-GigabitEthernet 0/1)# exit
Device D(cfg)# interface gigabitethernet 0/2
Device D(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device D(cfg-if-GigabitEthernet 0/2)# ipv6 enable
Device D(cfg-if-GigabitEthernet 0/2)# ipv6 address 4003::1/112
Device D(cfg-if-GigabitEthernet 0/2)# ipv6 router-protocol isis
Device D(cfg-if-GigabitEthernet 0/2)# ip address 192.168.4.4 255.255.255.0
Device D(cfg-if-GigabitEthernet 0/2)# ip router-protocol isis
```

5. Verification

Run the display command on Device A to view IPv4 route information and check whether the next hop of the IPv4 route destined for Device D is Device B.

```
Device A# display ip route-config
Codes:   C - connected, S - static, R - RIP, B - BGP
         O - OSPF, IA - OSPF inter area
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         ia - IS-IS inter area, * - candidate default
Gateway of last resort is no set
C       192.168.1.0/24 is directly connected, GigabitEthernet 0/1
C       192.168.1.1/32 is local host.
C       192.168.2.0/24 is directly connected, GigabitEthernet 0/2
C       192.168.2.1/32 is local host.
i L1 192.168.3.0/24 [115/20] via 192.168.1.2, 00:13:14, GigabitEthernet 0/1
i L1 192.168.4.0/24 [115/23] via 192.168.2.3, 00:02:40, GigabitEthernet 0/2
```

Run the display command on Device A to view IPv6 route information and check whether the next hop of the IPv6 route destined for Device D is Device C.

```
Device A# display ipv6 route
IPv6 routing table name is - Default - 16 entries
Codes: C - Connected, L - Local, S - Static, R - RIP, B - BGP
       I1 - ISIS L1, I2 - ISIS L2, IA - ISIS interarea, IS - ISIS summary
       O - OSPF intra area, OI - OSPF inter area,  OE1 - OSPF external type 1, OE2 - OSPF external type 2
       ON1 - OSPF NSSA external type 1, ON2 - OSPF NSSA external type 2
L      ::1/128 via Loopback, local host
C       1002::/112 via GigabitEthernet 0/1, directly connected
L      1002::1/128 via GigabitEthernet 0/1, local host
C       1003::/112 via GigabitEthernet 0/2, directly connected
L      1003::1/128 via GigabitEthernet 0/2, local host
I1     3001::/112 [115/21] via FE80::C806:5FF:FEE8:38, GigabitEthernet 0/2
I1     3004::/112 [115/21] via FE80::C806:5FF:FEE8:38, GigabitEthernet 0/2
I1     4002::/112 [115/31] via FE80::C806:5FF:FEE8:38, GigabitEthernet 0/2
I1     4003::/112 [115/31] via FE80::C806:5FF:FEE8:38, GigabitEthernet 0/2
L      FE80::/10 via ::1, Null0
C       FE80::/64 via GigabitEthernet 0/2, directly connected
L      FE80::1614:4BFF:FE12:ADFC/128 via GigabitEthernet 0/2, local host
C       FE80::/64 via GigabitEthernet 0/1, directly connected
L      FE80::1614:4BFF:FE12:ADFD/128 via GigabitEthernet 0/1, local host
C       FE80::/64 via Local 0, directly connected
L      FE80::1614:4BFF:FE12:ADFC/128 via Local 0, local host
```

6. Configuration Files

Device A configuration file

```
!
router-protocol isis
 net 49.0001. 0000.0000.0001.00
 is-type level-1
 metric-class wide-metric
 address-family-mode ipv6
  ipv6-multitopology
!
interface gigabitethernet 0/1
 unset switch-port
 ipv6 enable
 ipv6 address 1002::1/112
 ipv6 router-protocol isis
 ip address 192.168.1.1 255.255.255.0
 ip router-protocol isis
!
interface gigabitethernet 0/2
 unset switch-port
 ipv6 enable
```

```
 ipv6 address 1003::1/112
 ipv6 router-protocol isis
 ip address 192.168.2.1 255.255.255.0
 ip router-protocol isis
 isis wide-metric 11
!
```

Device B configuration file

```
!
router-protocol isis
 net 49.0001. 0000.0000.0002.00
 is-type level-1
 metric-class wide-metric
 address-family-mode ipv6
  unset adjacency-support-check
!
 interface gigabitethernet 0/1
 unset switch-port
 ip address 192.168.1.2 255.255.255.0
 ip router-protocol isis
!
interface gigabitethernet 0/2
 unset switch-port
 ip address 192.168.3.2 255.255.255.0
 ip router-protocol isis
!
```

Device C configuration file

```
!
router-protocol isis
 net 49.0001. 0000.0000.0003.00
 is-type level-1
 metric-class wide-metric
 address-family-mode ipv6
  ipv6-multitopology
!
 interface gigabitethernet 0/1
 unset switch-port
 ipv6 enable
 ipv6 address 3001::1/112
 ipv6 router-protocol isis
 ip address 192.168.2.3 255.255.255.0
 ip router-protocol isis
 isis wide-metric 11
!
interface gigabitethernet 0/2
  unset switch-port
```

```
 ipv6 enable
 ipv6 address 3004::1/112
 ipv6 router-protocol isis
 ip address 192.168.4.3 255.255.255.0
 ip router-protocol isis
 isis wide-metric 12
!
```

Device D configuration file

```
!
router-protocol isis
 net 49.0001.0000.0000.0004.00
 is-type level-1
 metric-class wide-metric
 address-family-mode ipv6
  ipv6-multitopology
!
interface gigabitethernet 0/1
unset switch-port
 ipv6 enable
 ipv6 address 4002::1/112
 ipv6 router-protocol isis
 ip address 192.168.3.4 255.255.255.0
 ip router-protocol isis
!
interface gigabitethernet 0/2
unset switch-port
 ipv6 enable
 ipv6 address 4003::1/112
 ipv6 router-protocol isis
 ip address 192.168.4.4 255.255.255.0
 ip router-protocol isis
!
```

7. Common Errors

- The metric-style is not set to **Wide** or **Transition**.
- The protocol types used by two neighbors do not match; therefore, a neighbor relationship cannot be established.
- The IP addresses of the interfaces connected between neighbors are not in the same network segment.
- The **ip router-protocol isis** command is not executed on interfaces.
- No NET address is configured, or different NET addresses exist at Level-1.
- max-area-addresses is configured differently on both sides.
- metric-style is configured differently on both sides.
- The interface Levels on both sides are different. One side is Level-1, whereas the other side is Level-2.
- One side is configured with the P2P mode, whereas the other side is configured with the broadcast mode.

- One side is enabled with authentication, whereas the other side is not.

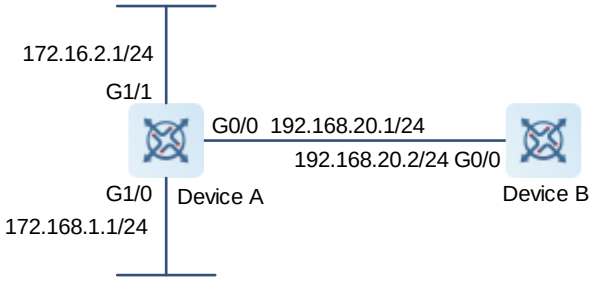
1.29.3 Configuring IS-IS Route Summarization

1. Requirements

Device A and Device B are connected through the Ethernet and run IS-IS. Device A is configured to advertise only the 172.16.0.0/22 route instead of the 172.16.1.0/24 and 172.16.2.0/24 routes.

2. Topology

Figure 1-13 Configuring IS-IS Route Summarization



3. Notes

- Configure IS-IS.
- Configure route summarization.

4. Procedure

(1) Configure IS-IS.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# router-protocol isis
Device A(cfg-router)# net 49.0001.0000.0000.0001.00
Device A(cfg-router)# exit
Device A(cfg)# interface gigabitethernet 0/0
Device A(cfg-if-GigabitEthernet 0/0)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/0)# ip address 192.168.20.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/0)# ip router-protocol isis
Device A(cfg-if-GigabitEthernet 0/0)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ip address 172.16.1.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device A(cfg-if-GigabitEthernet 0/1)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/2)# ip address 172.16.2.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/2)# ip router-protocol isis
Device A(cfg-if-GigabitEthernet 0/2)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# router-protocol isis
Device B(cfg-router)# net 49.0001.0000.0000.0002.00
Device B(cfg-router)# exit
Device B(cfg)# interface gigabitethernet 0/0
Device B(cfg-if-GigabitEthernet 0/0)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/0)# ip address 192.168.20.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/0)# ip router-protocol isis
```

(2) Configure route summarization.

Configure Device A.

```
Device A(cfg)# router-protocol isis
Device A(cfg-router)# summary-address 172.16.0.0/16 level-1-2
```

5. Verification

Run the display ip route-config command on Device B to check whether only one summary route exists.

```
Device B# display ip route-config
i L1      172.16.0.0/16 [115/20]    via 192.168.20.1, FastEthernet0/0
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol isis
 net 49.0001.0000.0000.0001.00
 summary-address 172.16.0.0/16 level-1-2
!
interface gigabitethernet 0/0
 unset switch-port
 ip address 192.168.20.1 255.255.255.0
 ip router-protocol isis
!
interface gigabitethernet 1/0
 unset switch-port
 ip address 172.16.1.1 255.255.255.0
 ip router-protocol isis
!
interface gigabitethernet 1/1
 unset switch-port
 ip address 172.16.2.1 255.255.255.0
 ip router-protocol isis
!
```

- Device B configuration file

```
!
```

```
router-protocol isis
 net 49.0001.0000.0000.0002.00
!
interface gigabitethernet 0/0
 unset switch-port
 ip address 192.168.20.2 255.255.255.0
 ip router-protocol isis
!
```

1.29.4 Correlating IS-IS with BFD

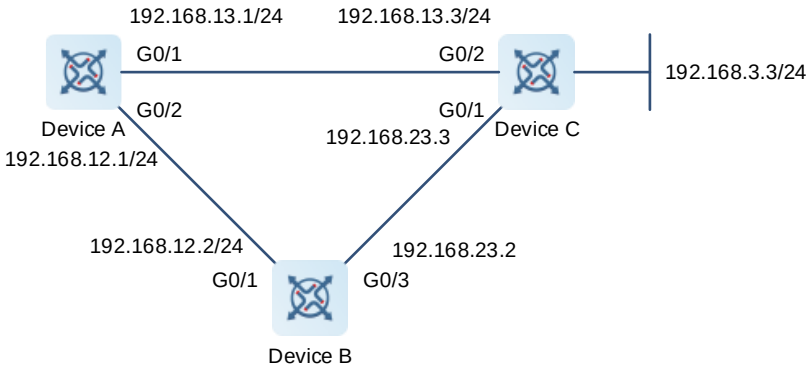
1. Requirements

Devices A, B, C, and D are interconnected through IS-IS routing protocol. The devices are Level-2 devices.

When a link from Device A to Device C fails, services can be quickly switched to Device B.

2. Topology

Figure 1-14 Correlating IS-IS with BFD



3. Notes

- Configure interface IP addresses for all the devices.
- Configure the basic IS-IS functions on all devices.
- Enable the BFD function in the IS-IS routing process mode.

4. Procedure

(1) Configure interface IP addresses and perform basic configuration of IS-IS.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/2)# ip address 192.168.12.1/24
Device A(cfg-if-GigabitEthernet 0/2)# ip router-protocol isis
```

```
Device A(cfg-if-GigabitEthernet 0/2)# exit
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ip address 192.168.13.1/24
Device A(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# router-protocol isis
Device A(cfg-router)# net 49.0001.0000.0000.0001.00
Device A(cfg-router)# is-type level-2-only
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.12.2/24
Device B(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/3
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/3)# ip address 192.168.23.2/24
Device B(cfg-if-GigabitEthernet 0/3)# ip router-protocol isis
Device B(cfg-if-GigabitEthernet 0/3)# exit
Device B(cfg)# router-protocol isis
Device B(cfg-router)# net 49.0001.0000.0000.0002.00
Device B(cfg-router)# is-type level-2-only
```

Configure Device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# interface gigabitethernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/2)# ip address 192.168.13.3/24
Device C(cfg-if-GigabitEthernet 0/2)# ip router-protocol isis
Device C(cfg-if-GigabitEthernet 0/2)# exit
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device C(cfg-if-GigabitEthernet 0/1)# ip address 192.168.23.3/24
Device C(cfg-if-GigabitEthernet 0/1)# ip router-protocol isis
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# router-protocol isis
Device C(cfg-router)# net 49.0001.0000.0000.0003.00
Device C(cfg-router)# is-type level-2-only
```

(2) Correlate IS-IS with BFD globally.

Configure Device A.

```
Device A(cfg-router)# bfd whole-interfaces
```

Configure Device B.

```
Device B(cfg-router)# bfd whole-interfaces
```

Configure Device C.

```
Device C(cfg-router)# bfd whole-interfaces
```

5. Verification

- Normal network environment

  Check the routing table of Device A.

```
Device A# display ip route-config
I L2    192.168.23.0/24 [115/2] via 192.168.13.3, 00:19:07, GigabitEthernet 0/1
                                [115/2] via 192.168.12.2, 00:19:07, GigabitEthernet 0/2
I L2    192.168.33.0/24 [115/1] via 192.168.13.3, 00:17:40, GigabitEthernet 0/1
```

  Run **traceroute 192.168.3.3** on Device A.

```
Device A# traceroute 192.168.3.3
   < press Ctrl+C to break >
Tracing the route to 192.168.3.3

 1          192.168.3.3        <1 msec       5 msec       9 msec
```

- Link from Device A to Device C fails

  Check the routing table of Device A.

```
Device A# display ip route-config isis
I L2    192.168.23.0/24 [115/2] via 192.168.12.2, 00:00:08, GigabitEthernet 0/2
I L2    192.168.33.0/24 [115/2] via 192.168.12.2, 00:00:08, GigabitEthernet 0/2
```

  Run **traceroute 192.168.3.3** on Device A.

```
Device A# traceroute 192.168.3.3
   < press Ctrl+C to break >
Tracing the route to 192.168.3.3

 1          192.168.12.2         9 msec       8 msec     13 msec
 2          192.168.3.3          8 msec     10 msec       9 msec
```

6. Configuration Files

- Device A configuration file

```
!
interface gigabitethernet 0/2
 unset switch-port
 ip address 192.168.12.1/24
 ip router-protocol isis
!
interface gigabitethernet 0/1
 unset switch-port
 ip address 192.168.13.1/24
 ip router-protocol isis
```

```
!
router-protocol isis
 net 49.0001.0000.0000.0001.00
 is-type level-2-only
!
```

- Device B configuration file

```
!
interface gigabitethernet 0/1
unset switch-port
 ip address 192.168.12.2/24
 ip router-protocol isis
!
interface gigabitethernet 0/3
 unset switch-port
 ip address 192.168.23.2/24
 ip router-protocol isis
!
router-protocol isis
 net 49.0001.0000.0000.0002.00
 is-type level-2-only
!
```

- Device C configuration file

```
!
interface gigabitethernet 0/2
 unset switch-port
 ip address 192.168.13.3/24
 ip router-protocol isis
!
interface gigabitethernet 0/1
 unset switch-port
 ip address 192.168.23.3/24
 ip router-protocol isis
!
router-protocol isis
 net 49.0001.0000.0000.0003.00
 is-type level-2-only
!
```

Contents

1 Configuring BGP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Concepts 1
1.1.3 Route Attributes of BGP 3
1.1.4 Route Exchange and Route Selection Policy of BGP 5
1.1.5 Neighbor Connection Establishment 7
1.1.6 Multi-Path Load Balancing of BGP 8
1.1.7 Route Reflector 8
1.1.8 BGP Alliance 9
1.1.9 Route Aggregation 9
1.1.10 Route Dampening 9
1.1.11 BGP Reliability 10
1.1.12 MP-BGP 12
1.1.13 BMP 13
1.1.14 BGP Neighbor Auto-Discovery 13
1.1.15 Others 13
1.1.16 Protocols and Standards 14
1.2 Configuration Task Summary 15
1.3 Configuring BGP Basic Features 16
1.3.1 Configuration Tasks 16
1.3.2 Configuring the BGP Version 17

1.3.3 Enabling BGP 17
1.3.4 Configuring a BGP Neighbor 18
1.3.5 Configuring a BGP IPv6 Address Family 20
1.3.6 Configuring a BGP Peer Group 21
1.3.7 Configuring the Dot Mode to Display 4-Byte AS Numbers 22
1.3.8 Injecting Network Information of Local AS into BGP 23
1.3.9 Obtaining Reachable Networks of Other AS from BGP 25
1.3.10 Configuring BGP Capacity Protection 26
1.3.11 Shutting Down BGP Connections Gracefully 27
1.3.12 Configuring BGP Soft Reset 28
1.3.13 Configuring Scope Configuration Mode 29
1.3.14 Configuring BGP Route Update Mechanisms 29
1.3.15 Configuring BGP Timers 31
1.4 Configuring a Large-Scale BGP Network 32
1.4.1 Configuration Tasks 32
1.4.2 Configuring BGP Route Reflector 32
1.4.3 Configuring BGP AS Alliance 33
1.4.4 Configuring Route Aggregation 34
1.4.5 Configuring Route Dampening 34
1.5 Configuring BGP Route Selection and Load Balancing 35
1.5.1 Configuration Tasks 35
1.5.2 Configuring BGP Multi-Path Load Balancing 35
1.5.3 Configuring BGP Routes to be Recursive Only to Host Routes 37
1.5.4 Configuring Outbound Loop Detection for a BGP Neighbor 38

1.5.5 Configuring BGP ADD-PATH 38
1.5.6 Configuring BGP Advertising Routes with the Lowest Priority upon Device Restart ..39
1.5.7 Configuring AS_PATH Attribute 40
1.5.8 Configuring MED Attribute 41
1.5.9 Configuring AIGP Attribute 42
1.5.10 Configure LOCAL_PREF Attribute 42
1.5.11 Configure Extended Community Attribute 43
1.5.12 Configuring BGP Best Path Selection 43
1.6 Configuring BGP Controlling Route Advertisement and Receiving 45
1.6.1 Configuration Tasks 45
1.6.2 Configuring BGP Filters 45
1.6.3 Configuring BGP Route Sending Interval and Count 47
1.6.4 Configuring BGP Modifying Propagated Routing Information 47
1.6.5 Configuring BGP Processing Received Routing Information 48
1.6.6 Configuring BGP Fast Withdrawing Specific Route 48
1.6.7 Configuring BGP Delaying Advertising Routing Information 49
1.7 Configuring the BGP Security Features 50
1.7.1 Overview 50
1.7.2 Configuring MD5 Authentication 51
1.7.3 Configuring BGP Peer GTSM Security Check 52
1.8 Configuring BGP Reliability 52
1.8.1 Configuration Tasks 52
1.8.2 Notes 52
1.8.3 Configuring BFD for BGP 53

1.8.4 Configuring BGP FRR....54
1.8.5 Configuring EBGP Multi-Path Bypass Protection....55
1.8.6 Configuring BGP GR....55
1.8.7 Configuring BGP NSR....57
1.8.8 Configuring BGP Session Keepalive....58
1.8.9 Configuring EBGP Fast Neighbor Detection....58
1.8.10 Configuring BGP Link-State Group Association....59
1.9 Configuring MP-BGP....60
1.9.1 Configuration Tasks....60
1.9.2 Configuring an EVI....60
1.9.3 Configuring EVPN....61
1.10 Configuring BGP Neighbor Auto-Discovery....63
1.10.1 Configuration Tasks....63
1.10.2 Configuring BGP Neighbor Auto-Discovery....63
1.10.3 Configuring Interface Neighbor Auto-Discovery....64
1.11 Configuring BGP Extended Features....65
1.11.1 Configuration Tasks....65
1.11.2 Configuring BMP Monitoring....65
1.11.3 Configuring BGP Administrative Distance....67
1.11.4 Configuring Enhanced VPN Route Import....68
1.11.5 Configuring BGP Link-State....69
1.12 Monitoring....69
1.13 Configuration Examples....80
1.13.1 Configuring BGP Basic Networking....80

1.13.2 Configuring BGP MD5 Authentication 89
1.13.3 Configuring a BGP Route Reflector 94
1.13.4 Configuring a BGP Alliance 99
1.13.5 Configuring IBGP Multi-Path Load Balancing 106
1.13.6 Configuring EBGP FRR 109
1.13.7 Configuring BGP to Rapidly Withdraw Specified Routes 114
1.13.8 Configuring the BGP Local AS 115
1.13.9 Configuring BGP GR 117
1.13.10 Configuring IPv6 Route Exchange Across ASs 121
1.13.11 Configuring the BGP EVPN 124
1.13.12 Configuring Compatibility Between BGP Devices Supporting 4-Byte AS Numbers and Those Supporting 2-Byte AS Numbers 128
1.13.13 Configuring an IPv6 Local Link Address 131
1.13.14 Configuring BGP NSR 134
1.13.15 Configuring BGP Routes to Be Recursive Only to Host Routes 135
1.13.16 Configuring Outbound Loop Detection for a BGP Neighbor 137
1.13.17 Shutting Down BGP Connections Gracefully 139
1.13.18 Configuring BGP Multi-Path Bypass Protection 142
1.13.19 Configuring Multi-Path Route Import Between VRF Instances 147
1.13.20 Configuring BMP Monitoring 152
1.13.21 Configuring BGP ADD-PATH 154
1.13.22 Configuring BGP Neighbor Auto-Discovery 161

1 Configuring BGP

Note

The L2VPN EVPN address family mode, EVPN configuration mode, and EVPN-VNI configuration mode of BGP is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.1 Introduction

1.1.1 Overview

The Border Gateway Protocol (BGP) is a dynamic routing protocol that automatically exchanges loop-free routing information between autonomous systems (ASs).

A network is divided into multiple ASs to facilitate management. In the early Internet, the Exterior Gateway Protocol (EGP) was used to switch routing information between ASs. This protocol was first mentioned in RFC 187 and then listed into RFC 904 in 1984. EGP was a simple network reachability protocol and applied only to tree networks. Therefore, it could not meet the increasingly complicated network management requirements and finally was replaced by BGP. Unlike EGP, BGP can avoid loops and select better routes, and efficiently transfers routes and supports higher scalability.

The development of BGP has undergone different stages, and BGP-1 (RFC 1105), BGP-2 (RFC 1163), and BGP-3 (RFC 1267) were previously released. The prevailing version is BGP-4 (RFC 4271), which is widely applied between ASs as a standard routing protocol.

BGP has the following features:

- Interior Gateway Protocols (IGPs) such as Open Shortest Path First (OSPF) and Routing Information Protocol (RIP) focus on discovery and route calculation. By contrast, BGP lays emphasis on the control of route spreading and is applicable to networks that carry massive routes.
- BGP uses Transmission Control Protocol (TCP) as the transmission protocol and the destination port is 179. BGP uses the reliable transmission mechanism of TCP to ensure transmission reliability.
- As a distance vector routing protocol, BGP transmits only the locally selected optimal routes to its neighbors.

Note

In this chapter, the term "router" refers to any network device that supports the routing function. These network devices can be Layer-3 (L3) switches, routers, or firewalls.

1.1.2 Concepts

1. BGP speaker

Devices that run BGP and send BGP packets are called BGP speakers.

2. BGP peer

BGP speakers with a BGP session established are BGP peers. Multiple peers constitute a peer group.

3. AS

Devices that manage and share the same routing policies in one organization are an AS. When BGP is enabled on a router, a local AS number must be specified for it. An AS number is a globally unique number allocated by IANA, consisting of two or four bytes.

A traditional AS number consists of 2 bytes, ranging from 1 to 65535. The new 4-byte AS numbers aims to resolve the resource exhaust problem, and their value range is from 1 to 4294967295.

4-byte AS numbers support two expression modes: decimal notation mode and dot mode. The decimal notation mode is the same as the traditional expression mode, that is, the 4 bytes of an AS number are expressed in decimal notation. The conversion relationship between the dot mode and decimal notation mode is as follows: 4-byte decimal AS number = x × 65536 + Y.

For example:

- For an AS number 65534 in decimal mode, it is 65,534 in dot mode. An AS number smaller than 65536 is the same in decimal mode and dot mode.
- For an AS number 65536 in decimal mode, it is 1.0 in dot mode.
- For an AS number 65538 in decimal mode, it is 1.2 in dot mode.

With introduction of 4-byte AS numbers, BGP connections may be established between BGP speakers supporting only 2-byte AS numbers and those supporting 4-byte AS numbers. To communicate with a device that supports 4-byte AS numbers, reserve the AS number 23456 for filling on the device that supports only 2-byte AS numbers.

4. Neighbor type and route type

BGP neighbors are classified into two types:

- Internal BGP (IBGP) neighborship: The neighborship between BGP speakers within an AS is called IBGP neighborship. Routes learned from IBGP neighbors are called IBGP routes. IBGP completes transition of routing information within the AS.
- External BGP (EBGP) neighborship: The neighborship between BGP speakers in different ASs is called EBGP neighborship. Routes learned from EBGP neighbors are called EBGP routes. EBGP is used to exchange routing information between different ASs.

5. BGP message types

BGP runs by sending the following five types of messages:

- OPEN message: for establishing a connection between BGP peers.
- UPDATE message: for exchanging routing information between BGP peers.
- NOTIFICATION message: for closing a BGP connection.
- KEEPALIVE message: for keeping a BGP connection alive.
- ROUTE-REFRESH message: for requesting peers to re-send routing information after the routing strategy is changed. Only BGP-enabled devices that support the route refresh function will send and respond to this message.

The above messages are event-triggered except the KEEPALIVE message, which is sent periodically.

1.1.3 Route Attributes of BGP

When a BGP speaker advertises a route to its neighbors, the BGP speaker also advertises the attributes carried by the route. Common BGP attributes are as follows:

1. Origin attribute

Specifies the origin of a BGP route, which is set to IGP, EGP, or INCOMPLETE.

- IGP: The value is 0, specifying the internal routing information of the originating AS.
- EGP: The value is 1, specifying the routing information obtained from EGP.
- INCOMPLETE: The value is 2, specifying the routing information learned through other means.

2. AS_PATH attribute

Lists the ASs passed by a route reversely. The last AS is placed at the beginning of the list.

The AS_PATH attribute is used for:

- When a router receives routing information with the local AS number in the AS_PATH, the router considers that this route has loop and discards the packet.
- Routing information distribution is controlled by using the AS_PATH-based filtering rules. These filtering rules use regular expressions to parse AS paths.
- According to RFC 1771, BGP does not consider the AS path length when selecting an optimal path. Generally, a shorter AS path means a higher path priority; therefore, devices consider the AS path length when selecting the optimal path. Based on the actual condition, you can determine whether to consider the AS path length when selecting the optimal path.

3. NEXT_HOP attribute

Specifies the IP address of the next hop to be reached by a BGP route.

The change in the NEXT_HOP attribute complies with the following rules:

- When the route forwarded to an IBGP peer is non-local, the device does not change the NEXT_HOP attribute.
- When the route forwarded to an IBGP peer is local, the local interface address that the device uses to establish a BGP neighbor relationship with the peer is the next hop address.
- When the route is forwarded to an EBGP peer, the local interface address that the device uses to establish a BGP neighbor relationship with the peer is the next hop address.

4. MULTI_EXIT_DISC attribute

Distinguishes multiple output or input points for reaching the same neighbor AS. A smaller value means a higher priority. The MULTI_EXIT_DISC attribute is short for MED.

BGP uses the MED value as the basis for comparing priorities of paths learned from EBGP peers. A smaller MED value means a higher path priority.

- The **MED** value is compared only for paths of peers from the same AS during the optimal path selection by default.
- The **MED** value is not compared for paths of peers from other member ASs in an AS alliance during the optimal path selection by default.

- If a path not configured with the MED attribute is received, the MED value of this path is **0** by default. Since a smaller MED value means a higher path priority, the MED value of this path has the highest priority.
- The **MED** value is not compared for paths from different ASs during the optimal path selection by default. Instead, the receiving sequence of the paths is compared.

5. LOCAL_PREF attribute

Distinguishes the priorities of IBGP routes in an AS. A larger value means a higher priority.

When sending routes received from EBGP peers to IBGP peers, a BGP speaker adds the LOCAL_PREF attribute. BGP uses the LOCAL_PREF value as the basis for comparing priorities of paths learned from IBGP peers. A larger LOCAL_PREF value means a higher path priority.

6. ATOMIC_AGGREGATE attribute

This attribute informs the receiver that this route is aggregated.

When the BGP Speaker receives two overlay routes, and the IP address contained in one of them is a subset of the other, the BGP Speaker usually prefers a more-specific route. However, if the BGP Speaker propagates a less-specific route, the propagated route must carry the ATOMIC-AGGREGATE attribute. If a route with the ATOMIC-AGGREGATE attribute is received, it indicates that a more-specific route is lost during propagation.

During route aggregation, the ATOMIC-AGGREGATE attribute will be added to the aggregated route.

7. AGGREGATOR attribute

As a supplement to the ATOMIC_ AGGREGATE attribute, the AGGREGATOR attribute contains the AS number that initiated the route aggregation, and the IP address of the BGP speaker that generated the aggregation route.

During route aggregation, for aggregated route, the AGGREGATOR attribute will be added when the ATOMIC-AGGREGATE attribute is added.

8. COMMUNITY attribute

The COMMUNITY attribute is another method of controlling the distribution of routing information.

A community is a set of destination addresses. The COMMUNITY attribute is intended to facilitate execution of a community-based routing policy so as to simplify the configuration of routing information distribution control on BGP speakers. By using the community attribute, you can control the receiving and distribution of routing information. A destination address can belong to multiple communities. An AS administrator can define the communities, to which a destination address belongs.

BGP pre-defines four common community attribute values:

- **Internet**: Indicates the Internet community. All paths belong to this community.
- **No-export**: Specifies that the path is not advertised to EBGP peers.
- **No-advertise**: Specifies that the path is not advertised to any BGP peer.
- **Local-as**: Indicates that a path is not advertised outside the local AS. When an alliance is configured, the path is not advertised to other ASs or member ASs.

By default, all destination addresses belong to the Internet community and are carried in the COMMUNITY attribute of paths.

BGP speakers can set, add, or modify the community attribute value when learning, advertising, or re-distributing routes. An aggregate path will contain the community attribute values of all aggregated paths.

9. AIGP metric attribute

The AIGP metric attribute can accumulate the IGP metrics of paths that a route passes through. Preferred route selection based on the AIGP attribute value can reflect the path cost more accurately. A smaller AIGP attribute value indicates a higher route priority. Routes carrying the AIGP attribute have a higher priority than those without the AIGP attribute.

If the next hop of a route carrying the AIGP attribute is changed to self or another value during route transmission, the IGP metric value of the route to the original next-hop address needs to be added to the AIGP attribute value.

1.1.4 Route Exchange and Route Selection Policy of BGP

1. Route exchange

By default, the following rules are followed when routing information is exchanged between BGP neighbors:

- BGP receives routing information from all neighbors.
- If the sender of BGP routing information is an IBGP neighbor, BGP will send the routing information to the EBGP neighbor.
- If the sender of BGP routing information is an EBGP neighbor, BGP will send the routing information to all IBGP and EBGP neighbors.
- BGP sends only BGP routes that need to be updated when updating routing information.
- If there are multiple valid routes on the device to the same destination address, BGP will only advertise the optimal route.

BGP cannot automatically discover or learn accessible networks. The accessible network information of a local AS must be imported to BGP so that BGP can advertise this information to neighbors.

Two methods can be used to import network information of the local AS to BGP:

- Manual configuration: Import accessible network information within a specified range to BGP.
- Configuring route re-distribution: Configure route re-distribution to re-distribute accessible network information of IGP to BGP.

BGP offers a robust route management function. By configuring a route exchange policy for a BGP peer, you can actively control routes to be received by and to be advertised to this peer.

Send routing information of other ASs exchanged by BGP to the routing table of a device so that the device can forward packets of other ASs. Import routing information of other ASs exchanged by BGP to IGP so that IGP distributes the information to other IGP devices in the local AS and these devices can forward packets of other ASs.

Generally, BGP speakers serving as IBGP neighbors of each other are not directly connected physically. IGP devices in the link between the BGP speakers may fail to learn routing information same as that on the BGP speakers. When a BGP speaker at the border of an AS forwards packets received from other domains to the next-hop IBGP neighbor, the packets pass an IGP device in the middle. In this case, the packets may be lost due to no routing information on the IGP device. To synchronize BGP with IGP, you must ensure that all routers within an AS learn routing information to be sent to another AS before the routing information is advertised to this AS. BGP and IGP should be synchronized in most cases, except that:

- Routing information passing through this AS is not configured. For example, the AS is a stub AS.
- All routers within the AS run BGP. Full mesh is established among all BGP speakers (neighborship is established between each two BGP speakers).

2. BGP route update group

The BGP route update group function is a technology used to enhance the performance for advertising routes to neighbors.

The BGP route update group technology automatically classifies neighbors using the same policy into the same update group. When sending a route to a neighbor, the device encapsulates the update packet based on the update group, and sends the update packet to all neighbors in the update group. In this way, the update packet is encapsulated only once but sent multiple times, improving the performance of route advertisement to neighbors.

3. Route selection policy

BGP policies for selecting the optimal route are:

(1) Invalid routing entries (entries with the next hop unreachable and flapping entries) are not used for optimal route selection.

(2) Ethernet Virtual Private Network (EVPN) route: When routes share the same route distinguisher (RD), routes with the route target (RT) matched take priority over those with RT filtered.

(3) EVPN routes: A static MAC address route is preferred.

(4) EVPN routes: An anycast MAC route is prior to a local route.

(5) EVPN routes: The sequence numbers of routes from the same VXLAN tunnel end point (VTEP) are compared and a route with a larger sequence number is preferred.

(6) If not, the route generated by the **network-advertise** command is preferred.

(7) Otherwise, a route with the maximum weight is preferred (the default weight value of a route originated locally is **32768**, and the weight value of a router advertised by a neighbor is **0**).

(8) If routes have the same weight, a route with a larger **LOCAL_PREF** value is preferred.

(9) If routes have the same local priority, a local BGP route from the local router is preferred.

(10) The types of local routes are compared, and the route priority is as follows: direct route > static route > aggregate route generated by BGP.

(11) If the local routes have the same type, the Accumulated Interior Gateway Protocol (AIGP) extended route is compared. A route that carries the AIGP attribute is prior to that without the AIGP attribute.

(12) If the AIGP attribute comparison is skipped, the link state (LS) extended attribute is compared:

a A route with the LS attribute is preferred.

b A route with a smaller IGP metric is preferred.

c A route of the wide type has a higher priority than that of the narrow type.

(13) Otherwise, a route with the shortest AS length is selected.

(14) Otherwise, a route with the lowest **ORIGIN Code** attribute value is selected.

(15) Otherwise, a route with the smallest **MED** value is selected.

(16) Otherwise, EBGP routes have a higher priority than IBGP routes and routes in an AS alliance, and the IBGP routes have the same priority as the routes in an AS alliance.

(17) Otherwise, a route with the smallest IGP metric value to the next hop is selected.

(18) Otherwise, an EBGP route that is received first is selected.

(19) Otherwise, a route advertised by a BGP speaker with a smaller router-ID is selected.

(20) Otherwise, a route with a smaller cluster length is selected.

(21) Otherwise, a route with a smaller neighbor address is selected.

1.1.5 Neighbor Connection Establishment

A BGP neighbor is manually configured by a user. Neighbor relationships can be established in two modes: IBGP and EBGP. Users can identify the connection mode between BGP speakers according to the AS carrying the BGP peer and the AS carrying the BGP speaker.

A BGP speaker proactively sends a TCP connection request to a specified BGP peer. After a TCP connection is established successfully, BGP speakers exchange BGP protocol packets to negotiate about connection parameters. The BGP neighbor relationship is successfully established after the negotiation succeeds.

1. Establishing a TCP connection

A BGP speaker initiates a TCP connection request to a neighbor. The destination IP address of the request is the peer IP address specified by the user and the port number is fixed to 179.

The BGP speaker also listens on port 179 of the local TCP connection to receive connection requests from its peers.

2. Negotiating about protocol parameters

After a TCP connection is successfully established, BGP speakers exchange OPEN packets to negotiate about BGP connection parameters. Negotiated parameters include the following:

- **Version**: Indicates the BGP version number. At present, only version 4 is supported.
- **Neighbor AS number**: Determines whether the AS number of a neighbor is consistent with the AS number specified locally. If no, the connection request is denied.
- **Hold time**: Indicates the timeout duration for a BGP connection. The default value is **180** seconds.
- **Neighbor capability**: Negotiates about various extended capabilities supported by the neighbor, including the address family, dynamic route update and GR functions.

3. Maintaining the neighbor relationship

BGP speakers send the Keepalive message to each other periodically. If a new Keepalive packet is not received from a BGP neighbor after the hold time expires, the BGP speaker considers the neighbor unreachable, disconnects the TCP connection from the neighbor, and attempts to reconnect to it. The interval for a BGP speaker to send the Keepalive message is one third of the negotiated hold time and is 60 seconds by default.

The system allows you to manually configure various timers within BGP to meet the neighbor keepalive and route management requirements in different network environments.

- BGP neighbor keepalive timer

  BGP uses the keepalive timer to maintain a valid connection with a peer and the holdtime timer to judge whether the peer is valid. The default value of the keepalive timer is **60** seconds and the default value of the holdtime timer is **180** seconds. When BGP speakers establish a BGP connection, both parties negotiate about the holdtime and the smaller holdtime value is selected. The keepalive timer is set to one third of the negotiated holdtime value or the configured keepalive value, whichever is smaller.

- Neighbor reconnection timer

To reduce the impact of frequent BGP reconnection to a failed neighbor on the network bandwidth, after a BGP speaker detects a neighbor connection failure, it attempts to reconnect to the neighbor after the Connect-Retry timer expires. The default value of the Connect-Retry timer is **15** seconds.

- Route advertisement timer

  To reduce the impact of route update packets on the network bandwidth, after a BGP speaker detects a network topology change, it does not advertise the route update to its neighbors immediately. Instead, the BGP speaker adopts a regular update mechanism to advertise all changed routing information to its neighbors.

1.1.6 Multi-Path Load Balancing of BGP

Multi-path load balancing means that there are multiple paths to the same network and data packets are evenly forwarded by these paths. In this case, one route has multiple next hops in the routing table.

According to the types of equal-cost routes, multi-path load balancing of BGP is classified into the following types:

- EBGP load balancing: Implements load balancing for routes learned from EBGP neighbors.
- IBGP load balancing: Implements load balancing for routes learned from IBGP neighbors.

When there are multiple paths to the same network in the routing table of BGP, BGP calculates a route with the highest priority by default. If there are multiple optimal routes with the same priority, BGP still selects a unique route according to comparison rules, advertises the route to the forwarding plane, and controls the forwarding of data streams. After multi-path load balancing is enabled, in addition to calculating a unique optimal route, BGP lists paths with the same priority as the optimal route as equal-cost routes. Then, BGP advertises the optimal route and equal-cost routes to the forwarding plane to implement load balancing.

Equal-cost routes have the same basic attributes and priority. That is, according to the optimal path selection rules of BGP, paths have the same priorities before router-IDs are compared.

- AS_PATH loose comparison

  Equal-cost routes must have the same AS-PATH attribute by default. Under such strict conditions, load balancing cannot be implemented in certain environments. In this case, you are advised to enable the AS-PATH loose comparison mode. In AS-PATH loose comparison mode, when other conditions for equal-cost routes are met, equal-cost route conditions are met as long as the AS-PATH lengths of routes and the AS-PATH lengths of the member ASs are equal respectively.

- Router-ID multi-path comparison

  Equal-cost routes do not need to come from the same device (that is, the sources of the routes do not need to have the same router ID) by default. After the router-ID multi-path comparison is enabled, only routes with the same router-ID can be equal-cost routes.

- Non-equal-cost load balancing

  Load-balanced routes are equal-cost routes by default. After the non-equal-cost load balancing is enabled, you can configure routes to carry the Link-Bandwidth attribute so that non-equal-cost routes form.

1.1.7 Route Reflector

Route reflectors are a method of reducing IBGP peer connections in an AS.

According to the principle of BGP route advertisement, all BGP speakers in an AS must establish a full mesh of connections (every two BGP speakers need to establish a neighbor relationship). Too many BGP speakers in

an AS will increase their resource cost, ramp up network administrators' configuration workload and complexity, and reduce the network expansion capacity.

You can configure a BGP speaker as a route reflector, which classifies IBGP peers in the AS into two types: clients and non-clients.

The rules for implementing a route reflector in an AS are as follows:

- Configure a route reflector and specify clients for the route reflector. The route reflector and its clients form a cluster. The route reflector will connect to its clients.
- The clients of a route reflector in a cluster cannot connect to BGP speakers outside the cluster.
- Within an AS, a full mesh of connections is established among IBGP peers of non-clients. The IBGP peers of non-clients include the following situations: Multiple route reflectors in a cluster; a route reflector in a cluster and BGP speakers (generally not supporting the route reflector function) not involved in the route reflector function out of the cluster; a route reflector in a cluster and route reflectors in other clusters.

The rules for processing a route received by a route reflector are as follows:

- A route update message received from an EBGP speaker will be sent to all clients and non-clients.
- A route update message received from a client will be sent to other clients and all non-clients.
- A route update message received from an IBGP non-client will be sent to all clients of the route reflector.

1.1.8 BGP Alliance

Alliances are another method of reducing IBGP peer connections in an AS.

You can divide an AS into multiple member ASs and configure a unified alliance ID (namely, alliance AS number) for these member ASs to form an alliance. Outside the alliance, the entire alliance is still considered as an AS and only the AS number of the alliance is visible. Inside the alliance, BGP speakers in a member AS still establish a full mesh of IBGP peer connections, BGP speakers in different member ASs establish EBGP connections. Though EBGP connections are established between BGP speakers in different member ASs, the NEXT_HOP, MED, LOCAL_PREF, and other path attributes keep unchanged during information exchange.

1.1.9 Route Aggregation

The great BGP routing table in a large network brings heavy load to BGP devices. BGP supports classless inter-domain routing (CIDR). Therefore, route aggregation entries can be created to reduce the size of the BGP routing table. BGP aggregation entries can be added to the BGP routing table only when valid paths are configured within the aggregation range. Routes can be manually aggregated on a device. Thus, this function can control whether to reserve AS path information within the aggregation address range, and whether to advertise only aggregated paths.

1.1.10 Route Dampening

If a route changes between the valid state and invalid state, route flapping occurs. Route flapping often causes the transmission of unstable routes in a network, and thereby causes network instability. BGP route dampening is a method of reducing route flapping. It reduces possible route flapping by monitoring routing information from EBGP peers.

Terms involved in BGP route dampening are as follows:

- **Route Flap**: A route changes between the valid sate and invalid state.
- **Penalty**: Once route flapping occurs, a BGP speaker, on which route dampening is enabled, adds a penalty

value to this route. The penalty value accumulates until **Suppress Limit** is reached.

- **Suppress Limit**: When the penalty of a route is greater than this value, the route is suppressed.
- **Half-life-time**: It indicates the time required for the penalty to be halved.
- **Reuse Limit**: When the penalty value of a route is smaller than this value, route suppression is canceled.
- **Max-suppress-time**(Upper limit of penalty time): The maximum penalty value depends on the configured upper limit of penalty time. The penalty value will not increase after reaching the upper limit. Formula: Maximum penalty value = 2 ^ (Max-suppress-time/Half-life-time) x Reuse Limit, and the maximum penalty value cannot be greater than **20000**.

A BGP speaker assigns a route a penalty (accumulated to the penalty value) each time the route flaps. When the penalty value reaches Suppress Limit, the route is suppressed. The penalty does not increase after it reaches a certain value. This value is called the upper limit of penalty. When the penalty value reaches Half-life-time, the penalty is halved. When the penalty value is reduced to Reuse Limit, the route is activated again.

By default, the penalty value increases by 1000 when a route is reactivated and increases by 500 when a route is updated.

1.1.11 BGP Reliability

1. GR

Graceful restart (GR) is intended to ensure that data forwarding is not interrupted during the BGP restart. The GR function keeps the network topology stable, maintains the forwarding table, and ensures that key services are not interrupted during the switching of master and slave supervisor engines.

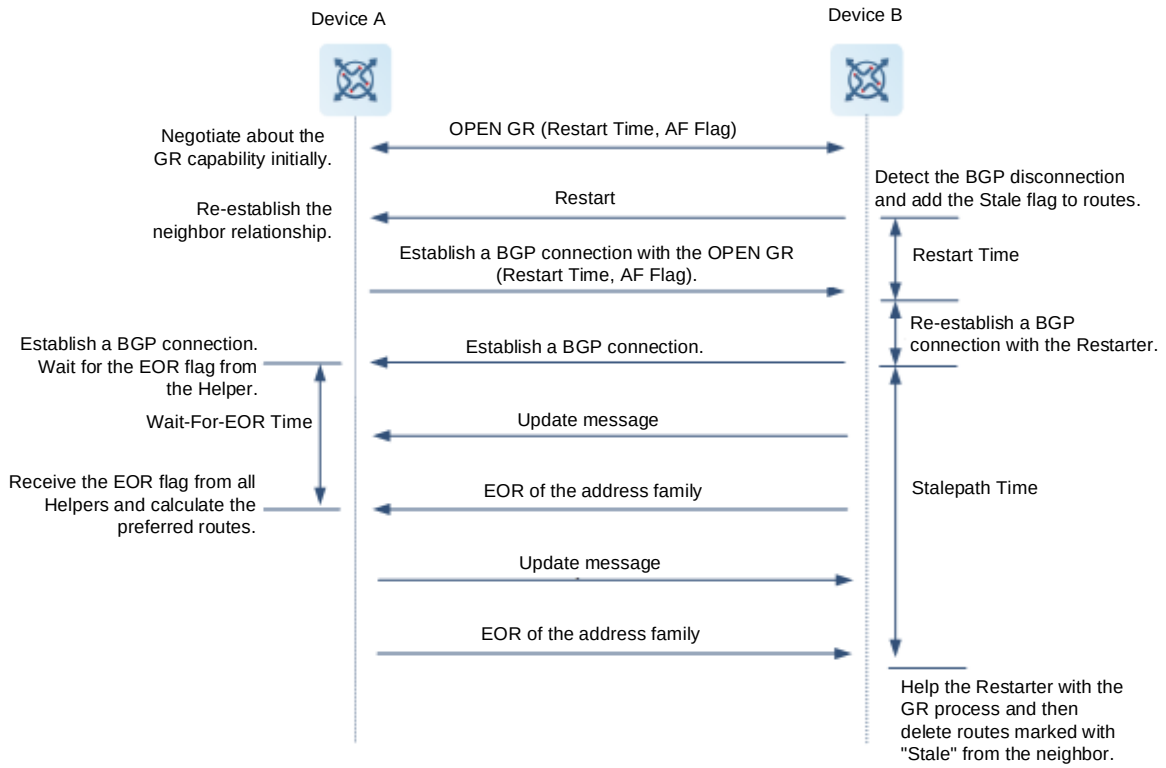
BGP GR is not an independent process, but is jointly accomplished by the Restarter and Helper.

- The Restarter performs restart and is capable of making the route forwarding plane to keep working when the route control plane is faulty.
- The Helper is the BGP neighbor of the Restarter and helps the Restarter complete GR.

To deliver graceful restart, BGP adds a capability field indicating to neighbors its support for GR to the OPEN message. The capability is called Graceful Restart Capability. During initialization of a BGP connection, both neighbors negotiate about the GR capability.

The End-of-RIB (EOR) flag is added to the update packets of BGP, to indicate that the routing information update to the neighbor is completed.

Figure 1-1 BGP GR Interaction Process



(1) When BGP establishes a neighbor relationship at the beginning, BGP uses the GR capability field in the OPEN message to negotiate about the GR capabilities of both neighbors.

(2) At a moment, the Restarter performs restart, and the BGP session is disconnected. The Helper detects the disconnection, keeps the route of the Restarter valid, but marks the Stale (aged but not updated).

(3) The Restarter completes restart.

(4) The Restarter connects to the Helper again.

(5) The Restarter waits for the route update message and EOR flag from the Helper.

(6) After receiving the EOR flag from all neighbors, the Restarter performs route calculation, updates routing entries, and then sends updated routes to the Helper.

(7) After receiving the updated routes, the Helper cancels their Stale flag. After receiving the EOR flag from the Restarter, the Helper deletes routes with the Stale flag (these routes are not updated), calculates routes calculation, and updates the routing entries. The entire GR process is completed.

BGP GR defines several important extended timers:

- **Restart-Timer**

  The value of this timer is advertised by the GR Restarter to the GR Helper. It indicates the maximum waiting time that the GR Restarter hopes the Helper to wait before a new connection is established between them. You can run the **bgp-instance graceful-restart restart-time** command to change the time value.

- **Wait-For-EOR Timer**

This timer indicates the maximum time for the GR Restarter to wait for the EOR flag from all GR Helpers. After the EOR flag is received from all GR Helpers or the Wait-For-EOR timer expires, the GR Restarter calculates the preferred routes and updates the routing entries. You can run the **bgp-instance advertise-delay** command to change the time value.

- **StalePath Timer**

  The timer indicates the maximum time for the GR-Helper to wait for the EOR flag from the GR Restarter after the connection between them is restored. Within this period, the Helper keeps the original route of the Restarter valid. After receiving the EOR flag or after the StalePath timer expires, the Helper clears the routing entries with the Stale tag. You can run the **bgp-instance graceful-restart staleroute-time** command to change the time value.

2. BGP NSR

None-stop-routing (NSR) is used to ensure uninterrupted routes during protocol restart upon a switchover between the master and slave supervisor engines. During the switchover, the NSR function keeps the network topology stable, maintains the neighbor status and forwarding table, and ensures that key services are not interrupted.

Unlike BGP GR, BGP NSR does not need the assistance of neighbors.

- Connections between BGP neighbors are not interrupted during active/standby switchover.
- Neighbors do not perceive the restart behavior of the peer device. The GR technology is used for restart restoration.

Neighbors with the BGP NSR function enabled do not need to negotiate about the GR capability nor need assistance from the peer. Therefore, no interworking problem arises.

3. Fast rerouting

With the fast development of IP technologies and the application of various complex services, higher requirements are posed for network security and stability. Some real-time services (audios and videos) are sensitive to the network running status and are largely affected by unstable networks. Therefore, network reliability receives more and more attention. In response to this requirement, the IP Fast Reroute (FRR) function appears. This function aims to use a backup link to forward data during route platform convergence after a faulty link is detected so as to ensure that intact packets are forwarded timely.

Thus, BGP FRR can avoid route disconnection due to a link fault before BGP route convergence. If a BGP routing table has multiple paths to the same network, BGP calculates the route with the highest priority by default. After the BGP FRR function is used, BGP selects a backup route for each optimal route. When detecting a fault occurring on the master link by using the BFD fast link detection mechanism, the device switches data to the selected backup link for data forwarding. After route convergence is completed, the device switches data to the re-calculated optimal path for data forwarding.

1.1.12 MP-BGP

Traditional BGP-4 only supports unicast IPv4 routes, and cannot propagate routing information of other network layer protocols (such as IPv6 unicast protocol). MP-BGP (Multi-Protocol BGP) is an extension of BGP-4, which can propagate the routing information of other network layer protocols without changing the original message mechanism and routing mechanism of the BGP protocol.

Compared with BGP, MP-BGP has two additional optional attributes:

- MP_REACH_NLRI (Multiprotocol Reachable NLRI): This attribute is used to carry the reachable route prefix

and the next-hop IP address information of multiple network layer protocols, in order to advertise the route to BGP neighbors.

- MP_UNREACH_NLRI (Multiprotocol Unreachable NLRI): This attribute is used to carry the unreachable route prefix information of multiple network layer protocols, in order to withdraw the route.

If the BGP Speaker does not support MP-BGP, these two attributes will be ignored after an UPDATE message with the above two attributes is received.

MP-BGP supports propagating the routing information for the following protocols:

1. EVPN

The EVPN is an extended layer-2 VPN technology based on the Multi-Protocol BGP (MP-BGP). The MP-BGP expands the EVPN address family to transfer the required layer-2 and layer-3 routing information.

MP-BGP adds not only the EVPN address family, but also the EVPN network layer reachability information, which includes the following:

- Ethernet Auto-discovery Route: Used to automatically discover Ethernet segment (ES) information in a multi-homed network.
- MAC/IP Advertisement Route: Used to advertise the MAC address and host ARP information.
- Inclusive Multicast Ethernet Tag (IMET) Route: Used to automatically discover VXLAN tunnel end points (VTEPs) and create VXLAN tunnels.
- Ethernet Segment Route: Used to advertise ES information and information about connected VTEPs.
- IP Prefix Route: Used to advertise routing information of IP network segments.

The preceding EVPN routing information includes the RD and RT information. The RD is used for EVPN routes of different VNIs and the RT is used for importing EVPN routes between different devices.

1.1.13 BMP

BGP Monitor Protocol (BMP) monitors BGP running on a device. It focuses on the status of BGP peers, peer entry transmission and receiving, and all kinds of BGP peer statistics.

A BGP speaker proactively sends a TCP connection request to a specified BMP server. After a TCP connection is established successfully, the BGP speaker advertises a BMP packet to the BMP server. In the TCP connection request initiated by a BGP speaker to a BMP server, the destination IP address is the IP address of the specified BMP server and the port number is a specified port number.

1.1.14 BGP Neighbor Auto-Discovery

BGP is based on the Transmission Control Protocol/Internet Protocol (TCP/IP) and requires IP addresses to establish BGP connections. In data center scenarios, a large number of IP addresses need to be assigned to devices. To save address resources without affecting BGP connections, you can deploy the BGP neighbor auto-discovery function. This function can save IP address resources, and supports neighbor auto-discovery and connection.

1.1.15 Others

1. Regular expression

A regular expression matches strings based on a rule. It is used to assess text data and return True or False to decide whether the expression can correctly describe the data.

The following table describes special characters involved in regular expressions used in BGP path attributes.

Table 1-1 Description of Special Characters in a Regular Expression

Character	Symbol	Special Meaning
Period	.	Matches any single character.
Asterisk	*	Matches zero or any sequence in a string.
Plus sign	+	Matches one or any sequence in a string.
Question mark	?	Matches zero or one symbol in a string.
Caret	^	Matches the start of a string.
Dollar sign	$	Matches the end of a string.
Underline	_	Matches commas, brackets, start and end of a string, and spaces.
Square brackets	[ ]	Matches a single character within a range.

1.1.16 Protocols and Standards

- RFC 1105: Border Gateway Protocol (BGP)
- RFC 1163: Border Gateway Protocol (BGP)
- RFC 1267: Border Gateway Protocol 3 (BGP-3)
- RFC 1771: A Border Gateway Protocol 4 (BGP-4)
- RFC 2918: Route Refresh Capability for BGP-4
- RFC 4273: Definitions of Managed Objects for BGP-4
- RFC 4360: PrBGP Extended Communities Attribute
- RFC 4364: Proposed Standard: BGP/MPLS IP Virtual Private Networks (VPNs)
- RFC 4486: Proposed Standard: Subcodes for BGP Cease Notification Message
- RFC 4724: Proposed Standard: Graceful Restart Mechanism for BGP
- RFC 4760: Draft Standard: Multiprotocol Extensions for BGP-4
- RFC 4893: BGP Support for Four-octet AS Number Space
- RFC 5396: Textual Representation of Autonomous System (AS) Numbers
- RFC 5492: Draft Standard: Capabilities Advertisement with BGP-4
- RFC 6198: Requirements for the Graceful Shutdown of BGP Sessions
- RFC 7313: Enhanced Route Refresh Capability for BGP-4
- RFC 7311: The Accumulated IGP Metric Attribute for BGP
- RFC 7432: Proposed Standard: BGP MPLS-based Ethernet VPN

1.2 Configuration Task Summary

BGP configuration includes the following tasks:

- Configuring BGP Basic Features
  - Configuring the BGP Version
  - Enabling BGP
  - Configuring a BGP Neighbor
  - (Optional) Configuring a BGP IPv6 Address Family
  - (Optional) Configuring a BGP Peer Group
  - (Optional) Configuring the Dot Mode to Display 4-Byte AS Numbers
  - (Optional) Injecting Network Information of Local AS into BGP
  - (Optional) Obtaining Reachable Networks of Other AS from BGP
  - (Optional) Configuring BGP Capacity Protection
  - (Optional) Shutting Down BGP Connections Gracefully
  - (Optional) Configuring BGP Soft Reset
  - (Optional) Configuring Scope Configuration Mode
  - (Optional) Configuring BGP Route Update Mechanisms
  - (Optional) Configuring BGP Timers
- (Optional) Configuring a Large-Scale BGP Network
  All the following configuration tasks are optional and may be selected as needed.
  - Configuring BGP Route Reflector
  - Configuring BGP AS Alliance
  - Configuring Route Aggregation
  - Configuring Route Dampening
- (Optional) Configuring BGP Route Selection and Load Balancing
  All the following configuration tasks are optional and may be selected as needed.
  - Configuring BGP Multi-Path Load Balancing
  - Configuring BGP Routes to be Recursive Only to Host Routes
  - Configuring Outbound Loop Detection for a BGP Neighborr
  - Configuring BGP ADD-PATH
  - Configuring BGP Advertising Routes with the Lowest Priority upon Device Restart
  - Configuring AS_PATH Attribute
  - Configuring MED
  - Configuring AIGP
  - Configure LOCAL_PREF
  - Configure Extended Community Attribute
  - Configuring BGP Best Path Selection
- (Optional) Configuring BGP Controlling Route Advertisement and Receiving

All the following configuration tasks are optional and may be selected as needed.

- Configuring BGP Filters
- Configuring BGP Route Sending Interval and Count
- Configuring BGP Modifying Propagated Routing Information
- Configuring BGP Processing Received Routing Information
- Configuring BGP Fast Withdrawing Specific Route
- Configuring BGP Delaying Advertising Routing Information

- (Optional) Configuring the BGP Security Features

  All the following configuration tasks are optional and may be selected as needed.
  - Configuring MD5 Authentication
  - Configuring BGP Peer GTSM Security Check

- (Optional) Configuring BGP Reliability

  All the following configuration tasks are optional and may be selected as needed.
  - Configuring BFD for BGP
  - Configuring BGP FRR
  - Configuring EBGP Multi-Path Bypass Protection
  - Configuring BGP GR
  - Configuring BGP NSR
  - Configuring BGP Session Keepalive
  - Configuring EBGP Fast Neighbor Detection
  - Configuring BGP Link-State Group Association

- (Optional) Configuring MP-BGP
  - Configuring an EVI
  - Configuring EVPN

- _(Optional) Configuring BGP Neighbor Auto-Discovery

  All the following configuration tasks are optional and may be selected as needed.
  - Configuring BGP Neighbor Auto-Discovery
  - Configuring Interface Neighbor Auto-Discovery

- (Optional)Configuring BGP Extended Features
  - Configuring BMP Monitoring
  - Configuring BGP Administrative Distance
  - Configuring Enhanced VPN Route Import
  - Configuring BGP Link-State

1.3 Configuring BGP Basic Features

1.3.1 Configuration Tasks

The configuration of BGP basic features includes the following tasks:

- Configuring the BGP Version
- Enabling BGP
- Configuring a BGP Neighbor
- (Optional) Configuring a BGP IPv6 Address Family
- (Optional) Configuring a BGP Peer Group
- (Optional) Configuring the Dot Mode to Display 4-Byte AS Numbers
- (Optional) Injecting Network Information of Local AS into BGP
- (Optional) Obtaining Reachable Networks of Other AS from BGP
- (Optional) Configuring BGP Capacity Protection
- (Optional) Shutting Down BGP Connections Gracefully
- (Optional) Configuring BGP Soft Reset
- (Optional) Configuring Scope Configuration Mode
- (Optional) Configuring BGP Route Update Mechanisms
- (Optional) Configuring BGP Timers

1.3.2 Configuring the BGP Version

1. Overview

After the TCP session is set up, the BGP Speaker will exchange the OPEN packets to negotiate BGP connection parameters, including the BGP version. Only BGP 4 is supported currently.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the BGP version number for a specified BGP peer.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **version 4**

No BGP version number is configured for a specified BGP peer by default.

1.3.3 Enabling BGP

1. Overview

Before configuring the BGP function, you need to enable the BGP protocol, and configure the AS number.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Configure the device ID of the device running the BGP protocol.

bgp-instance router-id *ipv4-address*

By default, the loopback interface of the device is preferred as the device ID. If no loopback interface is available, the Router-ID of the device is used.

(5) (Optional) Configure the default address family as IPv4 unicast address family.

bgp-instance default-config ipv4-unicast

By default, the default address family is the IPv4 unicast address family.

(6) (Optional) Configure the device to record the BGP status change information without the Debug function.

bgp-instance log-peer-changes

By default, the device records the BGP status change information without the Debug function.

(7) (Optional) Disable the TCP source IP address check function of BGP.

bgp-instance tcp-source-check disable

By default, the TCP source IP address check function is enabled.

1.3.4 Configuring a BGP Neighbor

1. Overview

This section describes how to configure a BGP speaker as a neighbor.

2. Restrictions and Guidelines

- If IBGP neighbors are not directly connected, you need to configure IGP or a static routing protocol to implement reachable routes between the neighbors.
- If EBGP neighbors are not directly connected, you need to configure the **ebgp-multihop** parameter for the neighbors.
- The source interface for establishing a neighbor relationship must be a valid local interface or IP address.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Configure the alarm function when the maximum number of BGP neighbors is exceed.

bgp-instance maximum-peer *maximum-neighbor-numbers* **warning**

By default, the alarm function the maximum number of BGP neighbors is exceed is not configured.

(5) (Optional) Configure the maximum number of sessions that can be established by a BGP peer.

bgp-instance network-peer-number *maximum-session-number*

By default, no maximum number of sessions that can be established by a BGP peer is configured.

(6) (Optional) Enable the BGP peer address family capability detection function.

bgp-instance notify unsupport-capability

By default, the BGP peer address family capability detection function is disabled.

(7) (Optional) Save original routing information sent by all BGP peers under the current BGP instance.

bgp-instance soft-resave inbound

No original routing information sent by all BGP peers under the current BGP instance is saved by default.

(8) (Optional) Set the connect-retry value used during BGP connection establishment.

bgp-instance timer connect-retry *connect-retry*

The connect-retry value used during BGP connection establishment is not configured by default.

(9) (Optional) Use RT attributes advertised by neighbors when RT is not configured for VRF.

bgp-instance keep vrf-route-target

By default, RT attributes advertised by neighbors are not used when RT is not configured for VRF.

(10) (Optional) Enter a required address family configuration mode. If you need to perform configuration for a specific address family, select one of the following tasks to configure.

- Enter the IPv4 unicast configuration mode.

  address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

- Enter the L2VPN configuration mode.

  address-family-mode l2vpn evpn

  BGP neighbors are not allowed to exchange L2VPN information by default.

- Enter the link-state configuration mode.

  address-family-mode link-state [**unicast**]

(11) Create a BGP neighbor. When the configured AS number is the same as the local AS number, an IBGP neighbor is created. When the configured AS number is different from the local AS number, an EBGP neighbor is created.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

No neighbor is specified for a BGP speaker by default.

(12) (Optional) Configure the description of the specified peer (group).

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **description** *text*

By default, the description of the specified peer (group) is not configured.

(13) (Optional) Configure a source interface or IP address for establishing a TCP connection with the neighbor. A loopback interface is recommended for IBGP neighbors while a directly connected interface is recommended for EBGP neighbors.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **update-source** { *interface-type interface-number* | *address* }

BGP automatically selects the source IP address for a TCP connection based on the IP address of a neighbor by default. The IP address of the outbound interface of local packets is generally used.

(14) (Optional) Configure the time to live (TTL) for TCL packets received from the neighbor. A larger TTL value indicates more hops to the neighbor. When the TTL is **1**, the BGP neighbor must be directly connected to the device.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **ebgp-multihop** [*ttl*]

The default TTL values of TCP packets to be sent to IBGP neighbors and EBGP neighbors are **255** and **1** respectively.

(15) (Optional) Configure the mode for BGP to establish a TCP connection with the neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **transport connection-mode** { **active-only** | **both** | **passive-only** }

By default, BGP neighbors support the active and passive modes of establishing TCP connections.

(16) (Optional) Activate the address family capability for the neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **activate**

Only the IPv4 unicast address family capability is activated for neighbors created in BGP configuration mode by default.

(17) (Optional) Configure the maximum number of sessions that can be established by a BGP peer.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* } **session-number** *maximum-session-number*

No maximum number of sessions that can be established by a BGP peer is configured by default.

(18) (Optional) Configure the DSCP value for sending BGP packets.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **dscp** [*dscp*]

The default DSCP value is **48**.

1.3.5 Configuring a BGP IPv6 Address Family

1. Overview

This section describes how to configure BGP to exchange IPv6 routes to implement IPv6 network access between different ASs.

2. Restrictions and Guidelines

- Generally, BGP uses IPv6 addresses to establish neighbor relationships and implement the exchange of IPv6 routes.
- When IPv6 link local address to establish peer connection with a BGP peer, the address is expressed as Linklocal address% interface name, for example, FE80:: 1% Gigabit Ethernet 0/1.
- In special scenarios, for example, the IPv6 provider edge router (6PE) function is configured, BGP supports

the exchange of IPv6 routes with the neighbors using IPv4 addresses.

- Configurations related to BGP IPv6 services must be completed in BGP IPv6 address family mode.
- Neighbors using IPv6 addresses are used to exchange IPv6 routes. When a neighbor is configured in BGP mode, BGP automatically activates the IPv4 unicast address family capability for the neighbor. You are advised to manually disable the IPv4 unicast address family capability.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(5) Configure the BGP IPv4 address family mode.

address-family-mode ipv4 unicast

(6) Disable the IPv4 address family capability for the BGP neighbor.

unset peer { *neighbor-ipv4-address* | *peer-group-name* } **activate**

- The IPv4 address family capability is enabled by default.

(7) Configure the BGP IPv6 address family mode.

address-family-mode ipv6 unicast

(8) Configure the IPv6 address family capability for the BGP neighbor.

peer { *neighbor-ipv6-address* | *peer-group-name* } **activate**

The IPv6 address family capability is disabled by default.

(9) (Optional) Configure BGP to advertise IPv6 routes.

network-advertise *network-ipv6-address/prefix-length* [**route-map** *route-map-name*] [**backdoor-route**]

No network routing information is configured for BGP by default.

(10) (Optional) Configure to use the link local address as the global next-hop IP address with sending IPv6 routing information to the BGP IPv6 link local peer (group).

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **global-nexthop-replace-local**

By default, when sending IPv6 routing information to the BGP IPv6 link-local address peer (group), the global next-hop IP address is not the link-local address.

1.3.6 Configuring a BGP Peer Group

1. Overview

BGP speakers in the same peer group have the same update strategy. Peer groups can simplify configuration and improve the running efficiency.

2. Restrictions and Guidelines

- Members of a peer group inherit all configurations of the peer group. Unified configuration of a peer group can be replaced by separate configuration of each member in the peer group, except the configuration that influences output update, for example, the configuration of **peer route-filtering**, **peer route-map**, and **peer filter-rule-list**.
- You can neither put neighbors belonging to different address families to the same peer group nor put IBGP and EBGP neighbors to the same peer group.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Enter a required address family configuration mode. If you need to perform configuration for a specific address family, select one of the following tasks to configure.

- Enter the IPv4 unicast configuration mode.

  address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

- Enter the L2VPN configuration mode.

  address-family-mode l2vpn evpn

  BGP neighbors are not allowed to exchange L2VPN information by default.

- Enter the link-state configuration mode.

  address-family-mode link-state [**unicast**]

(5) Create a peer group.

peer *peer-group-name* **peer**-**group**

(6) Specify a BGP speaker as a member of the BGP peer group.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* } **peer-group** *peer-group-name*

1.3.7 Configuring the Dot Mode to Display 4-Byte AS Numbers

1. Overview

The 4-byte AS numbers can be represented in dot mode.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the dot mode to display 4-byte BGP AS numbers.

bgp-instance as-symbol dot

1.3.8 Injecting Network Information of Local AS into BGP

1. Overview

BGP will not automatically discover or learn reachable networks. The information of reachable networks in local AS must be injected into BGP, so that BGP can propagate the information to BGP peers.

There are two ways to inject network information in local AS into BGP:

- Manual static configuration: inject the information of reachable networks within the specified range into BGP.
- Configure route redistribution: inject the information of reachable networks of IGP protocol into BGP.

2. Restrictions and Guidelines

- You are advised to enable the network information synchronization function. Otherwise, routing black holes may be incurred.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure to redistribute the routing information of other routing protocols to all BGP instances.

router-protocol bgp redistribute-route [**vrf** *vrf-name*] { **ipv4 unicast-af-modifier** | **ipv6 unicast-af-modifier** } *protocol-type* [**metric-value** *metric-value* | **route-map** *map-tag*] *

By default, redistributing the routing information of other routing protocols to all BGP instances is not configured.

(4) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(5) (Optional) Configure the current instance to not apply non-specified instance redistribution.

bgp-instance redistribute-route unspecified instance ignore

By default, a BGP instance applies non-specified instance redistribution.

(6) (Optional) Enter a required address family configuration mode. If you need to perform configuration for a specific address family, select one of the following tasks to configure.

- Enter the IPv4 unicast configuration mode.

address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

(7) Add static routing entries to a BGP routing table and advertise the routing entries to peers.

network-advertise *network-number* [**mask** *mask*] [**route-map** *route-map-name*] [**backdoor-route**]

No static routing entry is added to a BGP routing table and advertised to peers by default.

(8) (Optional) Configure a device to advertise routing information configured by running the **network-advertise** command after the device synchronizes with the local route.

network-advertise synchronization

By default, a device will advertise routing information configured by running the **network-advertise** command after the device synchronizes with the local route.

(9) Configure route redistribution. All the following configuration commands are optional and can be configured as needed.

(10) Configure to redistribute the routing information of other routing protocols to the BGP instance.

redistribute-route *protocol-type* [**metric-value** *metric-value* | **route-map** *map-tag*] *

By default, redistributing routing information of other routing protocols to a BGP instance is not configured.

(11) Configure to redistribute the routing information of the IS-IS routing protocol to the BGP instance.

redistribute-route isis [*isis-tag*] [{ **level-1** | **level-1-2** | **level-2** } | **metric-value** *metric-value* | **route-map** *map-tag*] *

By default, redistributing routing information of IS-IS routing protocol to a BGP instance is not configured.

(12) Configure to redistribute the routing information of OSPF routing protocol to the BGP instance.

redistribute-route ospf *process-id* [**filter** { { **external-type** | { **external-type 1** | **external-type 2** }* } | **internal-type** | { **nssa-external-type** | { **nssa-external-type 1** | **nssa-external-type 2** } * } } * | **metric-value** *metric-value* | **route-map** *map-tag*] *

By default, redistributing routing information of OSPF routing protocol to a BGP instance is not configured.

(13) Configure to enable to inject multiple next-hop routes of other protocols into the BGP protocol.

bgp-instance src-paths *protocol-type* **all**

By default, injecting multiple next-hop routes is disabled.

(14) (Optional) Configure to use the manually configured metric value when BGP redistributes a route.

default-metric *metric-value*

By default, the metric value of the route is used when the route is redistributed by BGP.

1.3.9 Obtaining Reachable Networks of Other AS from BGP

1. Overview

After the routing information of other ASs obtained by BGP is delivered to the routing table of the device, the device can forward messages sent to these ASs.

By injecting the routing information of other ASs obtained by BGP into the IGP which advertises the routing information to other devices in the local AS, these devices can forward messages sent to these ASs.

2. Restrictions and Guidelines

Generally, BGP Speakers that are IBGP neighbors to each other are not directly connected physically. IGP-enabled devices in the middle may not learn the same routing information as BGP Speakers. When the BGP Speaker at the boundary of the AS domain forwards the packets received from other domains to the next-hop IBGP neighbor, the packets will pass through the intermediate IGP-enabled device, and may be lost because there is no route on the IGP-enabled device. The synchronization of BGP and IGP can prevent such routing black hole. To achieve this, the routing information to be sent to another AS can only be propagated after all routing devices in this AS learn the routing information.

Under the following two cases, you can cancel the synchronization mechanism to ensure fast convergence of routing information:

- There is no route information traversing this AS, which in general, is a stub AS.
- All routers in the AS run the BGP protocol, and a full connection is established among all BGP speakers, that is, BGP speakers establish adjacency relationship in pairs.

3. Prerequisites

A route map must be configured. See *Configuring Routing Policies* for details about configuring a route map.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Control the routing information delivered by BGP to the core routing table.

table-map *route-map-name*

By default, BGP does not modify the attributes of all routing information delivered to the core routing table.

(5) Configure BGP to redistribute routes received from IBGP neighbors to IGP.

bgp-instance redistribute-inner-router

By default, BGP allows routes received from IBGP neighbors to be redistributed to IGP.

(6) (Optional) Enable the BGP and IGP routing information synchronization function.

synchronization

By default, the BGP and IGP routing information synchronization function is disabled.

(7) (Optional) Configure BGP to advertise the redistributed default route.

default-information originate

By default, BGP will not advertise the redistributed default route.

1.3.10 Configuring BGP Capacity Protection

1. Overview

Excessive BGP routing entries can overload the device capacity, especially on devices with a small memory. BGP capacity protection can help avoid non-predictable running status caused by excessive resource consumption of the device. BGP capacity protection includes the following two parts:

- Restricting the number of BGP routing entries
  The number of BGP routing entries can be restricted by configuring the maximum number of routing entries in a BGP address family and the maximum number of routing entries that are learned by a BGP neighbor.
- Enabling BGP to enter the overflow state in case of memory insufficiency
  After BGP enters the OVERFLOW state, it will discard the newly learned routes. This can effectively ensure that the memory will not be further consumed. If the BGP address family enters the OVERFLOW state, the newly learned route may be discarded, which may cause a network-wide routing loop. To reduce the probability of this phenomenon, BGP will generate a default route to the NULL interface, which will always exist in the OVERFLOW state.

  For the newly learned route, if the route is not a specific route which is a non-default route in the current routing table, the system will discard the route so that the memory can maintain at a relatively stable state. The system selectively discards routes to ensure that the overall network does not have routing loops. It is relatively safe for BGP to enter the OVERFLOW state.

2. Restrictions and Guidelines

- You can use the **cleanup bgp-instance** { *addressfamily* | **all** } command to reset a BGP session, and exit the OVERFLOW state of BGP address family.
- If the **unset overflow memory-lack** command is configured, BGP will not enter the OVERFLOW state when the memory is insufficient, which may cause memory resources to continue to be consumed. When the memory consumption exceeds the range allowed by the system, BGP will disconnect all neighbors and delete all learned routes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure BGP to enter the OVERFLOW state when the memory is insufficient.

overflow memory-lack

BGP enters the overflow state by default when the memory is insufficient.

(5) Enable the function of route entry limit in global configuration mode or for specified VRF instances.

bgp-instance maximum-prefix *maximum-prefix-numbers* [**vrf** *vrf-name*]

The function of route entry limit is disabled in global configuration mode or for specified VRF instances by default.

(6) Configure the maximum number of prefixes received from specified BGP peers.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **maximum-prefix** *maximum-prefix-value* [*maximum-prefixthreshold*] [**restart-time** *restart-time* | **warning** [**suppress**]]

The number of prefixes received from a specified BGP peer is not limited by default.

(7) Enter the corresponding address family configuration mode. Configure only one task.

- Enter the IPv4 unicast configuration mode.
  - **address-family-mode ipv4** [**unicast**]
- Enter the IPv4 VRF configuration mode.
  - **address-family-mode ipv4 vrf** *vrf-name*
- Enter the IPv6 unicast configuration mode.
  - **address-family-mode ipv6** [**unicast**]
- Enter the IPv6 VRF configuration mode.
  - **address-family-mode ipv6 vrf** *vrf-name*
- Enter the L2VPN configuration mode.
  - **address-family-mode l2vpn evpn**

BGP neighbors are not allowed to exchange L2VPN information by default.

(8) Configure the maximum number of prefixes in a BGP routing information base under the address family.

maximum-prefix *maximum-prefix-number*

The number of route prefixes in a BGP routing information base under an address family is not limited by default.

1.3.11 Shutting Down BGP Connections Gracefully

1. Overview

In a network, as a user isolates or upgrades a device, the RFC 6198 defines a function of shutting down BGP connections gracefully. This function demands the device to shut down connections with its neighbors when service traffic is not interrupted or is interrupted for a short time.

This feature adopts the "Make-Before-Break" mode to shut down BGP connections, to ensure that service traffic is not interrupted or is interrupted for a very short period of time in the process from the feature configuration to actual BGP connection shutdown. The steps are as follows:

(1) The feature enables the device to advertise a route with the lowest priority (the value of Local-Pref is **0** or the value of MED is 4,294,967,295), carrying Gshut Community, to a neighbor. Then, the neighbor updates the route and switches traffic to a backup link or other equal-cost links.

(2) The device shuts down the BGP connection with the neighbor after a period of time.

The delay time for BGP connection shutdown can be manually specified or automatically calculated. The automatically calculated delay time is based on 1 second for 1,000 routes.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Shut down all connections in a BGP instance gracefully.

bgp-instance shutdown [**graceful** [**append-community** *community-value*] [**delay** *delay-time*]]

None of the connections in a BGP instance is shut down by default.

(5) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(6) Shut down connections of the BGP neighbor gracefully.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **shutdown** [**graceful** [**append-community** *value*] [**delay** *time*]]

None of the connections of a BGP neighbor is shut down by default.

1.3.12 Configuring BGP Soft Reset

1. Overview

If routing policies (including peer route-filtering, peer route-map, peer prefix-list, and peer filter-rule-list) are modified, an effective method must be available to implement new routing policies. The conventional method is to configure BGP soft reset, which implements a new routing policy without terminating a BGP session connection. When an output routing policy is modified, BGP soft reset will re-advertise all routing information of a BGP speaker to its neighbors.

If an input routing policy is modified, the operation is more complex than that performed when an output routing policy is modified. This is because output routing policies are implemented on the routing table of the local BGP speaker, while the input routing policies are implemented on the routing information received from BGP peers. The local BGP speaker does not retain the original routing information received from BGP peers to reduce memory occupancy. If an input routing policy is modified and a neighbor supports the route update function, you can configure soft reset to send a route update request to the neighbor. After receiving the request, the neighbor re-advertises all routing information. You can also configure the local BGP speaker to store original routing information for each specified BGP peer and the original routing information is used as a basis for judging the modification of input routing policies.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Clear the specified session of the BGP IPv4 unicast address family.

cleanup bgp-instance ipv4 unicast-af-modifier [**vrf** vrf-name] { ***** | *as-number* | *neighbor-ipv4-address* | *neighbor-ipv6-address* } [**soft**] [**inbound** | **outbound**]

(3) Enter the global configuration mode.

config-system terminal

(4) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(5) (Optional) Save the original routing information of a specified BGP peer.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **soft-resave inbound**

No original routing information of a specified BGP peer is saved by default.

1.3.13 Configuring Scope Configuration Mode

1. Overview

BGP can be configured based on VRF by configuring the Scope configuration mode. The commands configured in the Scope configuration mode only take effect in the corresponding VRF.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure to enter the Scope configuration mode, and associate VRF with BGP.

scope vrf *vrf-name*

By default, the Scope address family is not configured.

1.3.14 Configuring BGP Route Update Mechanisms

1. Overview

BGP provides two route update mechanisms: regular scanning update and event triggering update.

- Regular scanning update

  BGP uses an internal timer to start scanning regularly and update the routing table.

- Event triggering update

  When BGP configuration commands are changed due to user configuration or the next hop of a BGP route changes, BGP is triggered to start the scanning mechanism and update the routing table.

  The function of triggering routing table update by the change of the next hop of BGP is a method of reducing the BGP convergence time. This function optimizes the method of monitoring the next hops of routes, to enable BGP to speed up route convergence when the network topology is stable. When BGP establishes a connection with a neighbor, it automatically monitors the next hops of the routes learned from the neighbor. When a next hop changes in the core routing table, BGP receives a next hop change advertisement and updates the BGP routing table. This optimization measure improves the convergence performance of BGP routes by reducing the time for detecting next-hop changes.

2. Restrictions and Guidelines

- Based on address families, this function is configured in the IPv4, IPv6, IPv4 VRF, and IPv6 VRF address family modes.
- If you set the BGP route update mechanism to event-triggering update, you must disable synchronization (by running the **unset synchronization** command) and enable the BGP next-hop triggering update function.
- The **bgp-instance nexthop-scan trigger delay** command and the **bgp-instance scan-interval** command control the same timer. If the BGP route update mechanism is configured as regular scanning update and the timer specified in the **bgp-instance nexthop-scan trigger delay** command is set to greater than 60 seconds, this command does not take effect, because the scanning timer is triggered prior to the delay of the next hop triggering update.
- In an unstable network environment with frequent hop changes, especially with many routes, the configuration of the next hop triggering update causes unnecessary route calculations, and thus consumes more CPU resources. Therefore, you are advised to disable this function in this environment.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure an interval of regular scanning update of BGP.

bgp-instance scan-interval *scan-time*

The default interval of regular scanning update of BGP is **60** seconds.

(5) (Optional) Enter the address family configuration mode. To configure this function for a specific address family, please configure only one task.

- Enter the IPv4 unicast configuration mode.

  address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

- Enter the L2VPN configuration mode.

  address-family-mode l2vpn evpn

  BGP neighbors are not allowed to exchange L2VPN information by default.

- Enter the link-state configuration mode.

  address-family-mode link-state [**unicast**]

(6) Configure the BGP route update mechanism as event triggering update. Run the following commands in turn:

- Configure the event triggering update mechanism to update the routing table.

  bgp-instance scan-rib disable

  BGP updates the routing table in regular scanning mode by default.

- Enable the next hop triggering update.

  bgp-instance nexthop-scan trigger enable

  The function of next hop triggering update is enabled by default.

- Configure a delay of updating the routing table before the next hop of a BGP route changes.

  bgp-instance nexthop-scan trigger delay *delay-time*

  The default delay of updating the routing table is **5** seconds after the next hop of a BGP route changes.

1.3.15 Configuring BGP Timers

1. Overview

By manually configuring timers in the BGP protocol, you can flexibly meet neighbor keepalive requirements in different network environments.

- Holdtime and Keepalive timers

BGP uses the Keepalive timer to maintain an effective connection with the peer, and uses the Holdtime timer to determine whether the peer is alive. When a BGP connection is established between BGP peers, the peers will negotiate the Holdtime interval and choose the smaller Holdtime value. The Keepalive timer value will be 1/3 of the negotiated Holdtime interval or the configured Keepalive value, whichever is smaller.

- Connect-Retry timer

To reduce the impact of BGP's frequent attempts to reconnect with disconnected neighbors on the network bandwidth, the BGP peer will initiate a Connect-Retry timer after detecting the failure to establish a connection with the neighbor, and try to reconnect after the interval expires.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the BGP timer interval.

timers-set bgp *keepalive-time holdtime* [*minimum-holdtime*]

By default, the interval for BGP to send keepalive messages is 60 seconds, and the hold time is 180 seconds. The minimum hold-time for neighbor notification is not limited.

(5) Configure the interval of Keepalive, Hold-Time, and Connect-Retry timers when establishing a BGP connection with a specified BGP peer.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **timers-set** { *keepalive-interval holdtime* [*minimum-holdtime*] | **connect** *connect-retry* }

By default, when establishing a BGP connection with a BGP peer, the interval of the Keepalive timer is 60 seconds, that of the Hold-Time timer is 180 seconds, and that of the Connect-Retry timer is 15 seconds.

(6) (Optional) Enable BGP internal timer accurate control function.

bgp-instance timer accuracy-control

By default, the BGP internal timer accurate control function is disabled.

1.4 Configuring a Large-Scale BGP Network

1.4.1 Configuration Tasks

The large BGP network configuration includes the following tasks: All configuration tasks are optional and may be selected as needed.

- Configuring BGP Route Reflector
- Configuring BGP AS Alliance
- Configuring Route Aggregation
- Configuring Route Dampening

1.4.2 Configuring BGP Route Reflector

1. Overview

A route reflector can be configured in an IBGP environment so that routing information is reflected between clients and the number of BGP neighbor connections is reduced.

2. Restrictions and Guidelines

- If you configure a route reflector in an address family, the reflector takes effect only in this address family.
- To facilitate configuration, you need to configure the reflector device. The client devices do not need to be configured as clients.
- When multiple reflection clusters are deployed in one AS domain, different reflection cluster identifiers must be configured for them.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(5) Configure the local device as a route reflector. Then, the BGP neighbor becomes the route reflection client.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **route-reflector-client**

(6) (Optional) Enable the route reflection function between clients of the devices.

bgp-instance client-reflection enable

The route reflection function between clients is enabled by default.

(7) (Optional) Configure a BGP reflection cluster ID.

bgp-instance cluster-id *cluster-id*

A BGP reflection cluster ID is the BGP router-ID by default.

(8) (Optional) Allow the BGP route reflector to modify route attributes.

bgp-instance route-reflector attribute-change

No route reflector is allowed to modify route attributes by default.

According to RFC4456, to prevent route loops, a BGP route reflector cannot modify route attributes when reflecting routes. However, relevant route attributes need to be modified when you re-plan network traffic paths. Users can configure a command to modify route attributes (including Route-map and Nexthop-self) on the route reflector.

1.4.3 Configuring BGP AS Alliance

1. Overview

On a large BGP network, you can configure a BGP alliance to divide an AS into multiple member ASs, to reduce the number of IBGP neighbor connections and facilitate AS management.

2. Restrictions and Guidelines

- You are advised to use private AS numbers for member ASs within an alliance. A private AS number is in the range from 64512 to 65535.
- Within a member AS of an alliance, BGP speakers must be connected in a full mesh manner or configured as reflectors and reflector clients.
- EBGP neighbor relationships must be established between member ASs of an alliance.
- All BGP speakers in an alliance must associate with a member AS of the alliance.
- All member ASs that have established EBGP relationships with the local device must be configured on members of a BGP alliance.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure a BGP alliance ID. The local BGP AS (specified by the **router-protocol bgp** *as-number* command) becomes the private AS inside the alliance and is invisible to other AS domains.

bgp-instance as-confederation id *as-number*

No alliance ID is configured for a BGP speaker by default.

(5) Configure members for the BGP alliance. The specified AS and local AS belong to the same alliance.

bgp-instance as-confederation peers *as-number*

No alliance member is configured for BGP by default.

1.4.4 Configuring Route Aggregation

1. Overview

One or more refined BGP routes are aggregated as a BGP route with a shorter network mask.

2. Restrictions and Guidelines

- By default, BGP advertises path information before and after aggregation. To enable BGP to advertise only path information after aggregation, run the **summary-address summary-only** command.
- When the **summary-address** command is used to configure route aggregation, the aggregated routes take effect immediately as long as any route falls into the configured address range.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the route aggregation entries of BGP.

summary-address { *ipv4-address mask* | *prefix* } [**advertise-route-map** *route-map-name* | **gen-sys-set** | **attribute-route-map** *route-map-name* | **summary-only** | **suppress-route-map** *route-map-name*] *

No route aggregation entry of BGP is configured by default.

1.4.5 Configuring Route Dampening

1. Overview

BGP route dampening is a method of reducing route flapping. It reduces possible route flapping by monitoring routing information from EBGP peers. The BGP speaker punishes a route once (added to the penalty) route flapping occurs. When the penalty reaches the upper limit, the route is suppressed. When the Half-life-time reaches, the penalty is halved. When the penalty is reduced to the Reuse Limit, the route is reactivated.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Enable the route dampening function and configure dampening parameters.

bgp-instance damp [*half-life* [*reusing suppressing maximun-supress-time*] [**withdrawal-ignore**] | **route-map** *route-map-name*]

The route dampening function is disabled by default.

1.5 Configuring BGP Route Selection and Load Balancing

1.5.1 Configuration Tasks

The configuration of BGP route selection and load balancing includes the following tasks: All configuration tasks are optional and may be selected as needed.

- Configuring BGP Multi-Path Load Balancing
- Configuring BGP Routes to be Recursive Only to Host Routes
- Configuring Outbound Loop Detection for a BGP Neighbor
- Configuring BGP ADD-PATH
- Configuring BGP Advertising Routes with the Lowest Priority upon Device Restart
- Configuring AS_PATH Attribute
- Configuring MED Attribute
- Configuring AIGP Attribute
- Configure LOCAL_PREF Attribute
- Configure Extended Community Attribute
- Configuring BGP Best Path Selection

1.5.2 Configuring BGP Multi-Path Load Balancing

1. Overview

This section describes how to implement multi-path load balancing for IBGP routes.

2. Restrictions and Guidelines

- Routes learned from an IBGP neighbor must have the same priority (the router-ID is not compared).
- This command does not allow IBGP and EBGP routes to form equivalent routes.
- All next hops must carry the Link-Bandwidth attribute so that non-equal-cost routes are generated. Otherwise, non-equal-cost load balancing cannot be implemented.
- Enable the function of sending extended community attributes so that the link bandwidth attribute can be sent to an IBGP neighbor.
- The function of carrying the link bandwidth attribute when a route from a specified neighbor is sent to an IBGP neighbor takes effect only on single-hop EBGP neighbors.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Enter a required address family configuration mode. If you need to perform configuration for a specific address family, select one of the following tasks to configure.

- Enter the IPv4 unicast configuration mode.

  address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

(5) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(6) Configure BGP load balancing. To configure equal-cost route load balancing by EBGP multiple paths of an alliance or by local inter-VRF imported routes, run the **max-ecmp-paths ebgp** command.

max-ecmp-paths { **ebgp** | **ibgp** } *maximum-paths-number*

BGP load balancing is disabled by default.

(7) (Optional) Enable the function of sending extended community attributes, to ensure that the link bandwidth attribute is transferred between neighbors.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **send-community** [**both** | **standard** | **extended**]

By default, when configuring a BGP peer for L2VPN address families, the extended community attributes will be advertised to this BGP neighbor at the same time. When a BGP peer is configured for other address families, the community attribute will not be advertised to this BGP neighbor.

(8) (Optional) Enable non-equal-cost load balancing. Non-equal-cost routes are generated based on the link bandwidth attribute.

bgp-instance link-bandwidth

Non-equal-cost load balancing is disabled by default.

(9) (Optional) Configure the unit of bandwidth transmission as kbytes.

bgp-instance link-bandwidth-kbytes

The transmission unit in the link bandwidth extended attribute is bytes by default.

(10) (Optional) Set the transmission mode of link bandwidth extended attribute to **non-transitive**.

bgp-instance link-bandwidth-non-transitive

The transmission mode of link bandwidth extended attribute is **transitive** by default.

(11) (Optional) Enable the function of carrying the link bandwidth attribute when a route from a specified neighbor is sent to an IBGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **link-bandwidth**

The function of carrying the link bandwidth attribute in a route from a specified neighbor is disabled by default.

(12) (Optional) Configure link bandwidth attribute to be carried in the specified neighbor's routes sent to other EBGP neighbors.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **link-bandwidth-ebgp**

No link bandwidth attribute is carried in the specified neighbor's routes sent to other EBGP neighbors by default.

(13) (Optional) Configure the link bandwidth extended community attribute received from the specified neighbor to be used for calculating non-equal-cost load balancing and cumulative bandwidth values.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **link-bandwidth-transitive**

The link bandwidth extended community attribute received from the specified neighbor is not configured to calculate non-equal-cost load balancing and cumulative bandwidth values by default.

1.5.3 Configuring BGP Routes to be Recursive Only to Host Routes

1. Overview

BGP routes use optimal matching for route recursion by default. BGP routes may be recursive to default routes or incorrect network segment routes, resulting in an error in the egress or next hop. After the function of making BGP routes recursive only to host routes is enabled, BGP routes are recursive only to 32-bit IPv4 host routes or 128-bit IPv6 host routes. This function is applied to routes learned by IBGP neighbors or multi-hop EBGP neighbors that have established neighbor relationships by using the loopback interface.

After the function of making BGP routes recursive only to host routes is enabled, the device checks the validity of the next-hop addresses of IBGP or multi-hop EBGP routes via exact matching, and conducts route recursion on the IBGP or multi-hop EBGP routes via exact matching when they are written into the forward information database (FIB). For example, the next-hop address of the IBGP route 192.168.2.0/24 is 1.1.1.1. The next-hop address of the IBGP route 192.168.2.0/24 is valid only when the route 1.1.1.1/32 exists, and the egress of the route written into the FIB is the egress of the route 1.1.1.1/32.

2. Restrictions and Guidelines

- This function takes effect only on IBGP or multi-hop EBGP routes. It is unavailable to direct EBGP routes or routes whose next-hop addresses are direct connection addresses.
- IBGP or multi-hop EBGP neighbor relationships of the device need to be established using the address of the loopback interface. Otherwise, this function cannot be enabled.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure BGP routes to be recursive only to host routes.

bgp-instance recursive host-route

BGP routes recurse via optimal matching by default.

1.5.4 Configuring Outbound Loop Detection for a BGP Neighbor

1. Overview

When receiving a BGP route from a neighbor, BGP conducts loop detection on the BGP route by default. If the AS-PATH attribute carried in a BGP route contains the local AS number, BGP filters out the BGP route. The BGP outbound loop detection is to conduct loop detection on routes before the routes are transmitted to a neighbor, so as to filter out loop routes.

When sending a route to an EBGP neighbor, the device judges whether the AS-PATH attribute carried in the BGP route contains the AS number of the neighbor. If yes, the route is looped and the device does not send the route to the EBGP neighbor.

2. Restrictions and Guidelines

This feature is available only to EBGP neighbors.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(5) Configure outbound loop detection for the BGP neighbor

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **as-loop-check out**

The outbound loop detection is disabled for a neighbor by default.

1.5.5 Configuring BGP ADD-PATH

1. Overview

The BGP additional paths (ADD-PATH) attribute is used to receive multiple routes with the same prefix from an IBGP neighbor. When receiving routes with the same prefix from the same peer, BGP replaces a previous route with the latest new one by default, that is, BGP regards the routes with the same prefix received from the same neighbor as one entry. After the BGP ADD-PATH function is configured, BGP adds a path-ID field to the route advertised to a neighbor that supports the ADD-PATH capability. The device judges whether routes are the same based on the route prefix and path-ID of a neighbor. When a received route is different from previous

routes, the device adds it as a new entry. In this way, the device can obtain multiple routes with the same prefix from the same neighbor.

2. Restrictions and Guidelines

Run the bgp-instance alternative-paths select command to select ADD-PATH alternative routes of a specific type. If the type of selected ADD-PATH alternative routes is different from that of to-be-advertised ones, the ADD-PATH alternative routes are not advertised but only the optimal routes are advertised.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Enable the ADD-PATH selection function.

bgp-instance alternative-paths select { **all** | **best** *number* | **ecmp** }

(5) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** as-number

(6) The ADD-PATH function is disabled by default. Configure the device to advertise the ADD-PATH alternative routes of a specific type to a peer when the ADD-PATH negotiation is successful.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **advertise alternative-paths** { **all** | **best** *number* | **ecmp** }

BGP does not advertise ADD-PATH alternative routes to its neighbors by default.

(7) Enable the ADD-PATH capability for a specified peer.

peer { *peer-address* | *peer-group-name* } **alternative-paths** { **send [recv] | recv** }

The ADD-PATH capability of a peer is disabled by default.

1.5.6 Configuring BGP Advertising Routes with the Lowest Priority upon Device Restart

1. Overview

By default, after a neighbor relationship is established upon device restart, a BGP peer can advertise routes to its neighbors. However, in certain cases, for example, a device has many neighbors, there are considerable routes during device startup, but writing entries into the hardware is slow. The neighbors have learned the routes and started forwarding traffic, but the local device has not completed the writing of entries into the hardware, causing a traffic forwarding failure. Therefore, it is necessary to adjust the priority of routes advertised by BGP to the lowest priority after the device restart. In this way, routes passing through the local device are not selected while other routes are available.

This function enables BGP to advertise routes with the lowest priority to neighbors after the local device is started and has established a neighbor relationship. For EBGP neighbors, the MED value of routes is adjusted to 4294967295. For IBGP neighbors, the LOCAL_PREF value of routes is adjusted to 0. After confirming that ARP entries have been delivered and forwarding entries have been written into hardware, run the cleanup

bgp-instance advertise lowest-priority startup command to restore the priority of advertised routes. If you do not configure the priority restoration for advertised routes, the device automatically restores the priority of advertised routes after a period of time.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure BGP to adjust the priority of advertised routes to the lowest level upon device restart.

bgp-instance advertise lowest-priority startup [*recover-time*]

BGP does not modify the priority of advertised routes by default.

1.5.7 Configuring AS_PATH Attribute

1. Overview

AS_PATH Lists the ASs passed by a route reversely. Routes are selected based on different AS_PATH attributes.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Enter the address family configuration mode. To configure this function for a specific address family, please configure only one task.

- Enter the IPv4 unicast configuration mode.

  address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

- Enter the L2VPN configuration mode.

  address-family-mode l2vpn evpn

- Enter the link-state configuration mode.

  address-family-mode link-state [**unicast**]

(5) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(6) (Optional) Configure to check the first AS number in the AS_PATH path field of the UPDATE message.

bgp-instance specify-first-as

By default, the device checks the first AS number in the AS_PATH path field of the UPDATE message.

(7) (Optional) Configure a BGP route to limit the number of AS numbers in the AS_PATH attribute.

bgp-instance max-as-limit *maxas-limit-number*

By default, the number of AS numbers in the AS_PATH attribute is not limited.

(8) (Optional)Configure a local AS number for all BGP peers of the current VRF instance.

bgp-instance local-as *as-number* [**no-prepend** [**change-as** [**dual-as**]]]

By default, no local AS number is configured for all BGP peers of the current VRF instance by default.

When the AS number of the local BGP system changes, there is no need to change the BGP configuration of the peer device. This function is used for AS migration and merging on large-scale networks while ensuring consistent configuration of other devices in interconnected ASs.

(9) Configure the local AS for the BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **local-as** *as-number* [**no-prepend** [**change-as** [**dual-as**]]]

No local AS is configured for any peer by default, and the local AS of a peer is the real AS of BGP.

After the local AS function is configured, when the real AS of the local BGP changes, you do not need to change the BGP configuration on the peer and a BGP connection can be still established. This function is used for AS migration and merging on large networks and ensures that the device configurations in other interconnected ASs are not affected.

(10) (Optional) Configure to delete the private AS number recorded in the AS path attribute of the route propagated to the specified EBGP peer.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **delete-private-as** [**force** [**ignore-remote-as**]] [**change-as**]

By default, the private AS number recorded in the AS path attribute of the route propagated to the specified EBGP peer is not deleted.

1.5.8 Configuring MED Attribute

1. Overview

The MED attribute is used to determine the best path when traffic enters the AS. When BGP Speaker obtains multiple routes with the same destination IP address but different next-hop IP addresses through different EBGP peers, if other conditions are the same, the priority will be given to the route with a smaller MED value when selecting the best path.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Enable BGP to always compare the MED value.

bgp-instance match-med

By default, the always compare MED value is disabled.

(5) (Optional) Configure to preferably compare the MED value of paths from peers of the same AS.

bgp-instance definitive-med

By default, the preferably compare MED value of paths from peers of the same AS is not configured.

1.5.9 Configuring AIGP Attribute

1. Overview

AIGP attribute can accumulate the IGP Metric of paths the route passes through. BGP uses the AIGP attribute for routing optimization. A route with a smaller AIGP attribute value has a higher priority.

By default, BGP route does not carry AIGP attribute. You can run the set-rule aigp-metric command of the route map to configure the AIGP attribute for the route.

2. Prerequisites

The AIGP weight attribute for routes that match the Match rule is configured. See *Configuring Routing Policy* for details.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Enable the AIGP function of BGP neighbors.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **aigp** [**transform med**]

By default, the AIGP function of BGP neighbors is disabled.

1.5.10 Configure LOCAL_PREF Attribute

1. Overview

When a BGP Speaker propagates a route received from the EBGP peer to another IBGP peer, it will add the LOCAL_PREF attribute. BGP will compare the LOCAL_PREF value when selecting the best path. A path with a higher LOCAL_PREF value has a higher priority.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the default LOCAL_PREF attribute value.

bgp-instance default local-preference *local-preference-value*

By default, the LOCAL_PREF value is 100.

1.5.11 Configure Extended Community Attribute

1. Overview

The extended community attribute is the extension of BGP community attribute for specific services.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure BGP to carry the non-transitive extended community attribute when advertising routes to EBGP neighbors.

bgp-instance advertise non-transitive extcommunity

By default, BGP does not carry the non-transitive extended community attribute when advertising routes to EBGP neighbors.

1.5.12 Configuring BGP Best Path Selection

1. Overview

When BGP selects the best path, it needs to compare some routing attributes. You can use the configuration commands to configure the policy for the system to compare routing attributes.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) The following steps are optional, and can be configured as needed.

- Configure to not compare the AIGP Metric value when selecting the best path.

- **bgp-instance optimal-path aigp ignore**
- By default, the AIGP Metric value is compared when selecting the best path.

- Configure to not compare the AS path length when selecting the best path.
  - **bgp-instance optimal-path as-path ignore**
  - By default, the AS path length is compared when selecting the best path.
- Configure BGP to relax the comparison of the AS_PATH attribute for BGP multipath load balancing.
  - **bgp-instance optimal-path as-path multipath-relax**
  - By default, relax comparison of the AS_PATH attribute for BGP multipath load balancing is not configured.
- Allow BGP to compare the alliance's AS AS_PATH length when selecting the best path.
  - **bgp-instance optimal-path compare-confed-aspath**
  - By default, the AS_PATH attribute value of an EBGP route within the same alliance is not compared when selecting the best path. The best path is selected based on other conditions.
- Configure to compare Router ID of paths when selecting the best path.
  - **bgp-instance optimal-path compare-router-id**
  - By default, when selecting the best path, if all path attributes of two paths received from different EBGP peers are the same, the path received first is considered to be the best path.
- Configure to not compare the serial number of EVPN table entries from the same VTEP when selecting the best path.
  - **bgp-instance optimal-path evpn-seq-relax**
  - By default, when selecting the best path, the device needs to compare the serial number of EVPN table entries from the same VTEP.

Note

This command is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

- Configure not to compare the IGP Metric value of the next hop when selecting the best path.
  - **bgp-instance optimal-path igp-metric ignore**
  - By default, the IGP Metric value of the next hop is compared when selecting the best path.
- Configure to compare the MED of the paths from the peers within the AS alliance when selecting the best path.
  - **bgp-instance optimal-path med confed** [**worst-med**]
  - By default, the MED value of the paths from peers in the AS alliance is not compared when selecting the best path.
- Set the lowest priority for paths without the MED attribute when selecting the best path.
  - **bgp-instance optimal-path med worst-med**
  - By default, when selecting the best path, the priority of paths without the MED attribute is set to the highest.
- Enable the Router ID comparison function for BGP multipath load balancing.

- **bgp-instance optimal-path multipath-compare-routerid**
- By default, the Router ID comparison function for BGP multipath load balancing is disabled.

- Reduce the priority of VTEP EVPN entry capacity comparison for optimal path election.

  bgp-instance optimal-path vtep-compare-delay

  Reducing the priority of VTEP EVPN entry capacity comparison for optimal path election is disabled by default.

1.6 Configuring BGP Controlling Route Advertisement and Receiving

1.6.1 Configuration Tasks

The configuration of route advertisement and receiving control of BGP includes the following tasks: All configuration tasks are optional and may be selected as needed.

- Configuring BGP Filters
- Configuring BGP Route Sending Interval and Count
- Configuring BGP Modifying Propagated Routing Information
- Configuring BGP Processing Received Routing Information
- Configuring BGP Fast Withdrawing Specific Route
- Configuring BGP Delaying Advertising Routing Information

1.6.2 Configuring BGP Filters

1. Overview

You can configure BGP filters to enable the device to advertise or receive only specified routes. BGP supports multiple route filtering policies.

2. Prerequisites

If the command parameter involves a route map, you need to configure the route map first. See *Configuring Routing Policies* for configuring a route map.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the global routing policy for matching for outbound route advertisement.

bgp-instance advertise-route-map *route-map-name*

By default, the global routing policy is not used for matching when sending outbound routes to BGP neighbors.

(5) Configure to use route matching policy for inbound or outbound routes.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **route-map** *map-tag* { **in** | **out** }

By default, route matching is not configured for inbound or outbound routes.

(6) Configure to implement routing policies according to ACL when sending and receiving routing information from or to a specified BGP peer.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **route-filtering** { *access-list-name* | *access-list-number* } { **in** | **out** }

By default, the routing policy is not configured according to ACL when sending and receiving routing information from or to the specified BGP peer.

(7) Configure to selectively advertise route information suppressed by route aggregation.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **unsuppress-route-map** *map-tag*

By default, routing information that is suppressed by routing aggregation is not selectively advertised.

(8) Allow BGP Speaker to advertise default routes to peers (groups).

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **default-originate** [**route-map** *map-tag*]

By default, BGP Speaker is not allowed to advertise default routes to peers (groups).

(9) Allow BGP Speaker to receive the UPDATE message containing the AS number of local device.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **allowas-in** [*occurrence-number*]

By default, BGP Speaker is allowed to receive the UPDATE message containing the AS number of local device.

(10) Configure that source-specific routes are not suppressed within the domain group to which the specified BGP peer belongs.

- Configure the domain group to which the specified BGP peer belongs.
  - **peer** { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **domain** *domain-name*
  - By default, the Domain group to which the specified BGP peer belongs is not configured.
- Configure the domain group to which the specified BGP peer belongs to not suppress source-specific routes.
  - **peer** { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **domain-unsuppress**
  - By default, the domain group to which the specified BGP peer belongs does not suppress source-specific routes.

(11) (Optional) Enable the Route-Target filtering function.

bgp-instance default-config routing-target filter

By default, the Route-Target filtering function is enabled.

Note

This command is not supported by the S6980-64QC, or S6580-24DC8QC.

1.6.3 Configuring BGP Route Sending Interval and Count

1. Overview

You can tune the outbound route advertisement speed of the BGP Speaker by configuring the time interval and number of BGP routes advertised.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the time interval for sending BGP route update messages.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **advertisement-interval** *advertisement-interval*

By default, the time interval for sending BGP route update message is 0 seconds for IBGP connections, and 30 seconds for EBGP connections.

(5) Configure the time interval to advertise the specified peer of the locally originated BGP route.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **as-origination-interval** *seconds*

By default, the time interval for advertising the locally originated BGP route to the specified peer is 1 second.

1.6.4 Configuring BGP Modifying Propagated Routing Information

1. Overview

You can use configuration commands to enable BGP to carry the specified information or modify the specified information in the route advertised by BGP.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the source site attribute values of the BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **soo** *soo-value*

By default, the source site attribute value of the BGP neighbor is not configured.

(5) Configure PE to override the AS number of a site.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **as-override**

By default, PE is not configured to override the AS number of a site.

(6) Configure the next hop when distributing routes to a specified BGP peer. The following tasks are mutually exclusive. Configure only one of them.

- Configure to set the next hop of route as this device when distributing routes to a specific BGP peer.
  - **peer** { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **nexthop-local**
  - By default, the next hop is the BGP Speaker when advertising routes to an EBGP peer, and the next hop is not modified when advertising routes to an IBGP peer.
- Configure to not modify the next hop when advertising routes to a peer (group).
  - **peer** { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **nexthop-unaltered**
  - By default, the next hop will be changed to the BGP Speaker when advertising routes to an EBGP peer, and the next hop will not be modified when advertising routes to an IBGP peer.

1.6.5 Configuring BGP Processing Received Routing Information

1. Overview

You can use configuration commands to configure BGP to process the received routing information.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure not to perform private PIC processing on routing information sent to and received by a BGP peer when distributing routes to and receiving routes from the designated BGP peer.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **pic-disable**

By default, the device performs private PIC processing on the routes sent and received by BGP peers.

(5) Configure the weight of the specified BGP peer.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **private-value** *weight-value*

By default, the weight of the specified BGP peer is not configured. The initial weight of routes learned from the neighbor is 0, and that of locally generated routes is 32768.

1.6.6 Configuring BGP Fast Withdrawing Specific Route

1. Overview

With the rapid development of IP technologies and the application of various complex services, higher requirements are posed for the network. Customers may need fast convergence of specific routes. This function can meet customers' requirements for preferential withdrawal of specific routes.

The route withdrawal sequence is calculated based on the routes in the routing table by default. This function adds routes that meet conditions to a high-priority queue for preferential processing, so as to achieve the purpose of rapidly withdrawing specified routes.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Enter the address family configuration mode. To configure this function for a specific address family, please configure only one task.

- Enter the IPv4 unicast configuration mode.

  address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

(5) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(6) Configure BGP to rapidly withdraw specified routes.

bgp-instance fast-withdraw { **access-control-list** { *access-list-number* | *access-list-name* } | **prefix-list** *prefix-list-name* | **route-map** *route-map-name* }

The rapid withdrawal of specified routes is disabled by default.

(7) Allow BGP Speaker to fast withdraw default routes to peers (groups).

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **default-fast-withdraw**

By default, BGP Speaker does not fast withdraw the default route to peers (groups).

1.6.7 Configuring BGP Delaying Advertising Routing Information

1. Overview

- Delayed advertisement upon system restart

  After the system is restarted and a neighbor relationship is established, a BGP peer can advertise routes to its neighbors. However, in certain cases, for example, a device has many neighbors or routes and the device is restarted. So, the device is slow in writing entries to hardware. As a result, the neighbors have learned the routes and started forwarding traffic, but the local device hardware has not completed writing of entries, causing a traffic forwarding failure.

  The BGP delayed advertisement upon system restart enables the device not to immediately advertise routes to its neighbors but wait for a period of time after the system is restarted and the BGP neighbor relationships are established. This function does not affect other behaviors of neighbors such as route

receiving. If some routes need to be kept from impact of the system delay, you can configure a prefix-list policy to match these routes, providing a flexible route advertisement method.

Delay-time specifies the waiting time before routes are advertised to neighbors and startup-time specifies the startup-time. Within the startup-time, BGP sends received routing information to the neighbors at an interval specified by delay-time. After the startup-time ends, the default route advertisement behavior recovers.

- Delayed first route advertisement

  Once the neighbors of BGP negotiate to reach the established status, they send update packets each other. If delayed first route advertisement of BGP is configured on the local device, the local device will send the route received from the specified neighbor to other neighbors after the delayed period.

2. Restrictions and Guidelines

- If delayed advertisement of BGP upon system restart and delayed first route advertisement of BGP are enabled at the same time, delayed advertisement of BGP upon system restart takes precedence over delayed first route advertisement for first routes of BGP.
- BGP GR is not affected by either delayed advertisement of BGP upon system restart or delayed first route advertisement of BGP, that is, the BGP GR route advertisement is not affected by the delay time.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Enable the function of delayed advertisement of BGP upon system restart.

bgp-instance initial-advertise-delay { *delay-time* [*startup-time*] [**wait-for-controller**] | **prefix-list** *prefix-list-name* }

The function of delayed advertisement of BGP upon system restart is disabled by default.

(5) Configure the maximum delay time for the BGP speaker to first send update messages.

bgp-instance advertise-delay *delay-time*

The default maximum delay time for the BGP speaker to first send update messages is **120** seconds.

1.7 Configuring the BGP Security Features

1.7.1 Overview

The configuration of the BGP security function includes the following tasks: All configurations are optional and may be selected as needed.

- Configuring MD5 Authentication
- Configuring BGP Peer GTSM Security Check

1.7.2 Configuring MD5 Authentication

1. Overview

Message-digest algorithm 5 (MD5) authentication passwords are used when BGP speakers establish a TCP connection. If the authentication fails, no TCP connection will be established.

2. Restrictions and Guidelines

- The same password must be set for the BGP speakers that need to establish a BGP connection. Otherwise, no neighbor relationship can be established between them.
- After the MD5 authentication command is configured, BGP speakers re-establishes a BGP connection.
- Only one authentication password can be configured for one neighbor.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Enter the address family configuration mode. To configure this function for a specific address family, please configure only one task.

- Enter the IPv4 unicast configuration mode.

  address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

- Enter the L2VPN configuration mode.

  address-family-mode l2vpn evpn

  BGP neighbors are not allowed to exchange L2VPN information by default.

- Enter the link-state configuration mode.

  address-family-mode link-state [**unicast**]

(5) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(6) Configure MD5 encryption for a TCP connection to be established with the BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **password** [**0** | **7**] *password-string*

By default, a BGP connection is not encrypted through MD5.

1.7.3 Configuring BGP Peer GTSM Security Check

1. Overview

Network devices consume CPU resources when processing IP packets on the control plane. An attacker continuously sends false BGP packets to a device. If the device finds the packets sent to the local device upon receipt of them and forwards the packets to the control plane for processing without identification, the device consumes many CPU resources, because it needs to process these packets.

Configuring BGP GTSM security check can protect devices from CPU type attacks. GTSM security check determines whether the TTL value in IP packet headers is within the specified range. If not, the packets are invalid and directly discarded.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the function of GTSM security check for BGP neighbors.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **ttl-safety hops** *hop-count*

The function of GTSM security check is not enabled by default.

1.8 Configuring BGP Reliability

1.8.1 Configuration Tasks

BGP reliability configuration includes the following tasks: All configurations are optional and may be selected as needed.

- Configuring BFD for BGP
- Configuring BGP FRR
- Configuring EBGP Multi-Path Bypass Protection
- Configuring BGP GR
- Configuring BGP NSR
- Configuring BGP Session
- Configuring EBGP Fast Neighbor Detection
- Configuring BGP Link-State Group Association

1.8.2 Notes

This section describes the precautions for configuring the BGP reliability function.

Whether the BGP reliability function is supported depends on devices.

Table 1-2 Products Supporting the BGP Reliability Function

Function	Device Supporting the Function
Configuring BGP NSR	S6520-64CQ, S6510-32CQ, S6510-48VS8CQ, S6510-4C, and S6250-48XS8CQ

1.8.3 Configuring BFD for BGP

1. Overview

You can configure bidirectional forwarding detection (BFD) for BGP to rapidly detect a network failure and switch data to a backup link, thereby improving the network use experience.

2. Restrictions and Guidelines

You need to configure BGP peer-specific BFD detection parameters for multi-hop BFD only. Single-hop BFD can use BGP peer-specific BFD detection parameters or BFD detection parameters configured on an interface. Peer-specific BFD detection parameters take precedence over the parameters configured on an interface.

3. Prerequisites

Before configuring BFD for BGP, make sure that GP peer-specific BFD session parameters have been configured. Peer-specific BFD detection parameters take precedence over the parameters configured on an interface. The REUP link-state group is configured. See *Configuring BFD* for details.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Enter the address family configuration mode. To configure this function for a specific address family, please configure only one task.

- Enter the IPv4 unicast configuration mode.

  address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

(5) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(6) Configure BFD for BGP to detect the neighbor changes.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **fall-over bfd** [**single-hop-prefer**] [**compatible**]

The BFD function is disabled by default.

(7) (Optional) Configure BFD detection parameters for the specified BGP peer (group).

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **fall-over bfd-detect interval** *interval-number* **min-rx** *min-rx-number* **multiplier** *multiplier-number*

No BFD detection parameters are configured for the specified BGP peer (group) by default.

1.8.4 Configuring BGP FRR

1. Overview

This section describes how to use a backup route to replace a failed route to rectify the fault locally.

2. Prerequisites

Configure a BFD session for a neighbor to implement fast link fault detection.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Enter the address family configuration mode. To configure this function for a specific address family, please configure only one task.

- Enter the IPv4 unicast configuration mode.

  address-family-mode ipv4 [**unicast**]

- Enter the IPv4 VRF configuration mode.

  address-family-mode ipv4 vrf *vrf-name*

- Enter the IPv6 unicast configuration mode.

  address-family-mode ipv6 [**unicast**]

- Enter the IPv6 VRF configuration mode.

  address-family-mode ipv6 vrf *vrf-name*

(5) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(6) Configure BGP FRR.

bgp-instance auto-frr

The FRR function is disabled by default.

1.8.5 Configuring EBGP Multi-Path Bypass Protection

1. Overview

After EBGP multi-path bypass protection is configured, when all ECMP routes fail, the bypass route becomes the primary route for data forwarding.

BGP FRR supports only 1:1 protection. After the equal-cost multi-path routing (ECMP) function is enabled, FRR becomes unavailable. The multi-path bypass protection enables BGP to support 1:N protection. Users can configure BGP multi-path bypass protection to enable the system to select a backup bypass path even if ECMP is configured. When all ECMP routes fail, the bypass route becomes the primary route for forwarding.

When there are multiple paths to the same network in the routing table of BGP, BGP calculates a route with the highest priority by default. After the BGP multi-path bypass protection function is enabled, BGP selects a preferred backup route even if ECMP routes exist. After the fast link detection mechanism of BFD detects that the master link is faulty, the system switches data to the originally calculated backup link for forwarding. After route convergence is completed, the device switches data to the re-calculated optimal path for data forwarding.

2. Restrictions and Guidelines

You can configure a neighbor BFD session to implement fast link fault detection.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure BGP multi-path bypass protection.

bgp-instance install standby-path

Multi-path bypass protection is disabled by default.

(5) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(6) Create a BFD session for the BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **fall-over bfd**

No BFD session is configured for a BGP neighbor by default.

1.8.6 Configuring BGP GR

1. Overview

This section describes how to configure BGP GR to ensure that data transmission is not interrupted during protocol restart and implement the deployment of highly reliable networks.

2. Restrictions and Guidelines

- To successfully deploy the BGP GR function, you need to configure a neighbor as the GR Helper.

- IGP GR also needs to be configured in an IBGP environment.
- After BGP GR is enabled, you need to reset a BGP neighbor connection to make BGP GR take effect.
- After the GR function is enabled in BGP configuration mode, the GR function will be enabled in all address families.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(5) Configure BGP GR.

bgp-instance graceful-restart

The GR function is enabled by default.

(6) Configure the BGP GR restart timer.

bgp-instance graceful-restart restart-time *restart-time*

The default timer value is **120** seconds.

(7) (Optional) Configure the BGP GR routing aging timer.

bgp-instance graceful-restart staleroute-time *time*

The default value of the aging timer is **360** seconds.

(8) (Optional) Configure the maximum wait time for EOR flags of BGP GR of all neighbors.

bgp-instance advertise-delay *delay*

The default maximum waiting time of EOR flags is **120** seconds.

(9) (Optional) Disable the GR function of BGP specified address family.

bgp-instance graceful-restart disable

By default, when the global BGP GR function is enabled, the GR function under each BGP address family is also enabled.

(10) (Optional) Configure the restart time of BGP long-lived GR.

bgp-instance graceful-restart long-lived-restart-time *time*

BGP long-lived GR is disabled by default.

(11) (Optional) Configure GR restart time for BGP peers.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **restart-time** *time* [**strictly**]

No GR restart time is configured for BGP peers by default.

1.8.7 Configuring BGP NSR

Note

This function is supported by only the S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ.

1. Overview

During switchover between master and slave supervisor engines, the NSR function keeps the network topology stable, maintains neighbor status and forwarding table, and ensures that key services are not interrupted.

2. Restrictions and Guidelines

- To deploy the BGP NSR function, rely on the device to support the function of dual engine redundancy.
- After the BGP NSR function is enabled, connections to the BGP neighbors will be reset for the function to take effect.
- The neighbor based NSR command is equivalent to the **bgp-instance nsr** command. This function enables the NSR function for all neighbors.
- You are not advised to use the neighbor-based NSR command as it is to be abandoned.
- When the BGP NSR function is enabled for a neighbor, the connection to the neighbor will be reset and the TCP none-stop-service (NSS) will be enabled for the neighbor. Then, relevant neighbor and routing information will be backed up to the slave supervisor engine. In this case, if the NSR function is enabled after a BGP speaker is in the established state, the neighbor relationship will be re-established, causing neighbor flapping. You are advised to enable NSR before a neighbor relationship is established between peers.
- The NSR function does not take effect on a neighbor that fails to establish a BGP connection. That is, if the BGP connection of a neighbor is not in the established state, the neighbor will not execute the NSR operation. Only a neighbor in the established state will perform NSR-related backup operations.
- Supporting the BGP NSR function does not mean that the device can perform switchovers between the active and standby management boards. To perform such switchovers, you must depend on the device hardware. NSR devices must support the function of dual engine redundancy.
- If the BGP or TCP backup is not complete, the current BGP NSR switchover may fail. When a NSR switchover failure occurs, routes can be still recovered through BGP GR. Therefore, you are advised to enable both BGP NSR and BGP GR.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the NSR function of BGP.

bgp-instance nsr [stable-mode]

By default, the NSR function is disabled.

(5) (Optional) Enable the NSR function of the specified BGP peer (group).

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **ha-mode nsr** [**stable-mode**]

By default, the NSR function of BGP peer (group) is disabled.

1.8.8 Configuring BGP Session Keepalive

1. Overview

For devices that support BGP, routing errors in an address family will affect route stability in other address families: When a device receives an update packet from a neighbor and an error on the multi-protocol routing attribute is detected, the device will disconnect the BGP session. As a result, flapping occurs on routes in all address families of this neighbor.

After the BGP session retention function is enabled, if an error occurs in the routing attribute of an address family, only the routing information related to the neighbor in this address family is deleted. The BGP session and other address families are not affected, which enhances the stability of BGP.

Recovery-time specifies the waiting time for automatic route recovery. A neighbor must support the route-refresh capability so that the time can be set. After the recovery-time elapses, BGP sends the route-refresh message of the address family to the neighbor and re-advertises all routing information in the address family to this neighbor.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure BGP to retain BGP sessions when it detects a multiprotocol path error.

bgp-instance mp-error-handle session-retain [**refresh-timer** *refresh-timer*]

BGP terminates BGP sessions by default when it detects a multiprotocol route error.

1.8.9 Configuring EBGP Fast Neighbor Detection

1. Overview

After the EBGP neighbor fast detection is configured, the device will quickly close the BGP session connection when the network interface for directly connecting the EBGP neighbor fails.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Enable the EBGP neighbor fast link detection function.

bgp-instance ebgp-interface-fallover

By default, the EBGP neighbor fast link detection function is enabled.

1.8.10 Configuring BGP Link-State Group Association

1. Overview

Rapid Ethernet Uplink Protection (REUP) is a fast uplink protection function. In the dual-uplink networking mode, REUP is used to ensure normal communication of links, block redundant links, avoid link loops, acting as a fast backup. REUP binds the link state of multiple downlink interfaces to the interface state of multiple uplinks by configuring the downlink and uplink interfaces of the link-state group. When all uplink links in the group fail, all downlink interfaces in the group are forced to be down. However, in the actual scenario, the state of the uplink interface may be normal, but the BGP connection established through the uplink interface is disconnected. In this case, REUP will not close the downlink, and the device still needs to handle the downlink traffic. Since the BGP connection has been disconnected and there is no corresponding routing table entry on the device, the device will discard this part of traffic. If the device can automatically close the downlink interface when it detects that the BGP connection is disconnected, the device will not need to process the downlink traffic.

After the BGP link-state group association function is configured, if all BGP neighbors associated with this link-state group are disconnected, the device will automatically close all the downlink interfaces in the link-state group to avoid wasting device resources for processing traffic.

2. Prerequisites

The link-state group of REUP must be configured. See *Configuring REUP* for details.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Associate a BGP peer (group) with a link-state group.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **link-state group** *link-state-group-number*

By default, a BGP peer (group) is not associated with a link-state group.

(5) (Optional) Configure the delay time for disassociating the BGP neighbor with the link-state group due to link failure.

bgp-instance link-state-group up-delay *delay-time*

By default, the delay time for disassociating the BGP neighbor with the link-state group due to link failure is 5 seconds.

1.9 Configuring MP-BGP

1.9.1 Configuration Tasks

1. EVPN

The configuration tasks of BGP EVPN function are as follows:

- Configuring an EVI
- Configuring EVPN

1.9.2 Configuring an EVI

1. Overview

This section describes how to configure an EVPN instance (EVI).

2. Restrictions and Guidelines

- When the number of configured VNIs exceeds the capacity or some VNI instances are being deleted, the configuration fails.
- The RD values of different VNI instances must be different. Therefore, VNI instances must be configured in batches by running only the **rd auto** command.
- If an EVI is defined and an RD value has been configured for it, the RD value cannot be modified. To modify the RD value, delete the EVI and then configure a new RD. One EVI instance can have only one RD value.
- RD values automatically generated are in the format of ipv4-address:nn. In the format, "ipv4-address" indicates the BGP router-ID and *nn* is the index value allocated by the BGP software. The value generated by the **rd auto** command keeps unchanged after generation, and it remains unchanged even if the BGP router-ID changes.
- An EVI instance can be configured with multiple imported and exported RT values.
- The automatically generated RT value is in the format of AS2:nn. In the format, "AS2" indicates the 2-byte AS number. If a 4-byte AS number is used, it is truncated to two bytes and then placed in the RT value; "nn" indicates the VNI-ID, which occupies four bytes and can fully accommodate the complete the VNI-ID value.
- If the BGP AS number changes, the automatically generated RT value also changes.
- If the manually configured RT value is the same as the automatically generated RT value, both values are displayed.
- Automatically generated RT values cannot be deleted even if you configure an RT value deletion command. Instead, you need to run the **unset route-target auto** command to delete the configuration.
- The routing policy in an EVI supports only one **route-map**, and the new configuration will overwrite the earlier configuration if multiple route maps are configured.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the EVPN function and enter the EVPN configuration mode.

evpn

The EVPN function is disabled by default.

(4) Enter the EVI configuration mode. Configure one of the following tasks.

- Create an EVI.
  - **vni** *vni-id*
  - No EVI is configured by default.
- Create EVIs in batches.
  - **vni range** *vni-id-list*

(5) Configure an RD.

rd { **auto** | *rd-value* }

No RD is configured for an EVI by default.

(6) Configure an RT.

routing-target { **both** | **import** | **export** } { **auto** | *rt-value* }

No RT value is configured for an EVI instance by default.

(7) (Optional) Configure a routing policy for the EVI. Configure one of the following tasks.

- Configure policy rules for extended community attributes of EVPN routes distributed from the local end to the remote end.
  - **export map** *route-map-name*
  - No policy rule is configured by default for extended community attributes of EVPN routes distributed from the local end to the remote end.
- Filter out remote EVPN routes that are imported to the local VNI instance, or modify the attribute of the remote EVPN routes to be imported to the local VNI instance.
  - **import map** *route-map-name*
  - No routing policy rule is configured for importing remote EVPN routes to the local VNI by default.

(8) (Optional) Configure the EVPN Type 5 route mode.

ip-prefix model { **interface-less** | **interface-full** { **sbd-irb** | **unnumber-sbd-irb** } }

By default, EVPN Type 5 routes support the Interface-Less mode.

1.9.3 Configuring EVPN

1. Overview

This section describes how to configure a policy related to the L2VPN EVPN address family.

2. Restrictions and Guidelines

- Mobility may occur on one MAC address many times. When a provider edge router (PE) detects an MAC address mobility event through local learning, it will start a timer of *mac-mobility-interval* (the default value is **180** seconds). If the number of detected MAC address mobility events exceeds the value of *mac-mobility-threshold* (the default value is **5**) before the timer times out, it is considered that a MAC address conflict occurs. After a MAC address conflict occurs, the EVPN route corresponding to the MAC address is not advertised to a neighbor and a delay recovery timer with the value of *retry-timer* (**180** seconds by default) is started. When the timer expires, the MAC address conflict is deleted and the EVPN route

corresponding to the MAC address is advertised again.

- Generally, the inner MAC address of an EVPN packet is the router MAC address of the egress gateway. You can also modify attributes to configure the anycast MAC address as the inner MAC address of EVPN packets.
- If the VNI instance, to which an EVPN route belongs, has no corresponding VNI instance on the local device, the modified attributes do not take effect.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Create a BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **remote-as** *as-number*

(5) Configure the L2VPN EVPN address family mode.

address-family-mode l2vpn evpn

BGP neighbors are not allowed to exchange L2VPN information by default.

(6) Activate the EVPN capability for the BGP neighbor.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **activate**

The EVPN address family is disabled by default.

(7) (Optional)Configure MAC mobility detection parameters.

bgp-instance mac-mobility *mac-mobility-interval mac-mobility-threshold* [*retry-timer*]

By default, the timer time of MAC address move detection is set to **180** seconds, the detection times is **5**, and the delayed timer time of MAC conflict restoration is **180** seconds.

(8) (Optional)Set the router MAC attribute of EVPN routing information to the router MAC address or anycast MAC address of the local device for EVPN VXLAN scenarios across data centers.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **router-mac-self**

The router MAC attribute of EVPN routes is not modified when the routes are distributed to a specified BGP peer by default.

(9) (Optional) Configure to advertise only VNI Mapping table entries to a specified peer.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **send-mapping-only**

By default, advertising only VNI Mapping table entries to a specified peer is not configured.

Note

This command is not supported by the S6980-64QC, S6920-4C, or S6580-24DC8QC switch.

(10) (Optional) Configure to redistribute IPv4 unicast route.

advertise ipv4 unicast-af-modifier

Redistributing IPv4 unicast route is not configured by default.

(11) Configure to redistribute IPv6 unicast route.

advertise ipv6 unicast-af-modifier

Redistributing IPv6 unicast route is not configured by default.

1.10 Configuring BGP Neighbor Auto-Discovery

1.10.1 Configuration Tasks

The BGP neighbor auto-discovery configuration includes the following tasks:

- Configuring BGP Neighbor Auto-Discovery
- Configuring Interface Neighbor Auto-Discovery

1.10.2 Configuring BGP Neighbor Auto-Discovery

1. Overview

This section describes how to configure BGP neighbor auto-discovery.

2. Restrictions and Guidelines

- If an auto-discovered neighbor conflicts with a common neighbor or network segment neighbor, retain the neighbor based on the priority. Priorities: Common neighbor > auto-discovered neighbor > network segment neighbor.
- The **peer update-source** command cannot be configured for neighbor groups used by interfaces.
- Interfaces with BGP neighbor auto-discovery enabled send BGP hello packets and learn peer routes through negotiation. The **distance** command is used to change the administrative distance of routes discovered by the BGP auto-discovery mechanism.
- You can run **exit**-**auto-discovery** command to exit the BGP neighbor auto-discovery configuration mode.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Create a BGP neighbor group.

peer *peer-group-name* **peer-group**

The neighbor group is used when a neighbor is created for an interface.

(5) Enable the neighbor auto-discovery function and enter the neighbor auto-discovery configuration mode.

auto-discovery

The neighbor auto-discovery function is disabled by default.

(6) (Optional) Configure the remote neighbor AS number list acceptable by BGP neighbor auto-discovery.

accept-as-list *as-list*

By default, there is no restriction on the remote neighbor AS number list acceptable by BGP neighbor auto-discovery.

(7) (Optional) Configure the administrative distance for routes learned through auto-discovery.

distance *distance*

The default administrative distance of a route learned through auto-discovery is **18**.

1.10.3 Configuring Interface Neighbor Auto-Discovery

1. Overview

This section describes how to configure interface neighbor auto-discovery.

2. Restrictions and Guidelines

- Neighbor auto-discovery can be configured only on the L3 interfaces and cannot be configured on loopback interfaces.
- Neighbor auto-discovery can take effect on an interface only after an IPv4/IPv6 address or the **ipv6 enable** command is configured for the interface.
- *as-number* must be the same as the AS number of a BGP instance.
- The **peer update-source** command cannot be configured on a specified peer group of an interface.
- IBGP and EBGP neighbors need to use different peer groups.
- The value of *hello-interval* cannot be greater than 1/3 of the value of *hold-time*.
- When *hold-time* of hello packets is up, the BGP auto-discovered peer disconnects the session.
- When **source-interface** is set to a loopback interface, the IPv6 address sent to the peer for establishing a session uses the smallest IPv6 address configured for the loopback interface. After a large IPv6 address has been configured for the loopback interface and the peer has used the address to establish a BGP session, if a small IPv6 address is configured for the loopback interface, the auto-discovery mechanism sends the small IPv6 address to the peer. Then, the peer uses the small IPv6 address to establish a BGP session and deletes the previously auto-discovered neighbor.
- When **source-interface** is set to **source-self**, the IPv6 address sent to the peer for establishing sessions is the IPv6 LinkLocal address of the interface to establish sessions with the peer.
- When *source-interface* is set to a loopback interface and the loopback interface uses an IPv4 address, the layer 3 interface needs to use the IPv4 address or the IPv6 address (**ipv6 enable** *and* **ip l3protocol default-up**)

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure a local AS number for the BGP auto-discovered peer.

bgp-instance auto-discovery local-as *as-number*

No local AS number is configured for BGP auto-discovered peers by default.

(5) Configure the source interface and peer group to be used for establishing a BGP connection with a peer.

bgp-instance auto-discovery source-interface { *interface-type interface-number* | **source-self** } **peer-group** *peer-group-name* [**ipv4** | **ipv6**]

By default, no source interface or peer group is specified for establishing a BGP connection with a peer.

(6) (Optional) Configure the timer duration of the BGP neighbor auto-discovery.

bgp-instance auto-discovery hello-period *hello-interval* **hold-time** *hold-time*

The default interval for BGP to send hello messages is **15** seconds and the default hold time is **180** seconds.

1.11 Configuring BGP Extended Features

1.11.1 Configuration Tasks

The configuration of BGP extended functions includes the following tasks: All configuration tasks are optional and may be selected as needed.

- Configuring BMP Monitoring
- Configuring BGP Administrative Distance
- Configuring Enhanced VPN
- Configuring BGP Link-State

1.11.2 Configuring BMP Monitoring

1. Overview

After BMP monitoring is configured, BMP is monitored by the BMP server, which focuses on the status of BGP peers, peer entry transmission and receiving, and all kinds of peer statistics.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure neighbor to be monitored by the specified BMP server.

peer { *neighbor-ipv4-address* | *neighbor-ipv6-address* | *peer-group-name* } **bmp-active** { **all** | **server** *server-number*&<1-8> }

BGP peers are not monitored by the BMP server by default.

3. Configure BMP server instances.

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a BMP server instance.

bmp server *server-number*

(4) Configure the address and port number of the BMP server.

address [*bmp-server-ipv4-address* | *bmp-server-ipv6-address*] **port** *port-number*

The address and port number of the BMP server are not configured in BMP mode by default.

(5) Configure the MD5 authentication password for establishing a TCP connection with the BGP server.

password [**0** | **7**] *string*

By default, the MD5 authentication password for TCP connections to the BGP server is not configured.

(6) Configure an outbound interface for establishing a TCP connection with the BMP server.

update-source *interface-type interface-number*

The system automatically selects the outbound interface for a TCP connection based on the IP address of the BMP server by default. The outbound interface of local packets is generally used.

(7) Configure the VRF instance, to which a TCP connection to the BMP server belongs.

vrf *vrf-name*

A TCP connection is established with the BMP server in the default VRF instance by default.

(8) (Optional) Configure the interval for re-establishing a TCP connection with the BMP server.

bmp-reconnect-delay *time*

The interval for reconnecting to the BMP server gradually increases from the initial **30** seconds by default.

(9) (Optional) Configure a description for the BMP server instance.

description *text*

No description is configured for a BMP server instance by default.

(10) Configure the BMP server to monitor routes that are received from peers and have a routing policy applied.

adj-rib-in post-policy

The BMP server does not monitor routes that are received from a BGP peer and have a routing policy applied by default.

(11) Configure the BMP server to monitor routes that are received from a peer and are not renamed.

adj-rib-in pre-policy

The BMP server does not monitor routes that are received from a BGP peer and are not renamed by default.

(12) Configure the BMP server to monitor routes that are sent to a peer and have a routing policy applied.

adj-rib-out post-policy

The BMP server does not monitor routes that are sent to a BGP peer and have a routing policy applied by default.

(13) Configure the BMP server to monitor routes that are sent to a peer before a routing policy is applied.

adj-rib-out pre-policy

The BMP server does not monitor routes that are sent to a BGP peer before a routing policy is applied by default.

(14) Configure the monitoring of local-rib routes for IPv4 or IPv6.

local-rib { **ipv4** | **ipv6** } **unicast-af-modifier** [**vrf** *vrf-name*] [**add-paths**] [**route-map** *map-tag*]

The BMP server does not monitor local-rib routes for IPv4 or IPv6 by default.

(15) Enable the packet mirroring function of BGP.

route mirroring

The BMP server does not send route mirroring packets by default.

(16) Configure the interval for BMP server to send status statistics.

stats-reporting-period *time*

When status statistics of a monitored peer are changed, BMP delays a period of time in sending the status statistics to the BMP server by default.

(17) (Optional) Configure the maximum number of cached packet bytes of a BMP instance.

buffer-size *buffer-size-maximum* [**warning**]

1.11.3 Configuring BGP Administrative Distance

1. Overview

The administrative distance is used to evaluate the reliability of various route sources. A smaller administrative distance indicates a better route.

- Administrative distance of BGP

  The administrative distance indicates the reliability of a route source. The value range is from 1 to 255. A larger administrative distance indicates a lower route reliability. BGP sets different administrative distances for sources, from which routing information is learned. Administrative distances are classified into external-distances, internal-distances, and local-distances.

  - External-distance: Indicates the administrative distances of routes learned from EBGP peers.
  - Internal-distance: Indicates the administrative distances of routes learned from IBGP peers.
  - Local-distance: Indicates the administrative distances of routes learned from peers but it is considered that better routes can be learned from IGP. Generally, these routes can be indicated by the **network-advertise backdoor-route** command.

- Backdoor route

  If you prefer an IGP route without using this EBGP route, you can configure this EBGP route as the backdoor route. The default administrative distance of routes learned from a BGP speaker that has established an EBGP connection is **20**. You can run the **network-advertise backdoor-route** command to set the administrative distance of EBGP routes to 200 so that routes learned from IGP have a higher priority. Routes learned from IGP are considered as backdoor routes and are not advertised externally.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable BGP to enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Configure the administrative distance of a BGP route.

distance bgp *external-distance internal-distance local-distance*

The default administrative distance of a route learned by BGP from an EBGP peer is **20** and that of a route learned by BGP from an IBGP peer is **200**.

(5) Configure a backdoor route.

network-advertise *network-number* [**mask** *mask*] [**route-map** *map-tag*] **backdoor-route**

No backdoor route is configured by default.

1.11.4 Configuring Enhanced VPN Route Import

1. Overview

By default, when VRF routes are imported into each other, L3VPN remote routes are imported into VRF, or EVPN routes are imported into the IP routing table, only the routes with a preferred next hop are imported. If the enhanced VPN route importing function is configured, all next-hop routes or equal-cost next-hop routes can be imported. The imported routes can realize ECMP multi-path load sharing.

2. Restrictions and Guidelines

- IBGP routes and EBGP routes cannot be equal-cost routes.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) (Optional) Enter the corresponding address family configuration mode. Configure one of the items below for a specific address family.

(5) Enter the IPv4 unicast configuration mode.

address-family-mode ipv4 [**unicast**]

(6) Enter the IPv4 VRF configuration mode.

address-family-mode ipv4 vrf *vrf-name*

(7) Enter the IPv6 unicast configuration mode.

address-family-mode ipv6 [**unicast**]

(8) Enter the IPv6 VRF configuration mode.

address-family-mode ipv6 vrf *vrf-name*

(9) Configure the route import policy.

import route strategy { **all** | **optimal-path** | **multipath** }

By default, only preferred routes are imported.

1.11.5 Configuring BGP Link-State

1. Overview

BGP Link-State can collect link state information and summarize it to the controller.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the BGP protocol, and enter the BGP configuration mode.

router-protocol bgp *as-number*

(4) Allow BGP neighbors to interact routing information with each other, and enter the Link-State address family configuration mode.

address-family-mode link-state [**unicast**]

By default, the Link-State address family is not configured.

(5) Configure the AS field and Router ID field in the LS table entries advertised by BGP.

domain-distinguisher *ipv4-address:nn*

By default, the AS field and Router ID field in the LS table entries advertised by BGP are the current AS number and Router ID of BGP.

1.12 Monitoring

Run the display commands to check the running status of a configured function to verify the configuration effect.

Run the cleanup commands to clear information.

⚠ Caution

Run the **cleanup** command during operation of the device may lose vital information and interrupt services.

Run the debugging commands to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging switch immediately after use.

Table 1-3 Monitoring

Command	Purpose
cleanup bgp-instance all [*as-number*] [**soft**] [**inbound** | **outbound**]	Restores the priority of the route advertised by BGP to the neighbor to the priority before it is adjusted to the lowest priority.
cleanup bgp-instance all peer-group *peer-group-name* [**soft**] [**inbound** | **outbound**]	Clears all address families of BGP.
cleanup bgp-instance all peer-group *peer-group-name* [**soft**] [**inbound** | **outbound**]	Clears a specified peer group in BGP.
cleanup bgp-instance all update-group-info [*neighbor-ipv4-address* | *neighbor-ipv6-address* | *update-group-index*] [**soft**] [**inbound** | **outbound**]	Clears sessions of all members in the update group.
cleanup bgp-instance ipv4 unicast-af-modifier [**vrf** vrf-name] { ***** | *as-number* | *neighbor-ipv4-address* | *neighbor-ipv6-address* } [**soft**] [**inbound** | **outbound**] **cleanup ip bgp-instance** [**vrf** *vrf-name*] { ***** | *as-number* | *neighbor-ipv4-address* | *neighbor-ipv6-address* } [**soft**] [**inbound** | **outbound**]	Clears specified sessions of BGP IPv4 unicast address family.
cleanup bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **damp** [*ipv4-address* [*mask*] **cleanup ip bgp-instance** [**vrf** *vrf-name*] **damp** [*ipv4-address* [*mask*]	Clears IPv4 unicast address family damping information and route suppression.
cleanup bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **external** [**soft**] [**inbound** | **outbound**] **cleanup ip bgp-instance** [**vrf** *vrf-name*] **external** [**soft**] [**inbound** | **outbound**]	Clears EBGP connections of all IPv4 unicast address families.
cleanup bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **flap-statistics** [ipv4-*address* [*mask*] **cleanup ip bgp-instance** [**vrf** *vrf-name*] **flap-statistics** [*ipv4-address* [*mask*]	Clears routing flap statistics of the IPv4 unicast address family.
cleanup bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **peer-group** *peer-group-name* [**soft**] [**inbound** | **outbound**] **cleanup ip bgp-instance** [**vrf** *vrf-name*] **peer-group** *peer-group-name* [**soft**] [**inbound** | **outbound**]	Clears sessions of all members in the IPv4 unicast address family peer group.

cleanup bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **table-map** **cleanup ip bgp-instance** [**vrf** *vrf-name*] **table-map**	Clears and updates the Table-map settings of BGP IPv4 unicast address family.
cleanup bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **update-group** [*neighbor-ipv4-address* | *neighbor-ipv6-address* | *update-group-index*] [**soft**] [**inbound** | **outbound**] **cleanup ip bgp-instance** [**vrf** *vrf-name*] **update-group-info** [*neighbor-ipv4-address* | *neighbor-ipv6-address* | *update-group-index*] [**soft**]] [**inbound** | **outbound**]	Clears sessions of all members in the update group of the IPv4 unicast address family.
cleanup bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] { ***** | *as-number* | *neighbor-ipv4-address* | *neighbor-ipv6-address* } [**soft**] [**inbound** | **outbound**]	Clears specified sessions of BGP IPv6 unicast address family.
cleanup bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **damp** [*ipv6-address/prefix-length*]	Clears IPv6 unicast address family damping information and route suppression.
cleanup bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **external** [**soft**] [**inbound** | **outbound**]	Clears EBGP connections of all IPv6 unicast address families.
cleanup bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **flap-statistics** [*ipv6-address/prefix-length*]	Clears routing flap statistics of the IPv6 unicast address family.
cleanup bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **peer-group** *peer-group-name* [**soft**] [**inbound** | **outbound**]	Clears sessions of all members in the IPv6 unicast address family peer group.
cleanup bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **table-map**	Clears and updates the Table-map settings of BGP IPv6 unicast address family.
cleanup bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **update-group** [*neighbor-ipv4-address* | *neighbor-ipv6-addres* | *update-group-index*] [**soft**] [**inbound** | **outbound**]	Clears sessions of all members in the update group of the IPv6 unicast address family.
cleanup bgp-instance l2vpn evpn-af-modifier { ***** | *as-number* | *neighbor-ipv4-address* | *neighbor-ipv6-addres* } [**soft**] [**inbound** | **outbound**]	Clears specified sessions of L2VPN EVPN address family.
cleanup bgp-instance l2vpn evpn-af-modifier damp	Clears L2VPN EVPN address family damping information and route suppression.
cleanup bgp-instance l2vpn evpn-af-modifier external [**soft**] [**inbound** | **outbound**]	Clears EBGP connections of all L2VPN EVPN address families.

cleanup bgp-instance l2vpn evpn-af-modifier flap-statistics	Clears routing flap statistics of the L2VPN EVPN address family.
cleanup bgp-instance l2vpn evpn-af-modifier peer-group *peer-group-name* [**soft**] [**inbound** | **outbound**]	Clears sessions of all members in the L2VPN EVPN address family peer group.
cleanup bgp-instance l2vpn evpn-af-modifier update-group [*neighbor-ipv4-address* | *neighbor-ipv6-address* | *update-group-index*] [**soft**] [**inbound** | **outbound**]	Clears sessions of all members in the update group of the L2VPN EVPN address family.
cleanup evpn conflict mac [*vni-id*]	Clears EVPN conflict MAC address information.
cleanup bgp-instance link-state unicast-af-modifier { ***** | *as-number* | *neighbor-ipv4-address* | *neighbor-ipv6-addres* } [**soft**] [**inbound** | **outbound**]	Clears specified sessions of the Link-State address family.
cleanup bgp-instance link-state unicast-af-modifier external-type [**soft**] [**inbound** | **outbound**]	Clears EBGP connections of all Link-State address families.
cleanup bgp-instance link-state unicast-af-modifier ls-srv	Clears LS server connections of the Link-State address family.
cleanup bgp-instance link-state unicast-af-modifier peer-group *peer-group-name* [**soft**] [**inbound** | **outbound**]	Clears sessions of all members in the Link-State address family peer group.
cleanup bgp-instance link-state unicast-af-modifier update-group [*neighbor-ipv4-address* | *neighbor-ipv6-address* | *update-group-index*] [**soft**] [**inbound** | **outbound**]	Clears sessions of all members in the update group of the Link-State address family.
cleanup bmp { **all** | **server** *server-number* }	Clears BMP protocol.

display bgp-instance all [**aggregate** | **community** [*community-number*&<1-n> [**match**]] | **comm-type-list** *community-name* [**match**] | **extend-comm-list** *extcommunity-name* [**match**] | **filter-rule-list** *path-list-number* | **inconsistent-as** | **quoted-regular-expression** *regexp* | **regexp** *regexp*] **display bgp-instance all damp** { **damp-paths** | **flap-statistics** | **arguments** } **display bgp-instance all peers**[{ *neighbor-ipv4-address* | *neighbor-ipv6-address* } [**advertised-routes** [**check**] | **ha-mode** [**adj-in** [**detail**] | **adj-out**] | **hide-info** | **policy-based** [**detail**] | **received-routes** | **routes**]] **display bgp-instance all paths** **display bgp-instance all status** **display bgp-instance all update-group** [*neighbor-ipv4-address* | *neighbor-ipv6-address* | *update-group-index*] [**ha-mod adj-out** | **status**]	Displays all BGP routing information.
display bgp-instance auto-discovery [**interface** *interface-type interface-number*]	Displays BGP neighbor auto-discovery information.
display bgp-instance bmp { **peer** | **server** [*server-number*] [**detail**] | **status** }	Displays BMP server information.
display bgp-instance develop { **attr** { **color-nexthop** | **prefix-sid** } | **bgp-instance** | **cap** | **connected** | **evi-rd-hash** | **ifx-link-group** | **interface nd** [*ifx-number* [*mac-address*]] | **io-process** [**log-info** | **log-warn** | **master** | **memory** | **peer** [*neighbor-ipv4-address* [*vrf-name*] | *neighbor-ipv6-address* [*vrf-name*] | *peer-id*]] | **link-state-group** | **mom** | **peer-as** | **pic-info** | **route-shake** | **segment-routing policy** [**color** *color-number*] [**endpoint** *endpoint-address*] | **segment-routing ipv6 local-sid** [**sid** *sid-address*] | **segment-routing ipv6 local-sid-retry** | **thread** | **unspec-instance-redist-conf** | **vr** | **vrf-rd-hash** }	Displays BGP development information.
display bgp-instance evpn-af-modifier [**evi** [*vni-id*] | **evi-hash** | **local-mac** [*vni-id*] [**detail**] | **remote-mac** [*vni-id*] [**detail**] | **route-list** { **ethernet-ad** | **ethernet-segment** | **inclusive-multicast** } [*vni-id*] | **share-ip** [*vni-id*] [**detail**] | **status** | **vtep-nbr** [*vni-id*]]	Displays EVPN related information.
display bgp-instance grst { **bfd** | **evi** | **label** | **label-set** | **mdc** | **peer-info** | **status** }	Displays GR related information.
display bgp-instance hash-peer	Displays Hash tables of all neighbors.

display bgp-instance ipv4 **unicast-af-modifier** [*prefix-ipv4-address* [*network-mask* [**longer**-**prefixes**]]	**aggregate**	**cidr**-**only**	**community** [*community-number*&<1-n> [**match**]]	**comm-type**-**list** *community-name* [**match**]	**extend-comm**-**list** *extcommunity-name* [**match**]	**filter**-**rule-list** *path-list-number*	**global-vrf detail**	**inconsistent**-**as**	**labels**	**prefix**-**list** *ip-prefix-list-name*	**quoted-regular-expression** *regexp*	**regexp** *regexp*	**route-map** *map-tag*] **display bgp-instance ipv**4 **unicast-af-modifier vrf** *vrf-name* [*prefix-ipv4-address* [*network-mask* [**longer**-**prefixes**]]	**aggregate**	**cidr**-**only**	**community** [*community-number*&<1-n> [**match**]]	**comm-type**-**list** *community-name* [**match**]	**detail** [**all**]	**extend-comm**-**list** *extcommunity-name* [**match**]	**filter-rule**-**list** *path-list-number*	**inconsistent**-**as**	**prefix**-**list** *ip-prefix-list-name*	**quoted-regular-expression** *regexp*	**regexp** *regexp*	**route-map** *map-tag*]]	Displays IPv4 unicast routing information in BGP routes.
display bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **damp** { **damp-paths**	**flap-statistics**	**arguments** }	Displays BGP IPv4 unicast routing dampening information.
display bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **peers**[{ *neighbor-ipv4-address*	*neighbor-ipv6-address* } [**advertised**-**routes** [**check**]	**ha-mode** [**adj-in** [**detail**]	**adj-out**]	**hide-info**	**policy** [**detail**]	**received**-**routes**	**routes**]]	Displays BGP IPv4 unicast neighbor related information.
display bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **paths**	Displays IPv4 unicast paths in the route database.
display bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **status**	Displays IPv4 unicast paths in route database.
display bgp-instance ipv4 unicast-af-modifier [**vrf** *vrf-name*] **update-group-info** [*neighbor-ipv4-address*	*neighbor-ipv6-address*	*update-group-index*] [**ha-mod adj-out**	**status**]	Displays BGP IPv4 unicast update group related information.

display bgp-instance ipv6 **unicast-af-modifier** [*prefix-ipv6-address/prefix-length* [**longer**-**prefixes**]	**aggregate**	**community** [*community-number*&<1-n> [**match**]]	**comm-type**-**list** *community-name* [**match**]	**extend-comm**-**list** *extcommunity-name* [**match**]	**filter**-**rule-list** *path-list-number*	**global-vrf detail**	**inconsistent**-**as**	**labels**	**prefix**-**list** *ip-prefix-list-name*	**quoted-regular-expression** *regexp*	**regexp** *regexp*	**route-map** *map-tag*] **display bgp-instance ipv6 unicast-af-modifier vrf** *vrf-name* [*prefix-ipv6-address/prefix-length* [**longer-prefixes**]	**aggregate**	**community** [*community-number*&<1-n> [**match**]]	**comm-type-list** *community-name* [**match**]	**detail** [**all**]	**extend-comm-list** *extcommunity-name* [**match**]	**filter-rule-list** *path-list-number*	**inconsistent-as**	**prefix-list** *ip-prefix-list-name*	**quoted-regular-expression** *regexp*	**regexp** *regexp*	**route-map** *map-tag*]	Displays IPv6 unicast routing information in BGP routes.
display bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **damp** { **damp-paths**	**flap-statistics**	**arguments** }	Displays BGP IPv6 unicast routing dampening information.
display bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **peers**[{ *neighbor-ipv4-address*	*neighbor-ipv6-address* } [**advertised**-**routes** [**check**]	**ha-mode** [**adj-in** [**detail**]	**adj-out**]	**hide-info**	**policy** [**detail**]	**received**-**routes**	**routes**]]	Displays BGP IPv6 unicast neighbor related information.
display bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **paths**	Displays IPv6 unicast paths in route database.
display bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **status**	Displays BGP IPv6 unicast summary information.
display bgp-instance ipv6 unicast-af-modifier [**vrf** *vrf-name*] **update-group** [*neighbor-ipv4-address*	*neighbor-ipv6-address*	*update-group-index*] [**ha-mod adj-out**	**status**]	Displays BGP IPv6 unicast update group related information.

display bgp-instance l2vpn evpn-af-modifier { **evi** *vni-id* | **rd** *rd-value* } [{ **ethernet-ad** | **ethernet-segment** | **inclusive-multicast** [*ipv4-address* | *ipv6-address*] | **ip-prefix** [*ipv4-address* | *ipv6-address*] | **mac-ip** [*mac-address*] [*ipv4-address* | *ipv6-address*] } [**detail**]] **display bgp-instance l2vpn evpn-af-modifier all** [{ **ethernet-ad** | **ethernet-segment** | **inclusive-multicast** [*ipv4-address* | *ipv6-address*] | **ip-prefix** [*ipv4-address* | *ipv6-address*] | **mac-ip** [*mac-address*] [*ipv4-address* | *ipv6-address*] } [**from-neighbor** { *neighbor-ipv4-address* | *neighbor-ipv6-address* }] [**detail**]] **display bgp-instance l2vpn evpn-af-modifier all** { **peer** [{ *neighbor-ipv4-address* | *neighbor-ipv6-address* } [**ha-mode** { **adj-in** [**detail**] | **adj-out** } | **policy** [**detail**]]] | **status** } **display bgp-instance l2vpn evpn-af-modifier all update-group** [*neighbor-ipv4-address* | *neighbor-ipv6-address* | *update-group-index*] [**ha-mod adj-out** | **status**]]	Displays specific routes of BGP L2VPN EVPN address family.
display bgp-instance link-state unicast-af-modifier damp { **damp-paths** | **flap-statistics** | **arguments** } **display bgp-instance link-state unicast-af-modifier** [**detail** [*prefix-string*] | **ls-client** | **status**] **display bgp-instance link-state unicast-af-modifier peers**[{ *neighbor-ipv4-address* | *neighbor-ipv6-address* } [**advertised-routes** [**check**] | **ha-mode** { **adj-in** [**detail**] | **adj-out** } | **hide-info** | **policy** [**detail**] | **received-routes** | **routes**]] **display bgp-instance link-state unicast-af-modifier update-group** [*neighbor-ipv4-address* | *neighbor-ipv6-address* | *update-group-index*] [**ha-mod adj-out** | **status**]	Displays specific routes of BGP Link-State address family.
display bgp-instance log-info	Displays BGP logs.
display bgp-instance log-warn	Displays BGP alarms.
display bgp-instance memory	Displays BGP memory.
display bgp-instance nsr [**info** | **sock-cb** [*sock-cb-id*] | **sock-list** | **status**]	Displays NSR related information.
display bgp-instance route-block	Displays black-hole route statistics.
display bgp-instance rpi { **acl** [**detail** [*acl-name*]] | **as-access-list** | **comm-type-list** | **extend-comm--list** | **ip-prefix-list** | **ipv6-prefix-list** | **route-map** }	Displays RPI policy related information.
display bgp-instance statistics [**vrf** *vrf-name*]	Displays BGP statistics.
display evpn [*vni-id]* [**detail**]	Displays configured EVI instances.

display evpn mac { **conflict** | **mobility** } [*vni-id*]	Displays MAC addresses mobility and conflicts.
display ip bgp-instance [*prefix-ipv4-address* [*network-mask* [**longer-prefixes**]] | **aggregate** | **cidr-only** | **community** [*community-number*&<1-n> [**match**]] | **comm-type-list** *community-name* [**match**] | **extend-comm-list** *extcommunity-name* [**match**] | **filter-rule-list** *path-list-number* | **global-vrf detail** | **inconsistent-as** | **labels** | **prefix-list** *ip-prefix-list-name* | **quoted-regular-expression** *regexp* | **regexp** *regexp* | **route-map** *map-tag*] **display ip bgp-instance vrf** *vrf-name* [*prefix-ipv4-address* [*network-mask* [**longer-prefixes**]] | **aggregate** | **cidr-only** | **community** [*community-number*&<1-n> [**match**]] | **comm-type-list** *community-name* [**match**] | **detail** [**all**] | **extend-comm-list** *extcommunity-name* [**match**] | **filter-rule-list** *path-list-number* | **inconsistent-as** | **prefix-list** *ip-prefix-list-name* | **quoted-regular-expression** *regexp* | **regexp** *regexp* | **route-map** *map-tag*]] **display ip bgp-instance** [**vrf** *vrf-name*] **damp** { **damp-paths** | **flap-statistics** | **arguments** } **display ip bgp-instance** [**vrf** *vrf-name*] **neighbors** [{ *neighbor-ipv4-address* | *neighbor-ipv6-address* } [**advertised-routes** [**check**] | **ha-mode** [**adj-in** [**detail**] | **adj-out**] | **hide-info** | **policy** [**detail**] | **received-routes** | **routes**]] **display ip bgp-instance** [**vrf** *vrf-name*] **paths** **display ip bgp-instance** [**vrf** *vrf-name*] **status** **display ip bgp-instance** [**vrf** *vrf-name*] **update-group** [*neighbor-ipv4-address* | *neighbor-ipv6-address* | *update-group-index*] [**ha-mod adj-out** | **status**] **display ip bgp-instance scan** [**detail**]	Displays routes in BGP IPv4 unicast address family.
display bgp key-chain [*keychain-name*]	Displays BGP Key-chain information.
debugging ip bgp all	Debugs BGP.
debugging ip bgp auto-discovery	Debugs BGP neighbor auto-discovery.
debugging ip bgp bmp	Debugs BGP BMP.
debugging ip bgp damp	Debugs BGP route dampening.

debugging ip bgp event-info **debugging ip bgp event-infoagg** **debugging ip bgp event-infofrr** **debugging ip bgp event-infogrst** **debugging ip bgp event-infois** **debugging ip bgp event-infomsg** **debugging ip bgp event-infonexthop** **debugging ip bgp event-infoscan** **debugging ip bgp event-infothread**	Debugs BGP even handling.
debugging ip bgp filters	Debugs BGP route filtering.
debugging ip bgp fsm	Debugs BGP state machine.
debugging ip bgp keepalives	Debugs BGP neighbor keepalive.
debugging ip bgp mpls	Debugs BGP MPLS processing.
debugging ip bgp nsm	Debugs BGP core route processing.
debugging ip bgp update-info **debugging ip bgp update-infoin** **debugging ip bgp update-infoout**	Debugs BGP UPDATE message.
debugging ip bgp evpn	Debugs BGP EVPN.
debugging ip bgp vxlan	Debugs BGP VXLAN interaction.
debugging ip bgp update-group-info	Debugs BGP UPDATE-GROUP debugging switch.
debugging ip bgp alternative-paths	Debugs BGP ADD-PATH.
debugging ip bgp mbgp	Debugs MBGP routes sent by BGP to NSM.
debugging ip bgp CLI	Debugs BGP CLI commands.

debugging ip bgp debug address-family-mode link-state link-state **debugging ip bgp debug address-family-mode l2vpn evpn** **debugging ip bgp debug address-family-mode vpnv4 unicast-af-modifier** **debugging ip bgp debug address-family-mode vpnv6 unicast-af-modifier** **debugging ip bgp-instance info-filter address-family-mode ipv6 unicast-af-modifier** **debugging ip bgpdebug address-family-mode ipv4 unicast-af-modifier** **debugging ip bgp debug as** *as-number* **debugging ip bgp debug update-group-info** *update-group-index* **debugging ip bgp debug route-map** *map-tag* **debugging ip bgp debug srp-prefix** *srp-prefix* **debugging ip bgp idebug ls-prefix** *ls-prefix* **debugging ip bgp debug prefix** *prefix-address* **debugging ip bgp debug peer** [*neighbor-ipv4-address* | *neighbor-ipv6-address*] **debugging ip bgp info-filter ifx**	Debugs information filtering of BGP.
debugging ip bgp segment-routing	Debugs BGP SR.
debugging ip bgp sr-policy	Debugs BGP SR policies.
debugging ip bgp monitor	Debugs the BGP monitor.
debugging ip bgp ipc-msg	Debugs BGP IPC information.
debugging ip bgp pic	Debugs BGP PIC information.
debugging ip bgp bmp-fsm	Debugs BGP BMP state machine.
debugging ip bgp time-stat	Debugs the BGP timer.
debugging ip bgp dump-file	Debugs BGP and saves information to a file.
debugging ip bgp warn	Debugs BGP alarms.
debugging ip bgp fsm-trace	Debugs state machine stack information of BGP.

debugging ip bgp link-state-af **debugging ip bgp link-state-af srv6-sid-info** **debugging ip bgp link-state-af prefix-info** **debugging ip bgp link-state-af link-info** **debugging ip bgp link-state-af node-info**	Debugs link-state address family information of BGP.
debugging ip bgp rpi	Debugs BGP RPI.
debugging ip bgp show	Debugs BGP show information.
debugging ip bgp sock-cb warn **debugging ip bgp sock-cb event**	Debugs BGP SOCK-CB.

1.13 Configuration Examples

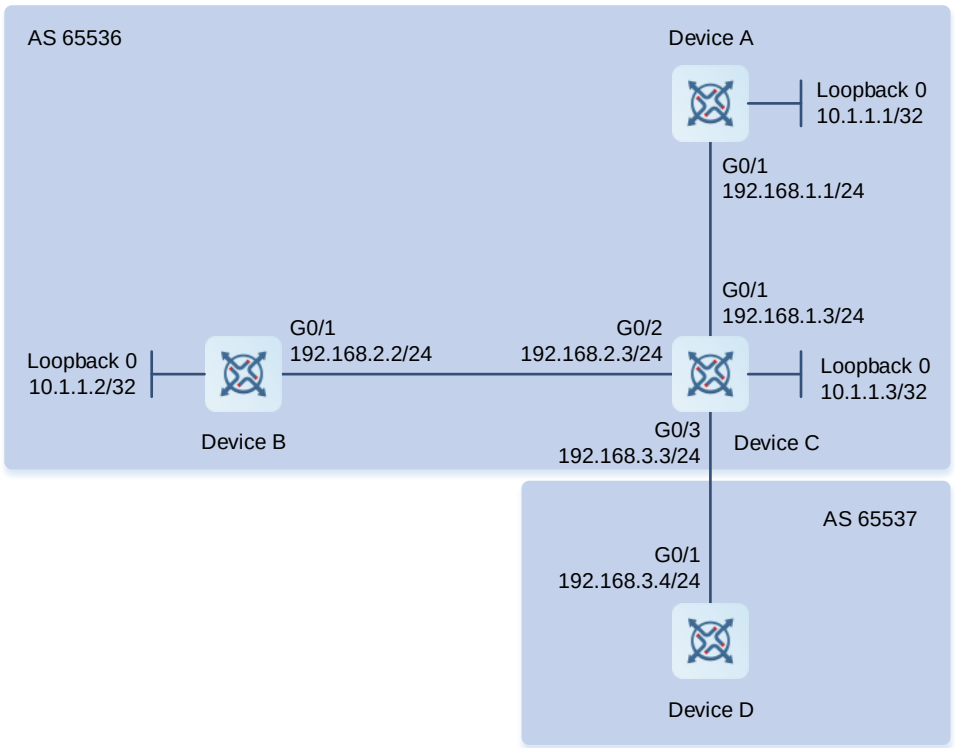
1.13.1 Configuring BGP Basic Networking

1. Requirements

BGP speakers need to establish IBGP neighbor relationships and EBGP neighbor relationships to implement basic networking.

2. Topology

Figure 1-2 Configuring BGP Basic Networking



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-2 (omitted).
- Configure a loopback interface on each of Devices A, B, and C, which establish IBGP neighbor relationships through the loopback interfaces.
- Configure Devices C and D to establish an EBGP neighbor relationship through directly connected interfaces.
- Create an IBGP peer group on Device C.

4. Procedure

(1) Configure loopback interfaces and IP addresses for the interfaces.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# interface loopback 0
Device A(cfg-if-Loopback 0)# ip address 10.1.1.1 255.255.255.255
Device A(cfg-if-Loopback 0)# exit
Device A(cfg)# interface GigabitEthernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/1)# exit
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# interface loopback 0
Device B(cfg-if-Loopback 0)# ip address 10.1.1.2 255.255.255.255
Device B(cfg-if-Loopback 0)# exit
Device B(cfg)# interface GigabitEthernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# exit
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# interface loopback 0
Device C(cfg-if-Loopback 0)# ip address 10.1.1.3 255.255.255.255
Device C(cfg-if-Loopback 0)# exit
Device C(cfg)# interface GigabitEthernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface GigabitEthernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# ip address 192.168.2.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/2)# exit
Device C(cfg)# interface GigabitEthernet 0/3
Device C(cfg-if-GigabitEthernet 0/3)# ip address 192.168.3.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/3)# exit
```

Configure Device D.

```
Device D# config-system terminal
Device D(cfg)# interface GigabitEthernet 0/1
```

```
Device D(cfg-if-GigabitEthernet 0/1)# ip address 192.168.3.4 255.255.255.0
Device D(cfg-if-GigabitEthernet 0/1)# exit
```

(2) Enable BGP and configure the loopback interfaces as neighbor interfaces.

Configure Device A.

```
Device A(cfg)# router-protocol bgp 65536
Device A(cfg-router)# peer 10.1.1.3 remote-as 65536
Device A(cfg-router)# peer 10.1.1.3 update-source loopback 0
```

Configure Device B.

```
Device B(cfg)# router-protocol bgp 65536
Device B(cfg-router)# peer 10.1.1.3 remote-as 65536
Device B(cfg-router)# peer 10.1.1.3 update-source loopback 0
```

Configure Device C.

```
Device C(cfg)# router-protocol bgp 65536
Device C(cfg-router)# peer ibgp-group peer-group
Device C(cfg-router)# peer ibgp-group remote-as 65536
Device C(cfg-router)# peer ibgp-group update-source loopback 0
Device C(cfg-router)# peer 10.1.1.1 peer-group ibgp-group
Device C(cfg-router)# peer 10.1.1.2 peer-group ibgp-group
Device C(cfg-router)# peer 192.168.3.4 remote-as 65537
```

Configure Device D.

```
Device D(cfg)# router-protocol bgp 65537
Device D(cfg-router)# peer 192.168.3.3 remote-as 65536
```

5. Verification

On Device A, run the **display ip bgp-instance peer** command to display the BGP neighbor status. By viewing the BGP neighbor status displayed in the BGP state in the corresponding neighbor table, you can verify whether the device has successfully established a BGP connection with the neighbor. The value Established indicates that the connection has been successfully established, and the value Active indicates that the neighbor has been activated but a connection has not yet been established.

```
Device A# display ip bgp-instance peer
BGP peer is 10.1.1.3, remote AS 65536, local AS 65536, internal link
  BGP version 4, remote router ID 10.1.1.3
  BGP state = Established, up for 00:00:05
  Last read          00:00:04, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 2 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:1
    refresh message:0 dynamic cap:0 notifications:0
  Sent 2 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:1
    refresh message:0 dynamic cap:0 notifications:0
```

```
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 0 seconds
  Update source is Loopback 0

 For address family: IPv4 Unicast
  BGP table version 1, peer version 1
  Index 0, Offset 0, Mask 0x1
  0 accepted prefixes
  0 announced prefixes

 Connections established 1; dropped 0
Local host: 10.1.1.1, Local port: 1039
Foreign host: 10.1.1.3, Foreign port: 179
Nexthop: 10.1.1.1
Nexthop global: ::
Nexthop local: ::
BGP connection: non shared network
Last Reset:             00:00:09, due to BGP Notification received
Notification Error Message: (Cease/Other Configuration Change.)
```

On Device B, run the **display** command to display the BGP neighbor status.

```
Device B# display ip bgp-instance peer
BGP peer is 10.1.1.3, remote AS 65536, local AS 65536, internal link
  BGP version 4, remote router ID 10.1.1.3
  BGP state = Established, up for 00:00:07
  Last read             00:00:05, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 2 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:1
    refresh message:0 dynamic cap:0 notifications:0
  Sent 2 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:1
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 0 seconds
  Update source is Loopback 0

 For address family: IPv4 Unicast
  BGP table version 1, peer version 1
  Index 0, Offset 0, Mask 0x1
  0 accepted prefixes
  0 announced prefixes

 Connections established 1; dropped 0
```

```
Local host: 10.1.1.2, Local port: 1041
Foreign host: 10.1.1.3, Foreign port: 179
Nexthop: 10.1.1.2
Nexthop global: ::
Nexthop local: ::
BGP connection: non shared network
Last Reset:             00:00:08, due to BGP Notification received
Notification Error Message: (Cease/Other Configuration Change.)
```

On device C, run the **display** command to display the BGP neighbor status.

```
Device C# display ip bgp-instance peer
BGP peer is 10.1.1.1, remote AS 65536, local AS 65536, internal link
 Member of peer-group ibgp-group for session parameters
  BGP version 4, remote router ID 10.1.1.1
  BGP state = Established, up for 00:01:13
  Last read             00:01:11, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 3 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:2
    refresh message:0 dynamic cap:0 notifications:0
  Sent 3 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:2
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 0 seconds
  Update source is Loopback 0

 For address family: IPv4 Unicast
  BGP table version 1, neighbor version 1
  Index 1, Offset 0, Mask 0x2
  ibgp-group peer-group member
  0 accepted prefixes
  0 announced prefixes

 Connections established 1; dropped 0
Local host: 10.1.1.3, Local port: 179
Foreign host: 10.1.1.1, Foreign port: 1039
Nexthop: 10.1.1.3
Nexthop global: ::
Nexthop local: ::
BGP connection: non shared network

BGP peer is 10.1.1.2, remote AS 65536, local AS 65536, internal link
 Member of peer-group ibgp-group for session parameters
```

```
  BGP version 4, remote router ID 10.1.1.2
  BGP state = Established, up for 00:01:17
  Last read          00:01:08, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 3 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:2
    refresh message:0 dynamic cap:0 notifications:0
  Sent 3 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:2
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 0 seconds
  Update source is Loopback 0

 For address family: IPv4 Unicast
  BGP table version 1, peer version 1
  Index 1, Offset 0, Mask 0x2
  ibgp-group peer-group member
  0 accepted prefixes
  0 announced prefixes

 Connections established 1; dropped 0
Local host: 10.1.1.3, Local port: 179
Foreign host: 10.1.1.2, Foreign port: 1041
Nexthop: 10.1.1.3
Nexthop global: ::
Nexthop local: ::
BGP connection: non shared network

BGP peer is 192.168.3.4, remote AS 65536, local AS 65536, internal link
 Member of peer-group ibgp-group for session parameters
  BGP version 4, remote router ID 192.168.3.4
  BGP state = Established, up for 00:01:01
  Last read          00:01:00, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 3 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:2
    refresh message:0 dynamic cap:0 notifications:0
  Sent 3 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:2
```

```
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 0 seconds
  Update source is Loopback 0

 For address family: IPv4 Unicast
  BGP table version 1, peer version 1
  Index 1, Offset 0, Mask 0x2
  ibgp-group peer-group member
  0 accepted prefixes
  0 announced prefixes

 Connections established 1; dropped 0
Local host: 192.168.3.3, Local port: 179
Foreign host: 192.168.3.4, Foreign port: 1018
Nexthop: 192.168.3.3
Nexthop global: ::
Nexthop local: ::
BGP connection: non shared network
```

On Device D, run the **display** command to display the BGP neighbor status.

```
Device D# display ip bgp-instance peer
BGP peer is 192.168.3.3, remote AS 65536, local AS 65536, internal link
 Member of peer-group ibgp-group for session parameters
  BGP version 4, remote router ID 10.1.1.3
  BGP state = Established, up for 00:01:01
  Last read           00:00:58, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 3 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:2
    refresh message:0 dynamic cap:0 notifications:0
  Sent 3 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:2
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 0 seconds
  Update source is Loopback 0

 For address family: IPv4 Unicast
  BGP table version 1, peer version 1
  Index 1, Offset 0, Mask 0x2
  ibgp-group peer-group member
  0 accepted prefixes
  0 announced prefixes
```

```
 Connections established 1; dropped 0
Local host: 192.168.3.4, Local port: 1018
Foreign host: 192.168.3.3, Foreign port: 179
Nexthop: 192.168.3.4
Nexthop global: ::
Nexthop local: ::
BGP connection: non shared network
```

6. Configuration Files

- Device A configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.1.1 255.255.255.0
!
interface Loopback 0
 ip address 10.1.1.1 255.255.255.255
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.3 remote-as 65536
 peer 10.1.1.3 update-source Loopback 0
 address-family-mode ipv4
  peer 10.1.1.3 activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.2.2 255.255.255.0
!
interface Loopback 0
 ip address 10.1.1.2 255.255.255.255
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.3 remote-as 65536
 peer 10.1.1.3 update-source Loopback 0
```

```
 address-family-mode ipv4
  peer 10.1.1.3 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.1.3 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 ip address 192.168.2.3 255.255.255.0
!
interface GigabitEthernet 0/3
 unset switch-port
 ip address 192.168.3.3 255.255.255.0
!
interface Loopback 0
 ip address 10.1.1.3 255.255.255.255
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer ibgp-group peer-group
 peer ibgp-group remote-as 65536
 peer ibgp-group update-source Loopback 0
 peer 10.1.1.1 peer-group ibgp-group
 peer 10.1.1.2 peer-group ibgp-group
 peer 192.168.3.4 remote-as 65537
 address-family-mode ipv4
  unset peer ibgp-group activate
  peer 192.168.3.4 activate
  exit-address-family-mode
!
```

- Device D configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.1.3 255.255.255.0
!
router-protocol bgp 65537
 bgp-instance log-peer-changes
```

```
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  exit-address-family-mode
!
```

7. Common Errors

- IGP is disabled and the local loopback address and the loopback address of an IBGP neighbor cannot interwork with each other, resulting in the failure to establish a neighbor relationship.
- When EBGP neighbors are not directly connected, **ebgp-multihop** is not configured, causing the failure to establish a TCP connection.

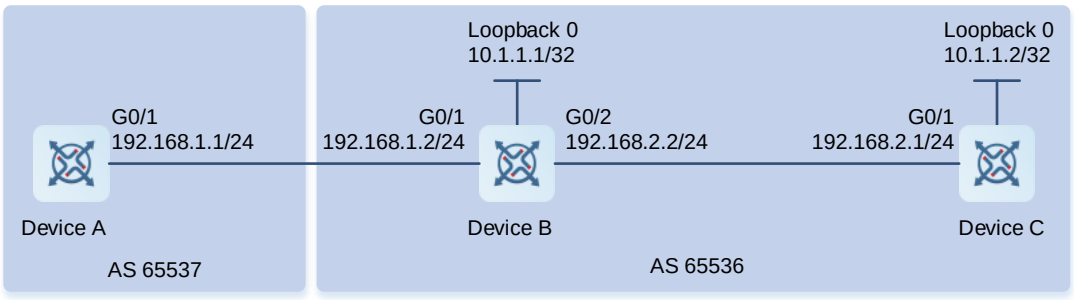
1.13.2 Configuring BGP MD5 Authentication

1. Requirements

The establishment of a TCP connection needs MD5 encryption to ensure data transmission security.

2. Topology

Figure 1-3 Configuring BGP MD5 Authentication



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-3 (omitted).
- Configure a loopback interface on each of Devices B and C, which establish an IBGP neighbor relationship through the loopback interfaces (omitted).
- Configure Devices A and B to establish an EBGP neighbor relationship through directly connected interfaces.
- On Devices A, B, and C, configure passwords for their neighbors.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65537
Device A(cfg-router)# peer 192.168.1.2 remote-as 65536
Device A(cfg-router)# peer 192.168.1.2 password 7 ebgpneighbor
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 65536
Device B(cfg-router)# peer 10.1.1.2 remote-as 65536
Device B(cfg-router)# peer 10.1.1.2 update-source loopback 0
Device B(cfg-router)# peer 10.1.1.2 password ibgpneighbor
Device B(cfg-router)# peer 192.168.1.1 remote-as 65537
Device B(cfg-router)# peer 192.168.1.1 password 7 ebgpneighbor
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# router-protocol bgp 65536
Device C(cfg-router)# peer 10.1.1.1 remote-as 65536
Device C(cfg-router)# peer 10.1.1.1 update-source loopback 0
Device C(cfg-router)# peer 10.1.1.1 password ibgpneighbor
```

5. Verification

On Device A, run the **display ip bgp-instance peers** command to display the BGP neighbor status.

```
Device A# display ip bgp-instance peers
BGP peer is 192.168.1.2, remote AS 65536, local AS 65537, external link
  BGP version 4, remote router ID 10.1.1.1
  BGP state = Established, up for 00:04:54
  Last read               , hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 7 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:6
    refresh message:0 dynamic cap:0 notifications:0
  Sent 7 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:6
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 30 seconds

 For address family: IPv4 Unicast
  BGP table version 1, peer version 0
  Index 1, Offset 0, Mask 0x2
  0 accepted prefixes
  0 announced prefixes

 Connections established 2; dropped 1
Local host: 192.168.1.1, Local port: 1026
Foreign host: 192.168.1.2, Foreign port: 179
Nexthop: 192.168.1.1
Nexthop global: ::
Nexthop local: ::
```

```
BGP connection: non shared network
Last Reset: 00:04:54, due to BGP Notification sent
Notification Error Message: (Cease/Administratively Reset.)
```

On Device B, run the **display ip bgp-instance peers** command to display the BGP neighbor status.

```
Device B# display ip bgp-instance peers
BGP peer is 10.1.1.2, remote AS 65536, local AS 65536, internal link
  BGP version 4, remote router ID 10.1.1.2
  BGP state = Established, up for 00:04:01
  Last read               , hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 8 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:7
    refresh message:0 dynamic cap:0 notifications:0
  Sent 8 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:7
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 30 seconds

 For address family: IPv4 Unicast
  BGP table version 1, peer version 0
  Index 1, Offset 0, Mask 0x2
  0 accepted prefixes
  0 announced prefixes

 Connections established 2; dropped 1
Local host: 10.1.1.1, Local port: 179
Foreign host: 10.1.1.2, Foreign port: 1038
Nexthop: 10.1.1.1
Nexthop global: ::
Nexthop local: ::
BGP connection: non shared network
Last Reset: 00:05:27, due to BGP Notification received
Notification Error Message: (Cease/Administratively Reset.)

BGP peer is 192.168.1.1, remote AS 65537, local AS 65536, external link
  BGP version 4, remote router ID 192.168.1.1
  BGP state = Established, up for 00:05:27
  Last read               , hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
```

```
  Received 8 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:7
    refresh message:0 dynamic cap:0 notifications:0
  Sent 8 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:7
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 30 seconds

 For address family: IPv4 Unicast
  BGP table version 1, peer version 0
  Index 1, Offset 0, Mask 0x2
  0 accepted prefixes
  0 announced prefixes

 Connections established 2; dropped 1
Local host: 192.168.1.2, Local port: 179
Foreign host: 192.168.1.1, Foreign port: 1026
Nexthop: 192.168.1.2
Nexthop global: ::
Nexthop local: ::
BGP connection: non shared network
Last Reset: 00:05:27, due to BGP Notification received
Notification Error Message: (Cease/Administratively Reset.)
```

On Device C, run the **display ip bgp-instance peers** command to display the BGP neighbor status.

```
Device C# display ip bgp-instance peers
BGP peer is 10.1.1.1, remote AS 65536, local AS 65536, internal link
  BGP version 4, remote router ID 10.1.1.1
  BGP state = Established, up for 00:04:01
  Last read               , hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
  Received 8 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:7
    refresh message:0 dynamic cap:0 notifications:0
  Sent 8 messages, 0 notifications, 0 in queue
    open message:1 update message:0 keepalive message:7
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 30 seconds

 For address family: IPv4 Unicast
  BGP table version 1, peer version 0
  Index 1, Offset 0, Mask 0x2
```

```
  0 accepted prefixes
  0 announced prefixes

 Connections established 2; dropped 1
Local host: 10.1.1.2, Local port: 1038
Foreign host: 10.1.1.1, Foreign port: 179
Nexthop: 10.1.1.2
Nexthop global: ::
Nexthop local: ::
BGP connection: non shared network
Last Reset: 00:05:27, due to BGP Notification received
Notification Error Message: (Cease/Administratively Reset.)
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65537
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.2 remote-as 65536
 peer 192.168.1.2 password 7 ebgpneighbor
 address-family-mode ipv4
  peer 192.168.1.2 activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.2 remote-as 65536
 peer 10.1.1.2 update-source Loopback 0
 peer 10.1.1.2 password ibgpneighbor
 peer 192.168.1.1 remote-as 65537
 peer 192.168.1.1 password 7 ebgpneighbor
 address-family-mode ipv4
  peer 10.1.1.2 activate
  peer 192.168.1.1 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.1 remote-as 65536
 peer 10.1.1.1 update-source Loopback 0
 peer 10.1.1.1 password ibgpneighbor
 address-family-mode ipv4
  peer 10.1.1.1 activate
  exit-address-family-mode
!
```

7. Common Errors

- The passwords for MD5 authentication are different on both sides of a BGP neighbor relationship.

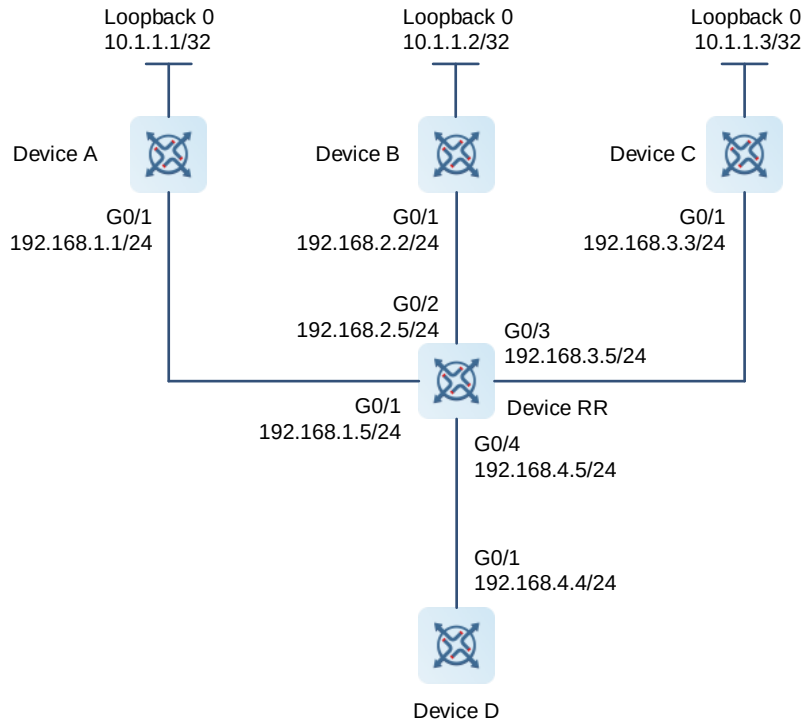
1.13.3 Configuring a BGP Route Reflector

1. Requirements

In a network involving many BGP speakers, a route reflector needs to be configured to simplify the management and synchronization of routing information and make routing more efficient.

2. Topology

Figure 1-4 Configuring a BGP Route Reflector



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-4 (omitted).
- Configure a loopback interface on each device and configure the devices to establish IBGP neighbor relationships through the loopback interfaces according to connection lines shown in the figure above (omitted).
- Configure route reflection on device RR and specify Devices A, B, C, and D as reflection clients.

4. Procedure

(1) Configure BGP neighbors to advertise network information.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65536
Device A(cfg-router)# peer 10.1.1.5 remote-as 65536
Device A(cfg-router)# peer 10.1.1.5 update-source loopback 0
Device A(cfg-router)# network-advertise 192.168.1.0 mask 255.255.255.0
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 65536
Device B(cfg-router)# peer 10.1.1.5 remote-as 65536
Device B(cfg-router)# peer 10.1.1.5 update-source loopback 0
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# router-protocol bgp 65536
Device C(cfg-router)# peer 10.1.1.5 remote-as 65536
Device C(cfg-router)# peer 10.1.1.5 update-source loopback 0
```

Configure Device D.

```
Device D# config-system terminal
Device D(cfg)# router-protocol bgp 65536
Device D(cfg-router)# peer 10.1.1.5 remote-as 65536
Device D(cfg-router)# peer 10.1.1.5 update-source loopback 0
```

(2) Configure a route reflector.

Configure Device RR.

```
Device RR# config-system terminal
Device RR(cfg)# router-protocol bgp 65536
Device RR(cfg-router)# peer 10.1.1.1 remote-as 65536
Device RR(cfg-router)# peer 10.1.1.1 update-source loopback 0
Device RR(cfg-router)# peer 10.1.1.1 route-reflector-client
Device RR(cfg-router)# peer 10.1.1.2 remote-as 65536
Device RR(cfg-router)# peer 10.1.1.2 update-source loopback 0
Device RR(cfg-router)# peer 10.1.1.2 route-reflector-client
Device RR(cfg-router)# peer 10.1.1.3 remote-as 65536
Device RR(cfg-router)# peer 10.1.1.3 update-source loopback 0
Device RR(cfg-router)# peer 10.1.1.3 route-reflector-client
```

```
Device RR(cfg-router)# peer 10.1.1.4 remote-as 65536
Device RR(cfg-router)# peer 10.1.1.4 update-source loopback 0
Device RR(cfg-router)# peer 10.1.1.4 route-reflector-client
```

5. Verification

On Device RR, check the BGP neighbor status.

```
Device RR# display ip bgp-instance status
BGP router identifier 10.1.1.5, local AS number 65536
BGP table version is 1
0 BGP AS-PATH entries
0 BGP Community entries
1 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down   State/PfxRcd
10.1.1.1        4        65536       8       9        1    0    0 00:05:11        1
10.1.1.2        4        65536       9       9        1    0    0 00:05:24        0
10.1.1.3        4        65536       8       7        1    0    0 00:05:10        0
10.1.1.4        4        65536       9       8        1    0    0 00:05:14        0

Device RR# display ip bgp-instance
BGP table version is 1, local router ID is 10.1.1.5
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric     LocPrf     Weight Path
*>i192.168.1.0      10.1.1.1                 0        100          0      i

Total number of prefixes 1
```

On Device D, run the **display ip bgp-instance status** command to display BGP details.

```
Device D# display ip bgp-instance status
BGP router-protocol id 10.1.1.4, local AS number 65536
BGP table version is 1
0 BGP AS-PATH entries
0 BGP Community entries
1 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down   State/PfxRcd
10.1.1.5        4        65536       8       9        1    0    0 00:05:20        1

Device D# display ip bgp-instance
BGP table version is 1, local router-protocol ID is 10.1.1.4
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete
```

```
   Network          Next Hop            Metric     LocPrf     Weight Path
* i192.168.1.0     10.1.1.1                 0        100          0      i

Total number of prefixes 1
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.5 remote-as 65536
 peer 10.1.1.5 update-source Loopback 0
 address-family-mode ipv4
  network-advertise 192.168.1.0
  peer 10.1.1.5 activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.5 remote-as 65536
 peer 10.1.1.5 update-source Loopback 0
 address-family-mode ipv4
  peer 10.1.1.5 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.5 remote-as 65536
 peer 10.1.1.5 update-source Loopback 0
 address-family-mode ipv4
  peer 10.1.1.5 activate
  exit-address-family-mode
```

```
!
```

- Device D configuration file

```
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.5 remote-as 65536
 peer 10.1.1.5 update-source Loopback 0
 address-family-mode ipv4
  peer 10.1.1.5 activate
  exit-address-family-mode
!
```

- Device RR configuration file

```
!
router-protocol bgp 65536
 bgp-instance as-confederation id 100
 bgp-instance as-confederation peers 65537
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.1 remote-as 65536
 peer 10.1.1.1 update-source Loopback 0
 peer 10.1.1.2 remote-as 65536
 peer 10.1.1.2 update-source Loopback 0
 peer 10.1.1.3 remote-as 65536
 peer 10.1.1.3 update-source Loopback 0
 peer 10.1.1.4 remote-as 65536
 peer 10.1.1.4 update-source Loopback 0
 address-family-mode ipv4
  network-advertise 192.168.1.0
  peer 10.1.1.1 activate
  peer 10.1.1.1 route-reflector-client
  peer 10.1.1.2 activate
  peer 10.1.1.2 route-reflector-client
  peer 10.1.1.3 activate
  peer 10.1.1.3 route-reflector-client
  peer 10.1.1.4 activate
  peer 10.1.1.4 route-reflector-client
  exit-address-family-mode
!
```

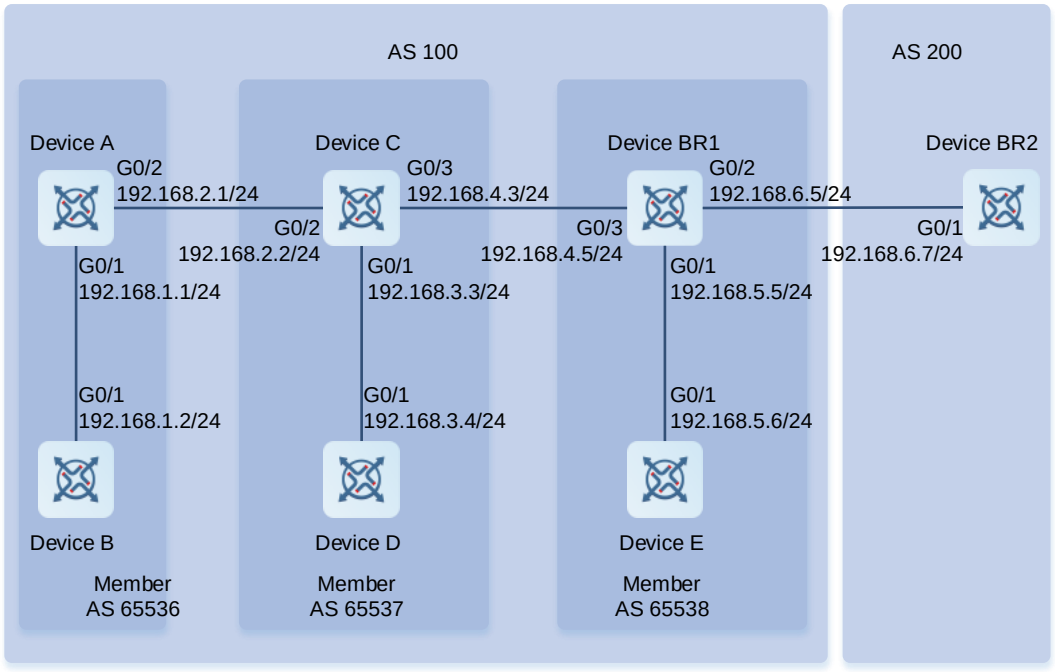
1.13.4 Configuring a BGP Alliance

1. Requirements

A BGP alliance needs to be configured to solve the problem of too many maintenance resources due to excessive IBGP connections in the network.

2. Topology

Figure 1-5 Configuring a BGP Alliance



3. Notes

- Configure IP addresses for device interfaces (omitted).
- On Devices A and B, configure BGP, set the member AS number to **65536**, and configure an IBGP neighbor relationship.
- On Devices C and D, configure BGP, set the member AS number to **65537**, and configure an IBGP neighbor relationship.
- On Devices BR1 and E, configure BGP, set the member AS number to **65538**, and configure an IBGP neighbor relationship.
- On Devices A, B, C, D, E, and BR1, set the Alliance ID to **100**.
- On Device A, configure 65537 as the alliance number, configure device C as an EBGP neighbor, and set the peer AS number to **65537**.
- On Device C, set the alliance member numbers to **65536** and **65538**, configure device A as an EBGP neighbor and set the peer AS number to **65536**, configure device BR1 as an EBGP neighbor and set the peer AS number to **65538**.
- On Device BR1, set the alliance member number to **65537**, configure device C as an EBGP neighbor and set the peer AS number to **65537**, configure device BR2 as an EBGP neighbor and set the peer AS number

to **200**.

- On Device BR2, configure BGP, set the AS number to **200**, configure device BR1 an EBGP neighbor, and set the peer AS number to **100**.

4. Procedure

(1) Configure IBGP neighbors and EBGP neighbors to advertise network information.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65536
Device A(cfg-router)# peer 10.1.1.2 remote-as 65536
Device A(cfg-router)# peer 10.1.1.2 update-source loopback 0
Device A(cfg-router)# peer 10.1.1.3 remote-as 65537
Device A(cfg-router)# peer 10.1.1.3 ebgp-multihop 2
Device A(cfg-router)# peer 10.1.1.3 update-source loopback 0
Device A(cfg-router)# network-advertise 192.168.1.0 mask 255.255.255.0
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 65536
Device B(cfg-router)# peer 10.1.1.1 remote-as 65536
Device B(cfg-router)# peer 10.1.1.1 update-source loopback 0
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# router-protocol bgp 65537
Device C(cfg-router)# peer 10.1.1.1 remote-as 65536
Device C(cfg-router)# peer 10.1.1.1 update-source loopback 0
Device C(cfg-router)# peer 10.1.1.1 ebgp-multihop 2
Device C(cfg-router)# peer 10.1.1.4 remote-as 65537
Device C(cfg-router)# peer 10.1.1.4 update-source loopback 0
Device C(cfg-router)# peer 10.1.1.5 remote-as 65538
Device C(cfg-router)# peer 10.1.1.5 update-source loopback 0
Device C(cfg-router)# peer 10.1.1.5 ebgp-multihop 2
```

Configure Device D.

```
Device D# config-system terminal
Device D(cfg)# router-protocol bgp 65537
Device D(cfg-router)# peer 10.1.1.3 remote-as 65537
Device D(cfg-router)# peer 10.1.1.3 update-source loopback 0
```

Configure Device E.

```
Device E# config-system terminal
Device E(cfg)# router-protocol bgp 65538
Device E(cfg-router)# peer 10.1.1.5 remote-as 65538
Device E(cfg-router)# peer 10.1.1.5 update-source loopback 0
```

Configure Device BR1.

```
Device BR1# config-system terminal
```

```
Device BR1(cfg)# router-protocol bgp 65538
Device BR1(cfg-router)# peer 10.1.1.3 remote-as 65537
Device BR1(cfg-router)# peer 10.1.1.3 update-source loopback 0
Device BR1(cfg-router)# peer 10.1.1.3 ebgp-multihop 2
Device BR1(cfg-router)# peer 10.1.1.6 remote-as 65538
Device BR1(cfg-router)# peer 10.1.1.6 update-source loopback 0
Device BR1(cfg-router)# peer 192.168.6.7 remote-as 200
```

Configure Device BR2.

```
Device BR2# config-system terminal
Device BR2(cfg)# router-protocol bgp 200
Device BR2(cfg-router)# peer 192.168.6.5 remote-as 100
Device BR2(cfg-router)# network-advertise 192.168.6.0 mask 255.255.255.0
```

(2) Configuring alliance members.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65536
Device A(cfg-router)# bgp-instance as-confederation id 100
Device A(cfg-router)# bgp-instance as-confederation peers 65537
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# router-protocol bgp 65537
Device C(cfg-router)# bgp-instance as-confederation id 100
Device C(cfg-router)# bgp-instance as-confederation peers 65536 65538
```

Configure Device BR1.

```
Device BR1# config-system terminal
Device BR1(cfg)# router-protocol bgp 65538
Device BR1(cfg-router)# bgp-instance as-confederation id 100
Device BR1(cfg-router)# bgp-instance as-confederation peers 65537
```

5. Verification

On Device A, check BGP information.

```
Device A# display ip bgp-instance status
BGP router identifier 10.1.1.1, local AS number 65536
BGP table version is 1
1 BGP AS-PATH entries
0 BGP Community entries
1 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V         AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down       State/PfxRcd
10.1.1.2        4      65536       3       3        1    0    0 00:00:05            0
10.1.1.3        4      65537       3       3        1    0    0 00:00:06            1

Total number of neighbors 1
```

```
Device A# display ip bgp-instance
BGP table version is 1, local router ID is 10.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop          Metric     LocPrf     Weight Path
*  192.168.6.0      192.168.6.7            0        100             0 (65537 65538) 200 i

Total number of prefixes 1
```

On Device BR1, check BGP information.

```
Device BR1# display ip bgp-instance status
BGP router identifier 10.1.1.5, local AS number 200
BGP table version is 2
2 BGP AS-PATH entries
0 BGP Community entries
2 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down       State/PfxRcd
10.1.1.3        4        65537       3       3        2    0    0 00:00:10            1
10.1.1.6        4        65538       3       3        2    0    0 00:00:08            0
192.168.6.7     4          200       3       3        2    0    0 00:00:05            1

Total number of neighbors 1

Device BR1# display ip bgp-instance
BGP table version is 1, local router ID is 10.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop          Metric     LocPrf     Weight Path
*  192.168.1.0      10.1.1.1               0        100             0 (65537 65536) i
*> 192.168.6.0      192.168.6.7            0        100             0 200 i

Total number of prefixes 1
```

On Device BR2, check BGP information.

```
Device BR2# display ip bgp-instance status
BGP router identifier 192.168.6.7, local AS number 200
BGP table version is 1
1 BGP AS-PATH entries
0 BGP Community entries
1 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down       State/PfxRcd
```

```
192.168.6.5     4          100        3        3        1     0     0 00:00:05          1

Total number of neighbors 1

Device BR2# display ip bgp-instance
BGP table version is 1, local router ID is 10.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                    S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

    Network          Next Hop            Metric     LocPrf     Weight Path
*> 192.168.1.0      192.168.6.5              0        100          0 (65537 65538) 200 i

Total number of prefixes 1
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65536
 bgp-instance as-confederation id 100
 bgp-instance as-confederation peers 65537
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.2 remote-as 65536
 peer 10.1.1.2 update-source Loopback 0
 peer 10.1.1.3 remote-as 65537
 peer 10.1.1.3 ebgp-multihop 2
 peer 10.1.1.3 update-source Loopback 0
 address-family-mode ipv4
  network-advertise 192.168.1.0
  peer 10.1.1.2 activate
  peer 10.1.1.3 activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.1 remote-as 65536
 peer 10.1.1.1 update-source Loopback 0
 address-family-mode ipv4
```

```
  peer 10.1.1.1 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
router-protocol bgp 65537
 bgp-instance as-confederation id 100
 bgp-instance as-confederation peers 65536 65538
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.1 remote-as 65536
 peer 10.1.1.1 ebgp-multihop 2
 peer 10.1.1.1 update-source Loopback 0
 peer 10.1.1.4 remote-as 65537
 peer 10.1.1.4 update-source Loopback 0
 peer 10.1.1.5 remote-as 65538
 peer 10.1.1.5 ebgp-multihop 2
 peer 10.1.1.5 update-source Loopback 0
 address-family-mode ipv4
  peer 10.1.1.1 activate
  peer 10.1.1.4 activate
  peer 10.1.1.5 activate
  exit-address-family-mode
!
```

- Device D configuration file

```
!
router-protocol bgp 65537
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.3 remote-as 65537
 peer 10.1.1.3 update-source Loopback 0
 address-family-mode ipv4
  peer 10.1.1.3 activate
  exit-address-family-mode
!
```

- Device E configuration file

```
!
router-protocol bgp 65538
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
```

```
 bgp-instance graceful-restart
 peer 10.1.1.5 remote-as 65538
 peer 10.1.1.5 update-source Loopback 0
 address-family-mode ipv4
  peer 10.1.1.5 activate
  exit-address-family-mode
!
```

- Device BR1 configuration file

```
!
router-protocol bgp 65538
 bgp-instance as-confederation id 100
 bgp-instance as-confederation peers 65537
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.3 remote-as 65537
 peer 10.1.1.3 ebgp-multihop 2
 peer 10.1.1.3 update-source Loopback 0
 peer 10.1.1.6 remote-as 65538
 peer 10.1.1.6 update-source Loopback 0
 peer 192.168.6.7 remote-as 200
 address-family-mode ipv4
  peer 10.1.1.3 activate
  peer 10.1.1.6 activate
  peer 192.168.6.7 activate
  exit-address-family-mode
!
```

- Device BR2 configuration file

```
!
router-protocol bgp 200
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.6.5 remote-as 100
 address-family-mode ipv4
  network-advertise 192.168.6.0
  peer 192.168.6.5 activate
  exit-address-family-mode
!
```

7. Common Errors

- No BGP alliance neighbor is configured.
- A full mesh of connections is not deployed within member ASs of an alliance.

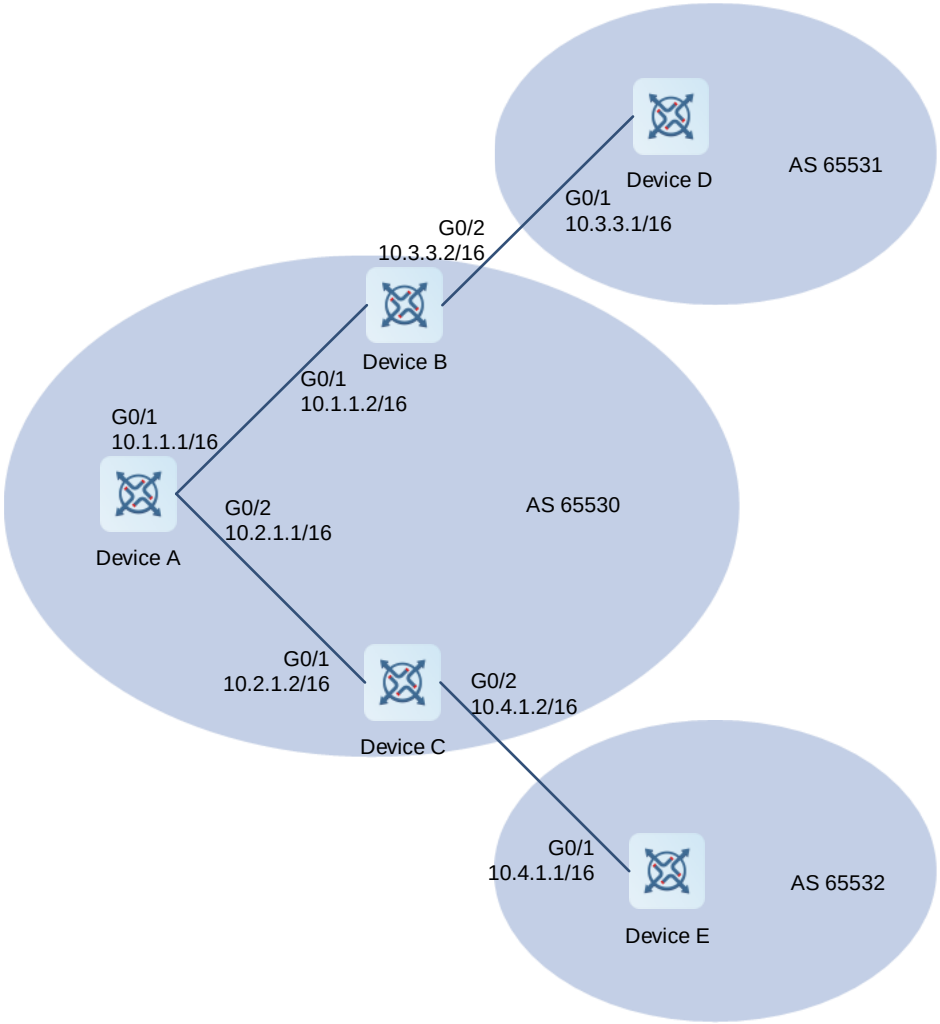
1.13.5 Configuring IBGP Multi-Path Load Balancing

1. Requirements

When there are two routes to the same destination network segment, load balancing needs to be configured to fully utilize network bandwidth resources.

2. Topology

Figure 1-6 Configuring IBGP Multi-Path Load Balancing



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-6 (omitted).
- Configure devices A and B and devices A and C to establish IBGP neighbor relationships through directly connected interfaces (omitted).
- Configure devices B and D and devices C and E to establish EBGP neighbor relationships through directly connected interfaces (omitted).
- Import the same routes to devices D and E.
- On device A, configure IBGP load balancing and enable the AS-PATH loose comparison mode.

4. Procedure

(1) Configure routing information on device A, and re-advertise directly connected route information on devices D and E.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# ip route-config 10.3.0.0 255.255.0.0 10.1.1.2
Device A(cfg)# ip route-config 10.4.0.0 255.255.0.0 10.2.1.2
```

Configure Device D.

```
Device D# config-system terminal
Device D(cfg)# router-protocol bgp 65531
Device D(cfg-router)# redistribute-route connected
```

Configure Device E.

```
Device E# config-system terminal
Device E(cfg)# router-protocol bgp 65532
Device E(cfg-router)# redistribute-route connected
```

(2) On Device A, configure load balancing in AS-PATH loose comparison mode.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65530
Device A(cfg-router)# max-ecmp-paths ibgp 2
Device A(cfg-router)# bgp-instance optimal-path as-path multipath-relax
```

5. Verification

On Device A, check BGP information.

```
Device A# display ip bgp-instance status
BGP router identifier 10.2.1.1, local AS number 65530
BGP table version is 9
2 BGP AS-PATH entries
0 BGP Community entries
3 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V    AS         MsgRcvd    MsgSent    TblVer    InQ    OutQ    Up/Down
State/PfxRcd
172.16.23.140   4    65530      29         25         8         0      0       00:18:48    2
172.16.23.141   4    65530      24         21         8         0      0       00:17:58    2

Device A# display ip bgp-instance
BGP table version is 9, local router ID is 10.2.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric     LocPrf     Weight Path
*>i10.3.0.0/16      10.3.1.1                 0        100          0 65531 ?
```

```
*>i10.4.0.0/16          10.4.1.1                 0          100          0 65532 ?
* i10.5.0.0/16          10.3.1.1                 0          100          0 65531 ?
*>i                      10.4.1.1                 0          100          0 65532 ?

Total number of prefixes 3
Device A# display ip bgp-instance 10.5.0.0
BGP routing table entry for 10.5.0.0/16
Paths: (2 available, best #1, table Default-IP-Routing-Table)
  Not advertised to any peer
  65532
    10.4.1.1 from 10.2.1.2 (172.16.24.1)
      Origin incomplete, metric 0, localpref 100, valid, internal, multipath, best
      Last update: Mon Mar 21 03:45:14 2011

  65531
    10.3.1.1 from 10.1.1.2 (172.16.25.1)
      Origin incomplete, metric 0, localpref 100, valid, internal, multipath
      Last update: Mon Mar 21 03:45:14 2011

Device A# display ip route-config

Codes:   C - connected, S - static, R - RIP, B - BGP
         O - OSPF, IA - OSPF inter area
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF   type 2
         i - IS-IS, su - IS-IS status, L1 - IS-IS level-1, L2 - IS-IS level-2
         ia - IS-IS inter area, * - candidate default

Gateway of last resort is unset set
C       10.1.0.0/16 is directly connected, FastEthernet 0/0
C       10.1.1.1/32 is local host-route.
C       10.2.0.0/16 is directly connected, FastEthernet 0/1
C       10.2.1.1/32 is local host-route.
S       10.3.0.0/16 [1/0] via 10.1.1.2
S       10.4.0.0/16 [1/0] via 10.2.1.2
B       10.5.0.0/16 [200/0] via 10.3.1.1, 00:27:56
                    [200/0] via 10.4.1.1, 00:27:56
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65530
 bgp-instance optimal-path as-path multipath-relax
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
```

```
 bgp-instance graceful-restart
 address-family-mode ipv4
  max-ecmp-paths ibgp 2
  exit-address-family-mode
!
ip route-config 10.3.0.0 255.255.0.0 10.1.1.2
ip route-config 10.4.0.0 255.255.0.0 10.2.1.2
!
```

- Device D configuration file

```
!
router-protocol bgp 65531
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  redistribute-route connected
  exit-address-family-mode
!
```

- Device E configuration file

```
!
router-protocol bgp 65532
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  redistribute-route connected
  exit-address-family-mode
!
```

7. Common Errors

The priorities of multi-hop BGP paths are different, which causes a load balancing failure.

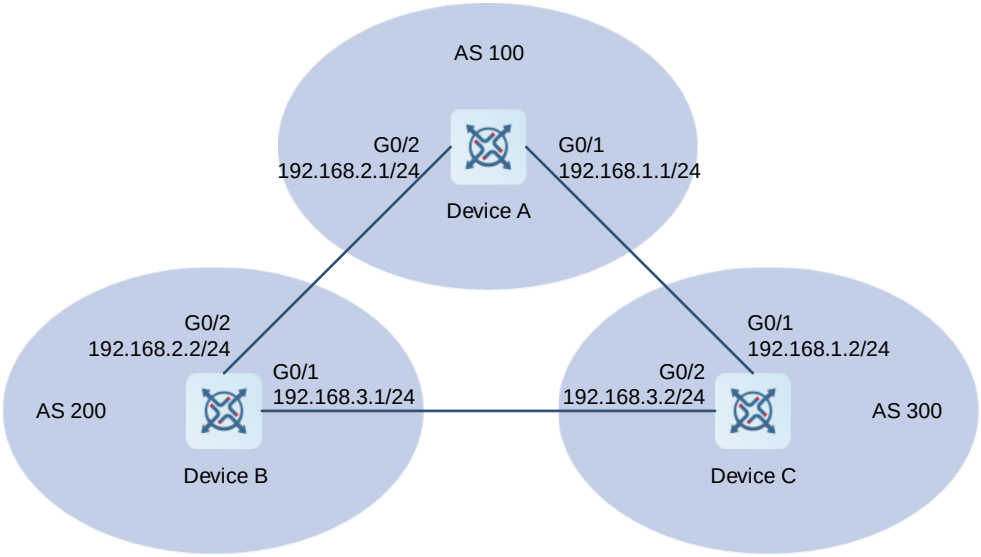
1.13.6 Configuring EBGP FRR

1. Requirements

The FRR function needs to be configured to back up and forward routing information and reduce the delay in service switching.

2. Topology

Figure 1-7 Configuring EBGP FRR



3. Notes

- Enable BGP on all devices (omitted).
- Configure addresses for directly connected interfaces on Devices A, B, and C, and configure them to establish EBGP neighbor relationships (omitted).
- Configure a BFD session for the EBGP neighbor relationship between Devices B and C.
- Configure FRR on Device C.

4. Procedure

(1) Configure interface addresses and interface BFD parameters on devices B and C.

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# interface GigabitEthernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.3.1 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# bfd min_tx_interval 200 min_rx_interval 200 multiplier 5
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface GigabitEthernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# ip address 192.168.2.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/2)# exit
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# interface GigabitEthernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.2 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface GigabitEthernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# ip address 192.168.3.2 255.255.255.0
```

```
Device C(cfg-if-GigabitEthernet 0/2)# bfd min_tx_interval 200 min_rx_interval 200 multiplier 5
Device C(cfg-if-GigabitEthernet 0/2)# exit
```

(2) Configure BFD for BGP and BGP FRR.

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 200
Device B(cfg-router)# peer 192.168.3.2 remote-as 300
Device B(cfg-router)# peer 192.168.3.2 fall-over bfd
Device B(cfg-router)# peer 192.168.2.1 remote-as 100
Device B(cfg-router)# redistribute-route connect
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# router-protocol bgp 300
Device C(cfg-router)# peer 192.168.1.1 remote-as 100
Device C(cfg-router)# peer 192.168.3.1 remote-as 200
Device C(cfg-router)# peer 192.168.3.1 fall-over bfd
Device C(cfg-router)# address-family-mode ipv4 unicast
Device C(cfg-router-af)# bgp-instance auto-frr
Device C(cfg-router-af)# redistribute-route connect
```

5. Verification

On Device C, check BGP information.

```
Device C# display ip bgp-instance status
BGP router identifier 10.10.10.10, local AS number 300
BGP table version is 12
4 BGP AS-PATH entries
0 BGP Community entries
3 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
192.168.1.1     4          100      76      77       12    12    0 00:59:27        3
192.168.3.1     4          200      30      30       12    12    0 00:19:03        3

Total number of neighbors 2

Device C# display ip bgp-instance
BGP table version is 12, local router ID is 10.10.10.10
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

     Network          Next Hop            Metric     LocPrf     Weight Path
*    192.168.1.0      192.168.3.1              0                    0 200 ?
*                     192.168.1.1              0                    0 100 ?
*>                    0.0.0.0                  0                32768 ?
```

```
*> 192.168.2.0          192.168.3.1                0                          0 200 ?
*b                       192.168.1.1                0                          0 100 ?
*   192.168.3.0          192.168.3.1                0                          0 200 ?
*                        192.168.1.1                0                          0 100 200 ?
*>                       0.0.0.0                    0                      32768        ?

Total number of prefixes 3
Device C# display ip bgp-instance 192.168.2.0
BGP routing table entry for 192.168.2.0/24
Paths: (2 available, best #1, table Default-IP-Routing-Table)
  Advertised to non peer-group peers:
  192.168.1.1
  200
    192.168.3.1 from 192.168.3.1 (3.3.3.3)
      Origin incomplete, metric-value 0, localpref 100, valid, external, best
      Last update: Tue Oct    5 00:26:52 1971

  100
    192.168.1.1 from 192.168.1.1 (44.44.44.44)
      Origin incomplete, metric-value 0, localpref 100, valid, external, backup
      Last update: Mon Oct    4 23:46:28 1971
Device C# display ip route-config

Codes:    C - connected, S - static, R - RIP, B - BGP
          O - OSPF, IA - OSPF inter area
          N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
          E1 - OSPF external type 1, E2 - OSPF external type 2
          i - IS-IS, su - IS-IS status, L1 - IS-IS level-1, L2 - IS-IS level-2
          ia - IS-IS inter area, * - candidate reset

Gateway of last resort is unset set
C       192.168.1.0/24 is directly connected, GigabitEthernet 1/9
C       192.168.1.2/32 is local host-route.
B       192.168.2.0/24 [20/0] via 192.168.3.1, 00:21:39
C       192.168.3.0/24 is directly connected, GigabitEthernet 1/11
C       192.168.3.2/32 is local host.
```

6. Configuration Files

- Device B configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.3.1 255.255.255.0
 bfd _interval interval 200 min_rx_interval 200 multiplier 5
!
interface GigabitEthernet 0/2
```

```
 unset switch-port
 ip address 192.168.2.2 255.255.255.0
!
router-protocol bgp 200
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.2.1 remote-as 100
 peer 192.168.3.2 remote-as 300
 peer 192.168.3.2 fall-over bfd
 address-family-mode ipv4
  redistribute-route connected
  peer 192.168.2.1 activate
  peer 192.168.3.2 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.1.2 255.255.255.0
 bfd min_tx_interval 200 min_rx_interval 200 multiplier 5
!
interface GigabitEthernet 0/2
 unset switch-port
 ip address 192.168.3.2 255.255.255.0
 bfd min_tx_interval 200 min_rx_interval 200 multiplier 5
!
router-protocol bgp 300
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.1 remote-as 100
 peer 192.168.3.1 remote-as 200
 peer 192.168.3.1 fall-over bfd
 address-family-mode ipv4
  bgp-instance auto-frr
  redistribute-route connected
  peer 192.168.1.1 activate
  peer 192.168.3.1 activate
  exit-address-family-mode
!
```

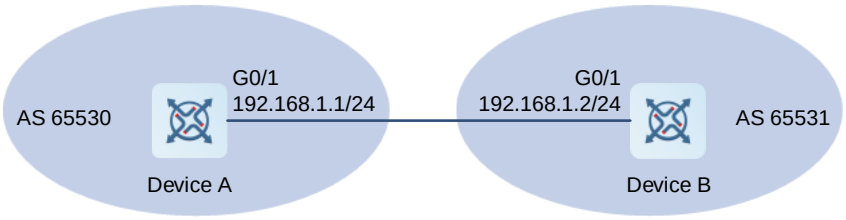
1.13.7 Configuring BGP to Rapidly Withdraw Specified Routes

1. Requirements

Specified routing information needs to be withdrawn.

2. Topology

Figure 1-8 Configuring Fast Withdrawal of Specified Routes



3. Notes

- Enable BGP on all devices (omitted).
- Configure devices A and B to establish a BGP neighbor relationship (omitted).
- Configure an ACL on device A.
- Configure fast withdrawal of specified routes on device A.

4. Procedure

(1) Configure an ACL to permit the address 1.1.1.1.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# ip access-control-list standard 1
Device A(cfg-stdacl)# permit 1.1.1.1 0.0.0.0
Device A(cfg-stdacl)# exit
```

(2) Configure fast withdrawal of the specified route 1.1.1.1.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65530
Device A(cfg-router)# peer 192.168.1.2 remote-as 65531
Device A(cfg-router)# bgp-instance fast-withdraw access-control-list 1
```

5. Verification

On Device A, check BGP routing information.

```
Device A# display current-config router-protocol bgp
router-protocol bgp 65530
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.2 remote-as 65531
```

```
 address-family-mode ipv4
   bgp-instance fast-withdraw access-control-list 1
   peer 192.168.1.2 activate
 exit-address-family-mode
```

6. Configuration Files

- Device A configuration file

```
!
ip access-control-list standard 1
  10 permit host-route 1.1.1.1
!
router-protocol bgp 65530
  bgp-instance log-peer-changes
  bgp-instance graceful-restart restart-time 120
  bgp-instance graceful-restart staleroute-time 360
  bgp-instance graceful-restart
  peer 192.168.1.2 remote-as 65531
  address-family-mode ipv4
    bgp-instance fast-withdraw access-control-list 1
    peer 192.168.1.2 activate
    exit-address-family-mode
!
```

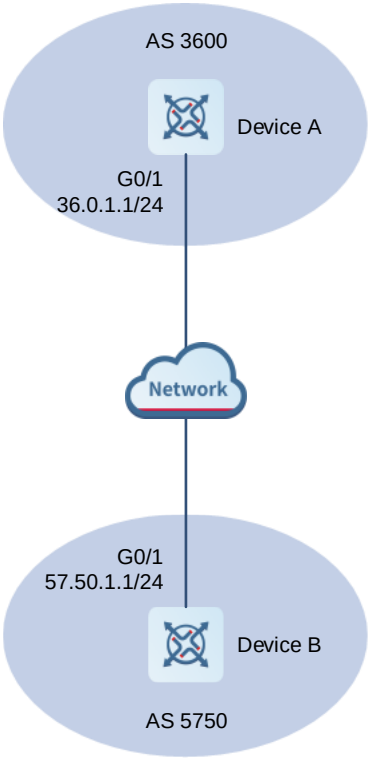
1.13.8 Configuring the BGP Local AS

1. Requirements

When the real AS of the local BGP changes, a new virtual AS 23 can be configured between peers so that the BGP configurations on peers do not need to be configured.

2. Topology

Figure 1-9 Configuring BGP Local AS



3. Notes

- Establish an EBGP neighbor relationship between devices A and B and set the local AS number of the EBGP neighbor to 23.
- Establish an EBGP neighbor relationship with devices A on B.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 3600
Device A(cfg-router)# peer 57.50.1.1 remote-as 5750
Device A(cfg-router)# peer 57.50.1.1 update-source loopback 0
Device A(cfg-router)# peer 57.50.1.1 ebgp-multihop 255
Device A(cfg-router)# peer 57.50.1.1 local-as 23 no-prepend change-as dual-as
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 5750
Device B(cfg-router)# peer 36.0.1.1 remote-as 23
Device B(cfg-router)# peer 36.0.1.1 update-source loopback 0
Device B(cfg-router)# peer 36.0.1.1 ebgp-multihop 255
```

5. Verification

On Device A, check the BGP neighbor status.

```
Device A# display ip bgp-instance peers 57.50.1.1
BGP peer is 57.50.1.1, remote AS 5750, local AS 23(using Peer's Local AS, no-prepend, change-as,
dual-as), external link
  BGP version 4, remote router ID 0.0.0.0
  BGP state = Idle
  Last read, hold time is 180, keepalive interval is 60 seconds
  Received 0 messages, 0 notifications, 0 in queue
    open message:0 update message:0 keepalive message:0
    refresh message:0 dynamic cap:0 notifications:0
  Sent 0 messages, 0 notifications, 0 in queue
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 3600
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 57.50.1.1 remote-as 5750
 peer 57.50.1.1 local-as 23 no-prepend change-as dual-as
 peer 57.50.1.1 ebgp-multihop 255
 peer 57.50.1.1 update-source Loopback 0
 address-family-mode ipv4
  peer 57.50.1.1 activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 5750
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 36.0.1.1 remote-as 23
 peer 36.0.1.1 ebgp-multihop 255
 peer 36.0.1.1 update-source Loopback 0
 address-family-mode ipv4
  peer 36.0.1.1 activate
  exit-address-family-mode
!
```

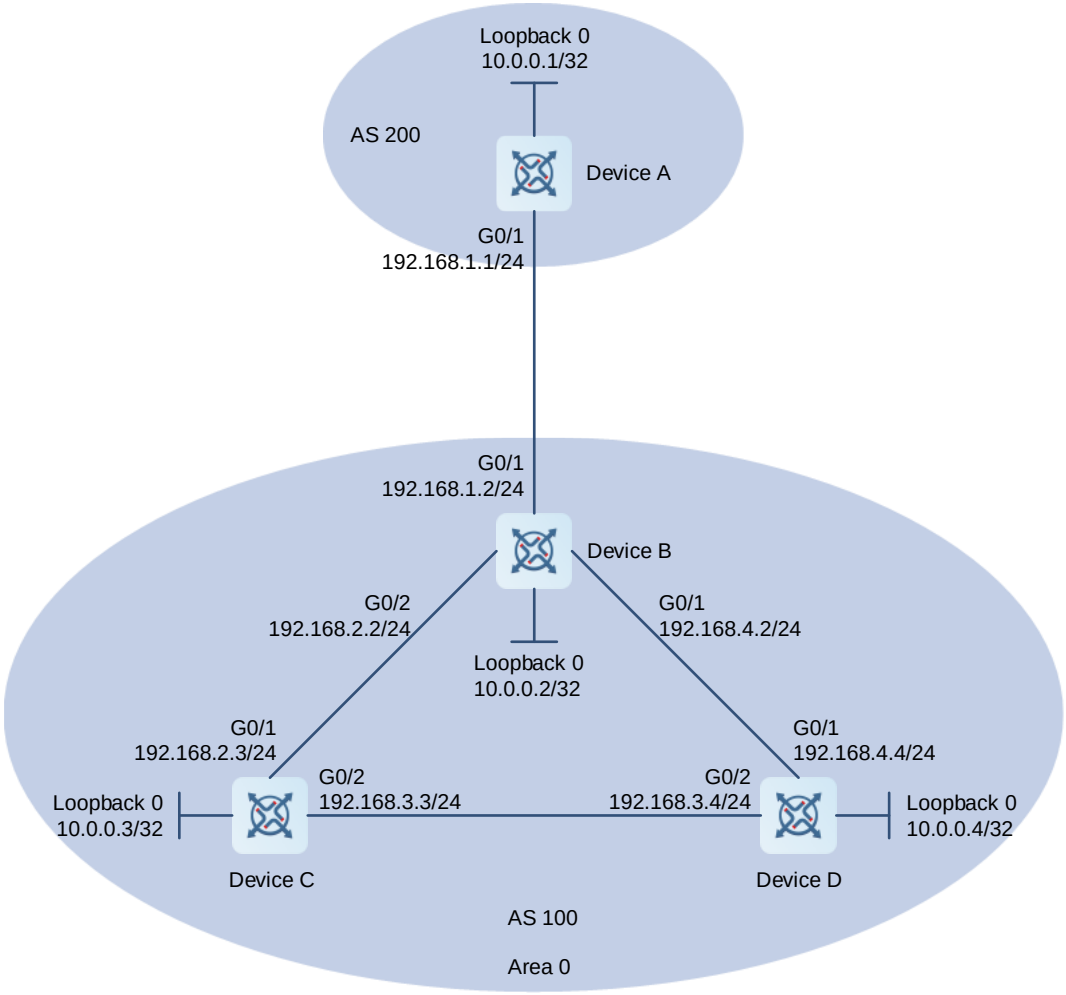
1.13.9 Configuring BGP GR

1. Requirements

Data forwarding should not be interrupted during the protocol restart.

2. Topology

Figure 1-10 Configuring BGP GR



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-10 (omitted).
- Configure a loopback interface on each of Devices B, C, and D and configure the devices to establish IBGP neighbor relationships through the loopback interfaces (omitted).
- Configure Devices A and B to establish an EBGP neighbor relationship through directly connected interfaces (omitted).
- On Devices A, B, C, and D, and enable the BGP GR function.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol ospf 1
Device A(cfg-router)# gr
Device A(cfg-router)# exit
Device A(cfg)# router-protocol bgp 200
```

```
Device A(cfg-router)# bgp-instance graceful-restart
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol ospf 1
Device B(cfg-router)# gr
Device B(cfg-router)# exit
Device B(cfg)# router-protocol bgp 100
Device B(cfg-router)# bgp-instance graceful-restart
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# router-protocol ospf 1
Device C(cfg-router)# graceful-restart
Device C(cfg-router)# exit
Device C(cfg)# router-protocol bgp 100
Device C(cfg-router)# bgp-instance graceful-restart
```

Configure Device D.

```
Device D# config-system terminal
Device D(cfg)# router-protocol ospf 1
Device D(cfg-router)# gr
Device D(cfg-router)# exit
Device D(cfg)# router-protocol bgp 100
Device D(cfg-router)# bgp-instance graceful-restart
```

5. Verification

On Device B, run the **display ip ospf-instance** command to display the OSPF routing status.

```
Device B# display ip ospf-instance
Routing Process "ospf 1" with ID 10.0.0.2
Process uptime is 4 minutes
Process bound to VRF default
Conforms to RFC2328, and RFC1583Compatibility flag isenabled
Supports only single TOS(TOS0) routes
Supports opaque LSA
This router is an ASBR (injecting external routing information)
SPF schedule delay 5 secs, Hold time between two SPFs 10 secs
LsaGroupPacing: 240 secs
Number of incomming current DD exchange neighbors 0/5
Number of outgoing current DD exchange neighbors 0/5
Number of external LSA 4. Checksum 0x0278E0
Number of opaque AS LSA 0. Checksum 0x000000
Number of non-default external LSA 4
External LSA database is unlimited.
Number of LSA originated 6
Number of LSA received 2
Log Neighbor Adjency Changes : Enabled
Graceful-restart enabled
```

```
Graceful-restart helper support enabled
Number of areas attached to this router: 1
Area 0 (BACKBONE)
```

On Device B, run the **display ip bgp-instance peers** command to display the BGP neighbor status.

```
Device B# display ip bgp-instance neighbors
BGP peer is 192.168.195.183, remote AS 200, local AS 100, external link
Using BFD to detect fast fallover - BFD session state up
  BGP version 4, remote router ID 10.0.0.1
  BGP state = Established, up for 00:06:37
  Last read 00:06:37, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
Address family IPv4 Unicast: advertised and received
Graceful restart: advertised and received
  Remote Restart timer is 120 seconds
  Address families preserved by peer:
    None
```

6. Configuration Files

- Device A configuration file

```
router-protocol bgp 200
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  exit-address-family-mode
!
router-protocol ospf 1
 gr
!
```

- Device B configuration file

```
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  exit-address-family-mode
!
router-protocol ospf 1
 gr
!
```

- Device C configuration file

```
router-protocol bgp 100
```

```
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  exit-address-family-mode
!
router-protocol ospf 1
 gr
!
```

- Device D configuration file

```
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  exit-address-family-mode
!
router-protocol ospf 1
 gr
!
```

7. Common Errors

- GR is not enabled for IGP.
- GR is not enabled for a BGP neighbor.

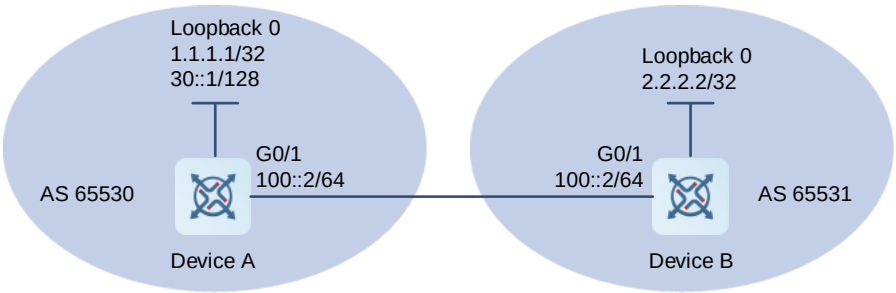
1.13.10 Configuring IPv6 Route Exchange Across ASs

1. Requirements

IPv6 routing information needs to be exchanged across different ASs.

2. Topology

Figure 1-11 Configuring IPv6 Route Exchange Across ASs



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-11 (omitted).

- Configure BGP neighbors, disable the IPv4 address family capability for the neighbors, and activate the IPv6 address family capability.
- Configure BGP to advertise IPv6 routes.

4. Procedure

(1) Configure BGP and activate the IPv6 address family.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65530
Device A(cfg-router)# peer 100::1 remote-as 65531
Device A(cfg-router)# address-family-mode ipv4
Device A(cfg-router-af)# unset peer 100::1 activate
Device A(cfg-router-af)# exit-address-family-mode
Device A(cfg-router)# address-family-mode ipv6
Device A(cfg-router-af)# peer 100::1 activate
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 65531
Device B(cfg-router)# peer 100::2 remote-as 65530
Device B(cfg-router)# address-family-mode ipv4
Device B(cfg-router-af)# unset peer 100::2 activate
Device B(cfg-router-af)# exit-address-family-mode
Device B(cfg-router)# address-family-mode ipv6
Device B(cfg-router-af)# peer 100::2 activate
```

(2) Configure the advertisement of IPv6 network information.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65530
Device A(cfg-router)# address-family-mode ipv6
Device A(cfg-router-af)# network-advertise 30::1/128
```

5. Verification

On Device A, check the BGP neighbor status.

```
Device A# display bgp-instance ipv6 unicast-af-modifier status
BGP router identifier 1.1.1.1, local AS number 65530
BGP table version is 1
1 BGP AS-PATH entries
0 BGP Community entries
1 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down       State/PfxRcd
100::1          4        65531       4       6        1    0    0 00:01:49            0

Total number of neighbors 1
```

On Device B, check the BGP neighbor status.

```
Device B# display bgp-instance ipv6 unicast-af-modifier
BGP table version is 4, local router ID is 2.2.2.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                    S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

    Network            Next Hop              Metric       LocPrf       Weight Path
*> 30::1/128          100::2                  0                              0 65530 i

Total number of prefixes 1
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65530
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 100::1 remote-as 65531
 address-family-mode ipv4
  unset peer 100::1 activate
  exit-address-family-mode
 address-family-mode ipv6
  network-advertise 30::1/128
  peer 100::1 activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 65531
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 100::2 remote-as 65530
 address-family-mode ipv4
  unset peer 100::2 activate
  exit-address-family-mode
 address-family-mode ipv6
  peer 100::2 activate
  exit-address-family-mode
!
```

7. Common Errors

- The IPv6 address family capability is not activated for BGP neighbors.

- In non-6PE scenarios, IPv4 addresses are used to establish a neighbor relationship to exchange IPv6 routes.

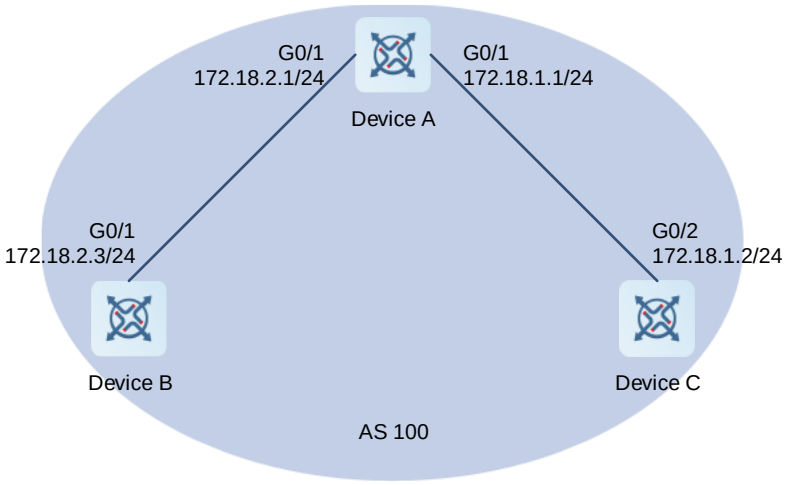
1.13.11 Configuring the BGP EVPN

1. Requirements

The L2VPN EVPN address family needs to be configured.

2. Topology

Figure 1-12 Configuring the BGP EVPN



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-12 (omitted).
- Configure BGP neighbors as shown in the figure (omitted).
- Configure device A as a BGP route reflector.
- Activate the L2VPN EVPN address family capability for the BGP neighbors.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 100
Device A(cfg-router)# peer 172.18.2.3 remote-as 100
Device A(cfg-router)# peer 172.18.1.2 remote-as 100
Device A(cfg-router)# address-family-mode l2vpn evpn
Device A(cfg-router-af)# peer 172.18.2.3 activate
Device A(cfg-router-af)# peer 172.18.1.2 activate
Device A(cfg-router-af)# peer 172.18.2.3 route-reflector-client
Device A(cfg-router-af)# peer 172.18.1.2 route-reflector-client
Device A(cfg-router-af)# exit-address-family-mode
```

Configure Device B.

```
Device B# config terminal
```

```
Device B(cfg)# evpn
Device B(cfg-evpn)# vni 100
Device B(cfg-evpn-vni)# rd auto
Device B(cfg-evpn-vni)# route-target both auto
Device B(cfg-evpn-vni)# exit
Device B(cfg-evpn)# exit
Device B(cfg)# router-protocol bgp 100
Device B(cfg-router)# peer 172.18.2.1 remote-as 100
Device B(cfg-router)# address-family-mode l2vpn evpn
Device B(cfg-router-af)# peer 172.18.2.1 activate
Device B(cfg-router-af)# exit-address-family-mode
```

Configure Device C.

```
Device C # config terminal
Device C(cfg)# evpn
Device C(cfg-evpn)# vni 100
Device C(cfg-evpn-vni)# rd auto
Device C(cfg-evpn-vni)# route-target both auto
Device C(cfg-evpn-vni)# exit
Device C(cfg-evpn)# exit
Device C(cfg)# router-protocol bgp 100
Device C(cfg-router)# peer 172.18.1.1 remote-as 100
Device C(cfg-router)# address-family-mode l2vpn evpn
Device C(cfg-router-af)# peer 172.18.1.1 activate
Device C(cfg-router-af)# exit-address-family-mode
```

5. Verification

On Device A, check L2VPN EVPN information.

```
Device A# display bgp-instance l2vpn evpn-af-modifier all status
For address family: L2VPN EVPN
BGP router identifier 3.3.3.3, local AS number 100
BGP table version is 51
1 BGP AS-PATH entries
0 BGP Community entries
6 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down  State/PfxRcd
172.18.2.3      4          100    1993    2033       49    0    0 1d04h40m           3
172.18.1.2      4          100     619     623       51    0    0 08:47:26           3

Total number of neighbors 2

Device A# display bgp-instance l2vpn evpn-af-modifier all
BGP table version is 51, local router ID is 3.3.3.3
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete
```

```
    Network              Next Hop              Metric       LocPrf       Weight Path
Route Distinguisher: 1.1.1.1:100
*>i0:6:0011.2233.2016:0:0.0.0.0/128
                         1.1.1.1                  0          100            0      i
*>i0:6:0011.2233.2016:32:100.1.1.2/128
                         1.1.1.1                  0          100            0      i
*>i0:32:1.1.1.1/72   1.1.1.1                  0          100            0      i

Total number of prefixes 3
Route Distinguisher: 2.2.2.2:100
*>i0:6:00d0.f822.33df:0:0.0.0.0/128
                         2.2.2.2                  0          100            0      i
*>i0:6:00d0.f822.33df:32:100.1.1.1/128
                         2.2.2.2                  0          100            0      i
*>i0:32:2.2.2.2/72   2.2.2.2                  0          100            0      i

Total number of prefixes 3
```

On Device B, check L2VPN EVPN information.

```
Device B# display bgp-instance l2vpn evpn all status
For address family: L2VPN EVPN
BGP router identifier 1.1.1.1, local AS number 100
BGP table version is 16
1 BGP AS-PATH entries
0 BGP Community entries
6 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down   State/PfxRcd
172.18.2.1      4          100     639     625       16    0    0 08:55:10        3

Total number of neighbors 1

Device B# display bgp-instance l2vpn evpn-af-modifier all
BGP table version is 16, local router ID is 1.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

    Network              Next Hop              Metric       LocPrf       Weight Path
Route Distinguisher: 1.1.1.1:100 (Default for EVI 100)
*> 0:6:0011.2233.2016:0:0.0.0.0/128
                         0.0.0.0                                         32768      i
*>i0:6:00d0.f822.33df:0:0.0.0.0/128
                         2.2.2.2                  0          100            0      i
*> 0:6:0011.2233.2016:32:100.1.1.2/128
                         0.0.0.0                                         32768      i
```

```
*>i0:6:00d0.f822.33df:32:100.1.1.1/128
                          2.2.2.2                  0          100          0      i
*> 0:32:1.1.1.1/72    0.0.0.0                                        32768      i
*>i0:32:2.2.2.2/72    2.2.2.2                  0          100          0      i

Total number of prefixes 6
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 172.18.1.2 remote-as 100
 peer 172.18.2.3 remote-as 100
 address-family-mode ipv4
  peer 172.18.1.2 activate
  peer 172.18.2.3 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 172.18.1.2 activate
  peer 172.18.1.2 route-reflector-client
  peer 172.18.1.2 send-community extended
  peer 172.18.2.3 activate
  peer 172.18.2.3 route-reflector-client
  peer 172.18.2.3 send-community extended
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 172.18.2.1 remote-as 100
 address-family-mode ipv4
  peer 172.18.2.1 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 172.18.2.1 activate
  peer 172.18.2.1 send-community extended
  exit-address-family-mode
!
```

```
evpn
 vni 100
  rd auto
  route-target both auto
!
```

- Device C configuration file

```
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 172.18.1.1 remote-as 100
 address-family-mode ipv4
  peer 172.18.1.1 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 172.18.1.1 activate
  peer 172.18.1.1 send-community extended
  exit-address-family-mode
!
evpn
 vni 100
  rd auto
  route-target both auto
!
```

7. Common Errors

- No neighbor is configured for the L2VPN EVPN address family.
- The EVPN address family capability of a BGP neighbor is not activated.

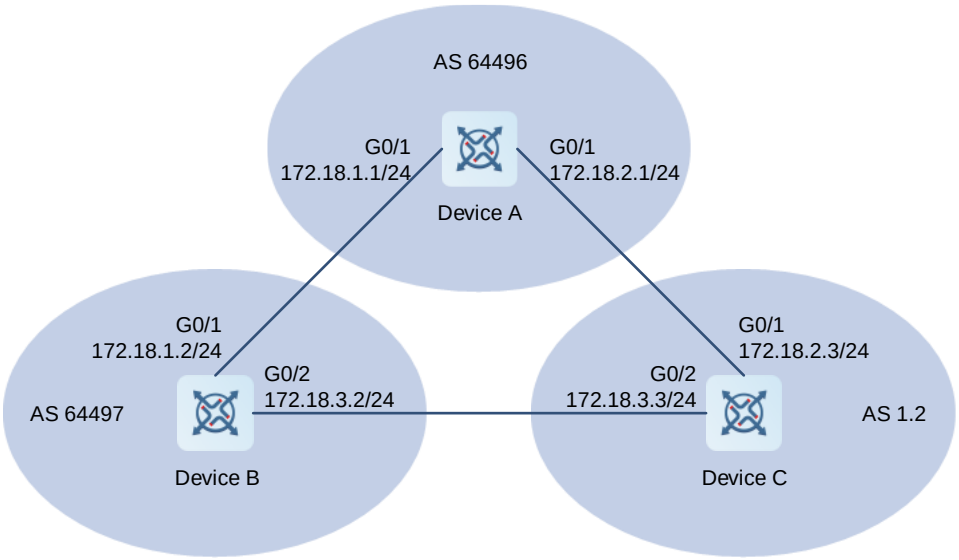
1.13.12 Configuring Compatibility Between BGP Devices Supporting 4-Byte AS Numbers and Those Supporting 2-Byte AS Numbers

1. Requirements

- Routers supporting 2-byte AS numbers and those supporting 4-byte AS numbers (using 2-byte AS numbers) can establish BGP connections.
- Routers supporting 2-byte AS numbers and those supporting 4-byte AS numbers (using 4-byte AS numbers) can establish BGP connections.
- Two routers supporting 4-byte AS numbers can establish a BGP connection, with one using a 2-byte AS number and the other using a 4-byte AS number.

2. Topology

Figure 1-13 Configuring Compatibility Between BGP Devices Supporting 4-Byte AS Numbers and Those Supporting 2-Byte AS Numbers



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-13 (omitted).
- Configure BGP neighbors.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 64496
Device A(cfg-router)# peer 172.18.1.2 remote-as 64497
Device A(cfg-router)# peer 172.18.2.3 remote-as 23456
Device A(cfg-router)# end
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 64497
Device B(cfg-router)# peer 172.18.1.1 remote-as 64496
Device B(cfg-router)# peer 172.18.3.3 remote-as 1.2
Device B(cfg-router)# bgp-instance as-symbol dot
Device B(cfg-router)# end
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# router-protocol bgp 1.2
Device C(cfg-router)# peer 172.18.2.1 remote-as 64496
Device C(cfg-router)# peer 172.18.3.2 remote-as 64497
Device C(cfg-router)# end
```

5. Verification

On Device A, check the BGP neighbor status.

```
Device A# display ip bgp-instance status

BGP router identifier 172.18.1.1, local AS number 64496
BGP table version is 1, main routing table version 1

Neighbor        V    AS        MsgRcvd    MsgSent    TblVer    InQ    OutQ    Up/Down      Statd
172.18.1.2      4    64497     7          7          1         0      0       00:03:04     0
172.18.2.3      4    23456     4          4          1         0      0       00:00:15     0
```

On Device B, check the BGP neighbor status.

```
Device B# display ip bgp-instance status

BGP router identifier 172.18.3.2, local AS number 64497
BGP table version is 1, main routing table version 1

Neighbor        V    AS        MsgRcvd    MsgSent    TblVer    InQ    OutQ    Up/Down      Statd
172.18.1.1      4    64496     7          7          1         0      0       00:00:04     0
172.18.3.2      4    1.2       4          4          1         0      0       00:00:16     0
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 64496
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 172.18.1.2 remote-as 64497
 peer 172.18.2.3 remote-as 23456
 address-family-mode ipv4
  peer 172.18.1.2 activate
  peer 172.18.2.3 activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 64497
 bgp-instance as-symbol dot
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 172.18.1.1 remote-as 64496
```

```
 peer 172.18.3.3 remote-as 1.2
 address-family-mode ipv4
  peer 172.18.1.1 activate
  peer 172.18.3.3 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
router-protocol bgp 65538
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 172.18.2.1 remote-as 64496
 peer 172.18.3.2 remote-as 64497
 address-family-mode ipv4
  peer 172.18.2.1 activate
  peer 172.18.3.2 activate
  exit-address-family-mode
!
```

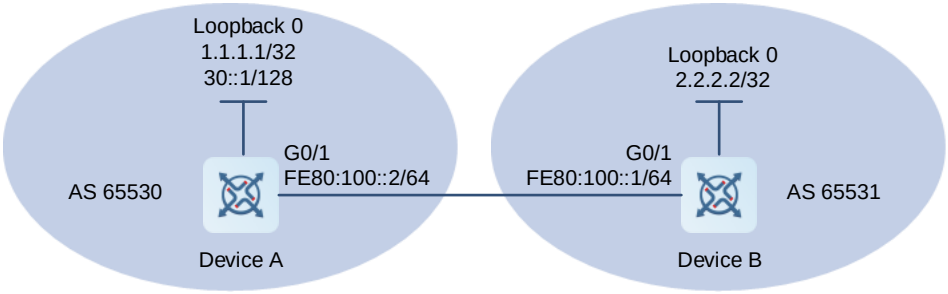
1.13.13 Configuring an IPv6 Local Link Address

1. Requirements

An IPv6 local link address needs to be configured as a neighbor update address.

2. Topology

Figure 1-14 Configuring IPv6 Local Link Address



3. Notes

- Enable BGP on all devices, configure IPv6 local link addresses for interfaces, and set the AS numbers for the devices as shown in Figure 1-14 (omitted).
- Configure BGP neighbors, disable the IPv4 address family capability for the neighbors, and activate the IPv6 address family capability.
- Configure BGP to advertise IPv6 routes.

4. Procedure

(1) Enable the IPv6 address family capability and disable the IPv4 address family capability.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65530
Device A(cfg-router)# peer fe80:100::1%Gi0/1 remote-as 65531
Device A(cfg-router)# address-family-mode ipv4
Device A(cfg-router-af)# unset peer fe80:100::1%Gi0/1 activate
Device A(cfg-router-af)# exit-address-family-mode
Device A(cfg-router)# address-family-mode ipv6
Device A(cfg-router-af)# peer fe80:100::1%Gi0/1 activate
Device A(cfg-router-af)# network-advertise 30::1/128
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 65531
Device B(cfg-router)# peer fe80:100::2%Gi0/1 remote-as 65530
Device B(cfg-router)# address-family-mode ipv4
Device B(cfg-router-af)# unset peer fe80:100::2%Gi0/1 activate
Device B(cfg-router-af)# exit-address-family-mode
Device B(cfg-router)# address-family-mode ipv6
Device B(cfg-router-af)# peer fe80:100::2%Gi0/1 activate
```

(2) Configure BGP to advertise IPv6 addresses.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65530
Device A(cfg-router)# address-family-mode ipv6
Device A(cfg-router-af)# network-advertise 30::1/128
```

5. Verification

On Device A, check the BGP neighbor status.

```
Device A# display bgp-instance ipv6 unicast-af-modifier status
BGP router identifier 1.1.1.1, local AS number 65530
BGP table version is 1
1 BGP AS-PATH entries
0 BGP Community entries
1 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor                V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down
State/PfxRcd
fe80:100::1%Gi0/1       4        65531       4       6        1    0    0 00:01:49
0

Total number of neighbors 1
```

On Device B, check the BGP neighbor status.

```
Device B# display bgp-instance ipv6 unicast-af-modifier
```

```
BGP table version is 4, local router ID is 2.2.2.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                    S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric     LocPrf     Weight Path
*> 30::1/128        fe80:100::2         0                     0 65530 i

Total number of prefixes 1
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65530
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer fe80:100::1%Gi0/1 remote-as 65531
 address-family-mode ipv4
  unset peer fe80:100::1%Gi0/1 activate
  exit-address-family-mode
 address-family-mode ipv6
  network-advertise 30::1/128
  peer fe80:100::1%Gi0/1 activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 65531
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer fe80:100::2%Gi0/1 remote-as 65530
 address-family-mode ipv4
  unset peer fe80:100::2%Gi0/1 activate
  exit-address-family-mode
 address-family-mode ipv6
  peer fe80:100::2%Gi0/1 activate
  exit-address-family-mode
!
```

7. Common Errors

- An IPv6 local link address is used to establish a neighbor relationship during neighbor configuration, but the interface, on which the IPv6 local link address is configured, is not configured as the update source during

local configuration.

- Only one end uses an IPv6 local link address for establishing a neighbor relationship.

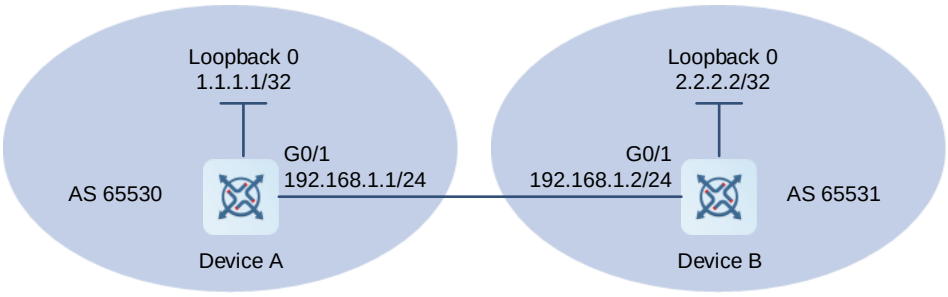
1.13.14 Configuring BGP NSR

1. Requirements

NSR can be configured to ensure that data switching and forwarding are not interrupted.

2. Topology

Figure 1-15 Configuring BGP NSR



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-15 (omitted).
- Configure devices A and B to establish an EBGP neighbor relationship through directly connected interfaces.
- Enable the BGP NSR function on device A.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65530
Device A(cfg-router)# peer 192.168.1.2 remote-as 65530
Device A(cfg-router)# peer 192.168.1.2 ha-mode nsr
```

5. Verification

On device A, check the BGP neighbor status.

```
Device A# display ip bgp-instance peers
BGP peer is 192.168.1.2, remote AS 65530, local AS 65531, external link
Using BFD to detect fast fallover - BFD session state up
  BGP version 4, remote router ID 2.2.2.2
  BGP state = Established, up for 00:06:37
  Last read 00:06:37, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
Address family IPv4 Unicast: advertised and received
Graceful restart: advertised and received
  Remote Restart timer is 120 seconds
```

```
  Address families preserved by peer:
    None
```

On device B, check the BGP neighbor status.

```
Device B# display ip bgp-instance peers
BGP peer is 192.168.1.1, remote AS 65530, local AS 65531, external link
Using BFD to detect fast fallover - BFD session state up
  BGP version 4, remote router ID 1.1.1.1
  BGP state = Established, up for 00:06:37
  Last read 00:06:37, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
Address family IPv4 Unicast: advertised and received
Graceful restart: advertised and received
  Remote Restart timer is 120 seconds
  Address families preserved by peer:
    None
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65530
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 bgp-instance nsr
 peer 192.168.1.2 remote-as 65530
 address-family-mode ipv4
  peer 192.168.1.2 activate
  exit-address-family-mode
!
```

7. Common Errors

- IGP is not configured.
- The NSR function is not enabled for BGP neighbors.

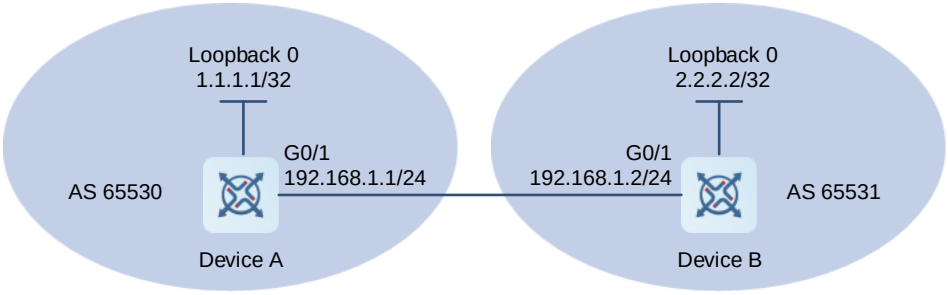
1.13.15 Configuring BGP Routes to Be Recursive Only to Host Routes

1. Requirements

Non-direct BGP routes need to be recursive to host routes.

2. Topology

Figure 1-16 Configuring BGP Routes to Be Recursive Only to Host Routes



3. Notes

- Enable BGP on all devices and set the AS numbers as shown in the figure above (omitted).
- Configure devices A and B to establish an EBGP neighbor relationship through loopback interfaces (omitted).
- On devices A and B, enable the function of making BGP routes recursive only to host routes.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65530
Device A(cfg-router)# bgp-instance recursive host-route
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 65531
Device B(cfg-router)# bgp-instance recursive host-route
```

5. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65530
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  bgp-instance recursive host-route
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 65531
 bgp-instance log-peer-changes
```

```
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  bgp-instance recursive host-route
  exit-address-family-mode
!
```

6. Common Errors

If an IBGP or multi-hop EBGP neighbor relationship is established not using loopback addresses, after the function is enabled, BGP cannot to find host routes matching the next-hop addresses. As a result, the next hops of BGP routes are invalid and routes cannot be preferentially selected.

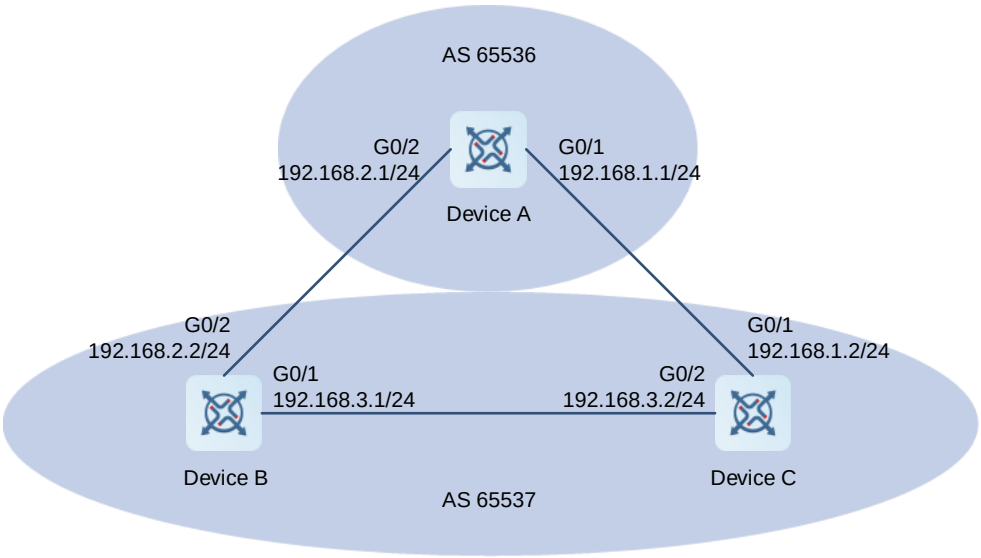
1.13.16 Configuring Outbound Loop Detection for a BGP Neighbor

1. Requirements

Outbound loop detection needs to be configured for a BGP neighbor to prevent the receiving of routing information carrying the AS number of the neighbor in the AS-PATH attribute.

2. Topology

Figure 1-17 Configuring Outbound Loop Detection for a BGP Neighbor



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-17 (omitted).
- Configure devices A and C and devices B and C to establish EBGP neighbor relationships (omitted).
- On device C, enable outbound loop detection for its neighbors: devices A and B.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65536
```

```
Device A(cfg-router)# peer 192.168.1.2 remote-as 65537
Device A(cfg-router)# peer 192.168.1.2 as-loop-check out
Device A(cfg-router)# peer 192.168.2.2 remote-as 65537
Device A(cfg-router)# peer 192.168.2.2 as-loop-check out
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 65537
Device B(cfg-router)# peer 192.168.2.1 remote-as 65536
Device B(cfg-router)# peer 192.168.3.2 remote-as 65537
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# router-protocol bgp 65537
Device C(cfg-router)# peer 192.168.1.1 remote-as 65536
Device C(cfg-router)# peer 192.168.3.1 remote-as 65537
```

5. Verification

On Device A, check the BGP neighbor status.

```
Device A# display ip bgp-instance peers 192.168.2.2
BGP peer is 192.168.2.2, remote AS 65537, local AS 65536, external link
Using as path loop detection in announcing route
  BGP version 4, remote router ID 192.168.2.2
  BGP state = Established, up for 00:06:37
  Last read 00:06:37, hold time is 180, keepalive interval is 60 seconds
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
Address family IPv4 Unicast: advertised and received
Graceful restart: advertised and received
  Remote Restart timer is 120 seconds
  Address families preserved by peer:
    None
……
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.2 remote-as 65537
 peer 192.168.2.2 remote-as 65537
 address-family-mode ipv4
  peer 192.168.1.2 activate
  peer 192.168.1.2 as-loop-check out
```

```
  peer 192.168.2.2 activate
  peer 192.168.2.2 as-loop-check out
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 65537
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.2.1 remote-as 65536
 peer 192.168.3.2 remote-as 65537
 address-family-mode ipv4
  peer 192.168.2.1 activate
  peer 192.168.3.2 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
router-protocol bgp 65537
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.1 remote-as 65536
 peer 192.168.3.1 remote-as 65537
 address-family-mode ipv4
  peer 192.168.1.1 activate
  peer 192.168.3.1 activate
  exit-address-family-mode
!
```

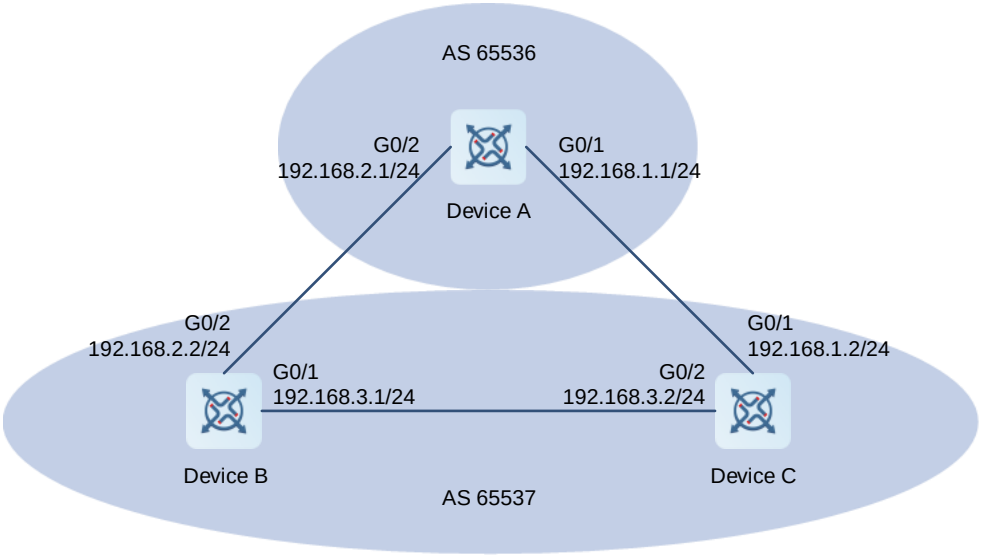
1.13.17 Shutting Down BGP Connections Gracefully

1. Requirements

Graceful shutdown of BGP connections needs to be configured on a device. Then, after a certain period of time (automatically calculated based on the number of advertised routes or specified by *delay time*), the device disconnects BGP connections from its neighbors to reduce the impact on the network.

2. Topology

Figure 1-18 Configuring Graceful Shutdown of BGP Connections



3. Notes

- Enable BGP on all devices and set AS numbers for the devices as shown in Figure 1-18 (omitted).
- Configure devices A and C and devices B and C to establish EBGP neighbor relationships (omitted).
- On device A, configure the graceful shutdown of all connections in the BGP instance.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65536
Device A(cfg-router)# peer 192.168.1.2 remote-as 65537
Device A(cfg-router)# peer 192.168.1.2 as-loop-check out
Device A(cfg-router)# peer 192.168.2.2 remote-as 65537
Device A(cfg-router)# peer 192.168.2.2 as-loop-check out
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 65537
Device B(cfg-router)# peer 192.168.2.1 remote-as 65536
Device B(cfg-router)# peer 192.168.3.2 remote-as 65537
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# router-protocol bgp 65537
Device C(cfg-router)# peer 192.168.1.1 remote-as 65536
Device C(cfg-router)# peer 192.168.3.1 remote-as 65537
Device C(cfg-router)# bgp-instance shutdown graceful
```

5. Verification

On Device C, check the BGP neighbor status.

```
Device C(cfg)#display ip bgp-instance peer 192.168.1.1
For address family: IPv4 Unicast
BGP peer is 192.168.1.1, remote AS 65536, local AS 65537, external link
 Administratively graceful shut down
  BGP version 4, remote router ID 0.0.0.0
  BGP state = Idle
  Last read               , hold time is 30, keepalive interval is 1 seconds
  Configured hold time is 30, keepalive interval is 1 seconds
  Received 0 messages, 0 notifications, 0 in queue
    open message:0 update message:0 keepalive message:0
    refresh message:0 dynamic cap:0 notifications:0
  Sent 0 messages, 0 notifications, 0 in queue
    open message:0 update message:0 keepalive message:0
    refresh message:0 dynamic cap:0 notifications:0
  Route refresh request: received 0, sent 0
  Minimum time between advertisement runs is 30 seconds
  Update source is Loopback 0

 For address family: IPv4 Unicast
  BGP table version 1, peer version 0
  Index 1, Offset 0, Mask 0x2
  0 accepted prefixes
  0 announced prefixes

 Connections established 0; dropped 0
  BGP peer may be up to 255 hops away.
Sock_fd: -1
Last Reset:               , due to BGP Notification received
Notification Error Message: (Cease/Other Configuration Change.)
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.2 remote-as 65537
 peer 192.168.2.2 remote-as 65537
 address-family-mode ipv4
  peer 192.168.1.2 activate
  peer 192.168.1.2 as-loop-check out
```

```
  peer 192.168.2.2 activate
  peer 192.168.2.2 as-loop-check out
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 65537
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.2.1 remote-as 65536
 peer 192.168.3.2 remote-as 65537
 address-family-mode ipv4
  peer 192.168.2.1 activate
  peer 192.168.3.2 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
router-protocol bgp 65537
 bgp-instance shutdown graceful
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.1 remote-as 65536
 peer 192.168.3.1 remote-as 65537
 address-family-mode ipv4
  peer 192.168.1.1 activate
  peer 192.168.3.1 activate
  exit-address-family-mode
!
```

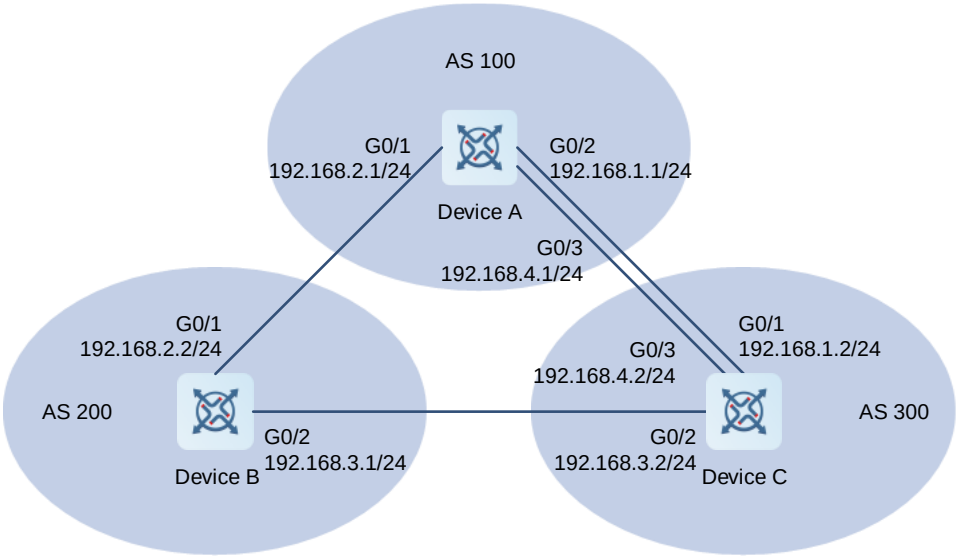
1.13.18 Configuring BGP Multi-Path Bypass Protection

1. Requirements

Multi-path bypass protection needs to be configured to enhance routing reliability.

2. Topology

Figure 1-19 Configuring BGP Multi-Path Bypass Protection



3. Notes

- Enable BGP on all devices (omitted).
- Configure addresses for directly connected interfaces on devices A, B, and C, and configure them to establish EBGP neighbor relationships (omitted).
- Configure a BFD session for the EBGP neighbor relationship between devices A and C (omitted).
- On device C, configure BGP multi-path bypass protection and ECMP.

4. Procedure

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 100
Device A(cfg-router)# peer 192.168.1.2 remote-as 300
Device A(cfg-router)# peer 192.168.4.2 remote-as 300
Device A(cfg-router)# peer 192.168.2.2 remote-as 200
Device A(cfg-router)# max-ecmp-paths ebgp 2
Device A(cfg-router)# redistribute-route connected
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 200
Device B(cfg-router)# peer 192.168.3.2 remote-as 300
Device B(cfg-router)# peer 192.168.3.2 fall-over bfd
Device B(cfg-router)# peer 192.168.2.1 remote-as 100
Device B(cfg-router)# redistribute-route connected
```

Configure Device C.

```
Device C# config-system terminal
```

```
Device C(cfg)# router-protocol bgp 300
Device C(cfg-router)# peer 192.168.1.1 remote-as 100
Device C(cfg-router)# peer 192.168.4.1 remote-as 100
Device C(cfg-router)# peer 192.168.3.1 remote-as 200
Device C(cfg-router)# peer 192.168.1.1 fall-over bfd
Device C(cfg-router)# peer 192.168.4.1 fall-over bfd
Device C(cfg-router)# max-ecmp-paths ebgp 2
Device C(cfg-router)# address-family-mode ipv4 unicast
Device C(cfg-router-af)# bgp-instance install standby-path
Device C(cfg-router-af)# redistribute-route connected
```

5. Verification

On Device C, check BGP information.

```
Device C# display ip bgp-instance status
BGP router identifier 192.168.4.2, local AS number 300
BGP table version is 12
4 BGP AS-PATH entries
0 BGP Community entries
4 BGP Prefix entries (Maximum-prefix:4294967295)

Neighbor        V           AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down       State/PfxRcd
192.168.1.1     4          100      76      77       12    12    0 00:59:27            3
192.168.3.1     4          200      30      30       12    12    0 00:19:03            3
192.168.4.1     4          100      76      77       12    12    0 00:59:01            3

Total number of neighbors 3

Device C# display ip bgp-instance
BGP table version is 12, local router ID is 10.10.10.10
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                    S Stale, b - backup entry, m - multipath
Origin codes: i - IGP, e - EGP, ? - incomplete

     Network            Next Hop            Metric     LocPrf     Weight Path
*    192.168.1.0        192.168.3.1              0                    0 200 ?
*                       192.168.4.1              0                    0 100 ?
*                       192.168.1.1              0                    0 100 ?
*>                      0.0.0.0                  0                32768 ?
*> 192.168.2.0         192.168.4.1              0                    0 200 ?
*m                      192.168.1.1              0                    0 200 ?
*b                      192.168.3.1              0                    0 100 ?
*    192.168.3.0        192.168.3.1              0                    0 200 ?
*                       192.168.4.1              0                    0 100 200 ?
*                       192.168.1.1              0                    0 100 200 ?
*>                      0.0.0.0                  0                32768 ?
*    192.168.4.0        192.168.3.1              0                    0 200 ?
```

```
*                    192.168.4.1              0                      0 100 ?
*                    192.168.1.1              0                      0 100 ?
*>                   0.0.0.0                  0                  32768     ?

Total number of prefixes 4
Device C# display ip bgp-instance 192.168.2.0
BGP routing table entry for 192.168.2.0/24
Paths: (3 available, best #3, table Default-IP-Routing-Table)
  Advertised to non peer-group peers:
  192.168.1.1
  200
    192.168.3.1 from 192.168.3.1 (3.3.3.3)
      Origin incomplete, metric-value 0, localpref 100, valid, external, backup
      Last update: Tue Oct  5 00:26:52 1971

  100
    192.168.1.1 from 192.168.1.1 (44.44.44.44)
      Origin incomplete, metric-value 0, localpref 100, valid, external, multipath
      Last update: Mon Oct  4 23:46:28 1971

  100
    192.168.4.1 from 192.168.4.1 (44.44.44.44)
      Origin incomplete, metric-value 0, localpref 100, valid, external, multipath, best
      Last update: Mon Oct  4 23:46:28 1971
Device C# display ip route-config

Codes:  C - connected, S - static, R - RIP, B - BGP
        O - OSPF, IA - OSPF inter area
        N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
        E1 - OSPF external type 1, E2 - OSPF external type 2
        i - IS-IS, su - IS-IS status, L1 - IS-IS level-1, L2 - IS-IS level-2
        ia - IS-IS inter area, * - candidate reset

Gateway of last resort is unset set
C       192.168.1.0/24 is directly connected, GigabitEthernet 0/1
C       192.168.1.2/32 is local host.
B       192.168.2.0/24 [20/0] via 192.168.1.1, 00:21:39
                       [20/0] via 192.168.4.1, 00:21:39
C       192.168.3.0/24 is directly connected, GigabitEthernet 0/2
C       192.168.3.2/32 is local host.
C       192.168.4.0/24 is directly connected, GigabitEthernet 0/3
C       192.168.4.2/32 is local host.
```

6. Configuration Files

- Device A configuration file

```
!
```

```
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.2 remote-as 300
 peer 192.168.2.2 remote-as 200
 peer 192.168.4.2 remote-as 300
 address-family-mode ipv4
  max-ecmp-paths ebgp 2
  redistribute-route connected
  peer 192.168.1.2 activate
  peer 192.168.2.2 activate
  peer 192.168.4.2 activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
router-protocol bgp 200
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.2.1 remote-as 100
 peer 192.168.3.2 remote-as 300
 peer 192.168.3.2 fall-over bfd
 address-family-mode ipv4
  redistribute-route connected
  peer 192.168.2.1 activate
  peer 192.168.3.2 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
router-protocol bgp 300
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.1 remote-as 100
 peer 192.168.1.1 fall-over bfd
 peer 192.168.3.1 remote-as 200
 peer 192.168.4.1 remote-as 100
 peer 192.168.4.1 fall-over bfd
 address-family-mode ipv4
```

```
 max-ecmp-paths ebgp 2
 bgp-instance install standby-path
 redistribute-route connected
 peer 192.168.1.1 activate
 peer 192.168.3.1 activate
 peer 192.168.4.1 activate
 exit-address-family-mode
!
```

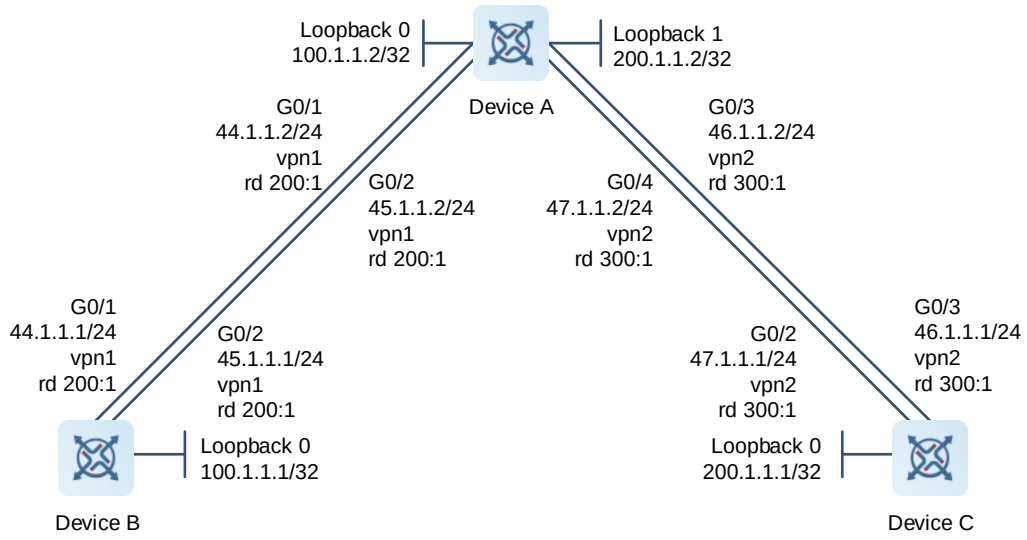
1.13.19 Configuring Multi-Path Route Import Between VRF Instances

1. Requirements

Device A is connected with two networks through the interfaces of two VRF instances (vpn1 and vpn2) so that cross-VRF access is implemented through device A.

2. Topology

Figure 1-20 Configuring Multi-Path Route Import Between VRF Instances



3. Notes

- Configure VRF instances and VRF static routes.
- Configure a VRF address family.
- Import VRF static routes to BGP.
- Enable the function of importing multi-path static routes to BGP.
- Configure BGP ECMP.
- Configure the import of all path routes between VRF instances.

4. Procedure

(1) Configure VRF instances and VRF static routes.

Configure Device A.

```
Device A# config-system terminal
```

```
Device A(cfg)# ip vrf vpn1
Device A(cfg-vrf)# rd 200:1
Device A(cfg-vrf)# route-target both 100:100
Device A(cfg-vrf)# exit
Device A(cfg)# ip vrf vpn2
Device A(cfg-vrf)# rd 300:1
Device A(cfg-vrf)# route-target both 100:100
Device A(cfg-vrf)# exit
Device A(cfg)# interface GigabitEthernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip vrf bindingvpn1
Device A(cfg-if-GigabitEthernet 0/1)# ip address 44.1.1.2 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface GigabitEthernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# ip vrf bindingvpn1
Device A(cfg-if-GigabitEthernet 0/2)# ip address 45.1.1.2 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/2)# exit
Device A(cfg)# interface GigabitEthernet 0/3
Device A(cfg-if-GigabitEthernet 0/3)# ip vrf bindingvpn2
Device A(cfg-if-GigabitEthernet 0/3)# ip address 46.1.1.2 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/3)# exit
Device A(cfg)# interface GigabitEthernet 0/4
Device A(cfg-if-GigabitEthernet 0/4)# ip vrf bindingvpn2
Device A(cfg-if-GigabitEthernet 0/4)# ip address 47.1.1.2 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/4)# exit
Device A(cfg)# ip route-config vrf vpn1 100.1.1.1 255.255.255.255 44.1.1.1
Device A(cfg)# ip route-config vrf vpn1 100.1.1.1 255.255.255.255 45.1.1.1
Device A(cfg)# ip route-config vrf vpn2 200.1.1.1 255.255.255.255 46.1.1.1
Device A(cfg)# ip route-config vrf vpn2 200.1.1.1 255.255.255.255 47.1.1.1
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# ip vrf vpn1
Device B(cfg-vrf)# rd 200:1
Device B(cfg-vrf)# route-target both 100:100
Device B(cfg-vrf)# exit
Device B(cfg)# interface GigabitEthernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip vrf bindingvpn1
Device B(cfg-if-GigabitEthernet 0/1)# ip address 44.1.1.1 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface GigabitEthernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# ip vrf bindingvpn1
Device B(cfg-if-GigabitEthernet 0/2)# ip address 45.1.1.1 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/2)# exit
Device B(cfg)# ip route-config vrf vpn1 100.1.1.2 255.255.255.255 44.1.1.2
Device B(cfg)# ip route-config vrf vpn1 100.1.1.2 255.255.255.255 45.1.1.2
```

Configure Device C.

```
Device C# config-system terminal
Device C(cfg)# ip vrf vpn2
Device C(cfg-vrf)# rd 300:1
Device C(cfg-vrf)# route-target both 100:100
Device C(cfg-vrf)# exit
Device C(cfg)# interface GigabitEthernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# ip vrf bindingvpn2
Device C(cfg-if-GigabitEthernet 0/1)# ip address 46.1.1.1 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface GigabitEthernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# ip vrf bindingvpn2
Device C(cfg-if-GigabitEthernet 0/2)# ip address 47.1.1.1 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/2)# exit
Device C(cfg)# ip route-config vrf vpn2 200.1.1.2 255.255.255.255 46.1.1.2
Device C(cfg)# ip route-config vrf vpn2 200.1.1.2 255.255.255.255 47.1.1.2
```

(2) Configure the function of importing multi-path static routes to BGP to implement inter-VRF route import.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 100
Device A(cfg-router)# address-family-mode ipv4 vrf vpn1
Device A(cfg-router-af)# redistribute-route static
Device A(cfg-router-af)# max-ecmp-paths ebgp 32
Device A(cfg-router-af)# bgp-instance src-paths static all
Device A(cfg-router-af)# import route strategy all
Device A(cfg-router-af)# exit-address-family-mode
Device A(cfg-router)# address-family-mode ipv4 vrf vpn2
Device A(cfg-router-af)# redistribute-route static
Device A(cfg-router-af)# max-ecmp-paths ebgp 32
Device A(cfg-router-af)# bgp-instance src-paths static all
Device A(cfg-router-af)# import route strategy all
Device A(cfg-router-af)# exit-address-family-mode
```

5. Verification

On Device A, check VRF routes.

```
Device A# display ip route-config vrf-info vpn1
Routing Table: vpn1

Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPF, B - BGP, I - IS-IS, V - Overflow route
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         SU - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area, EV - BGP EVPN, A - Arp to host-route
         * - candidate reset

Gateway of last resort is unset set
```

```
C       44.1.1.0/24 is directly connected, GigabitEthernet 0/1
C       44.1.1.2/32 is local host-route.
C       45.1.1.0/24 is directly connected, GigabitEthernet 0/2
C       45.1.1.2/32 is local host-route.
S       100.1.1.1/32 [1/0] via 44.1.1.1
                           [1/0] via 45.1.1.1
B       200.1.1.1/32 [20/0] via 47.1.1.1, 02:32:01
                           [20/0] via 46.1.1.1, 02:32:01

Device A# display ip route-config vrf-info vpn2
Routing Table: vpn2

Codes:   C - Connected, L - Local, S - Static
         R - RIP, O - OSPF, B - BGP, I - IS-IS, V - Overflow route
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         SU - IS-IS status, L1 - IS-IS level-1, L2 - IS-IS level-2
         IA - Inter area, EV - BGP EVPN, A - Arp to host-route
         * - candidate reset

Gateway of last resort is unset set
C       46.1.1.0/24 is directly connected, GigabitEthernet 0/3
C       46.1.1.2/32 is local host-route.
C       47.1.1.0/24 is directly connected, GigabitEthernet 0/4
C       47.1.1.2/32 is local host-route.
B       100.1.1.1/32 [20/0] via 45.1.1.1, 03:27:07
                           [20/0] via 44.1.1.1, 03:27:07
S       200.1.1.1/32 [1/0] via 46.1.1.1
                           [1/0] via 47.1.1.1
```

6. Configuration Files

- Device A configuration file

```
!
ip vrf vpn1
 rd 200:1
 route-target both 100:100
!
ip vrf vpn2
 rd 300:1
 route-target both 100:100
!
interface GigabitEthernet 0/1
 unset switch-port
 ip vrf bindingvpn1
 ip address 44.1.1.2 255.255.255.0
!
```

```
interface GigabitEthernet 0/2
 unset switch-port
 ip vrf bindingvpn1
 ip address 45.1.1.2 255.255.255.0
!
interface GigabitEthernet 0/3
 unset switch-port
 ip vrf bindingvpn2
 ip address 46.1.1.2 255.255.255.0
!
interface GigabitEthernet 0/4
 unset switch-port
 ip vrf bindingvpn2
 ip address 47.1.1.2 255.255.255.0
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 address-family-mode ipv4
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vpn1
  max-ecmp-paths ebgp 32
  bgp-instance src-paths static all
  import route strategy all
  redistribute-route static
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vpn2
  max-ecmp-paths ebgp 32
  bgp-instance src-paths static all
  import route strategy all
  redistribute-route static
  exit-address-family-mode
!
ip route-config vrf vpn1 100.1.1.1 255.255.255.255 44.1.1.1
ip route-config vrf vpn1 100.1.1.1 255.255.255.255 45.1.1.1
ip route-config vrf vpn2 200.1.1.1 255.255.255.255 46.1.1.1
ip route-config vrf vpn2 200.1.1.1 255.255.255.255 47.1.1.1
!
```

- Device B configuration file

```
!
ip vrf vpn1
 rd 200:1
```

```
 route-target both 100:100
!
interface GigabitEthernet 0/1
 unset switch-port
 ip vrf bindingvpn1
 ip address 44.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 ip vrf bindingvpn1
 ip address 45.1.1.1 255.255.255.0
!
ip route-config vrf vpn1 100.1.1.2 255.255.255.255 44.1.1.2
ip route-config vrf vpn1 100.1.1.2 255.255.255.255 45.1.1.2
!
```

- Device C configuration file

```
!
ip vrf vpn2
 rd 300:1
 routing-target both 100:100
!
interface GigabitEthernet 0/1
 unset switch-port
 ip vrf bindingvpn2
 ip address 46.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 ip vrf bindingvpn2
 ip address 47.1.1.1 255.255.255.0
!
ip route-config vrf vpn2 200.1.1.2 255.255.255.255 46.1.1.2
ip route-config vrf vpn2 200.1.1.2 255.255.255.255 47.1.1.2
!
```

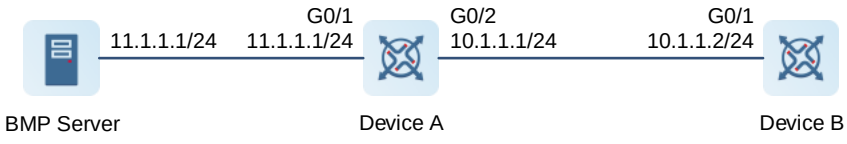
1.13.20 Configuring BMP Monitoring

1. Requirements

The BGP connections established between devices A and B need to be monitored by the BMP server.

2. Topology

Figure 1-21 Configuring BMP Monitoring



3. Notes

- Configure BGP basic features (omitted).
- Configure information relevant to a BMP server instance (address, port number, type of monitored routes, and MD5 authentication password for TCP connection).
- Configure the BMP server to monitor BGP peers.

4. Procedure

(1) Configure information related to a BMP server instance.

Configure device A.

```
Device A# config-system terminal
Device A(cfg)# bmp server 1
Device A(cfg-bmpsrvr)# address 11.1.1.2 port 12345
Device A(cfg-bmpsrvr)# update-source gigabitEthernet 0/1
Device A(cfg-bmpsrvr)# adj-rib-in post-policy
Device A(cfg-bmpsrvr)# password Red-Giant
```

(2) Configure the BMP server to monitor BGP peers.

Configure device A.

```
Device A(cfg)# router-protocol bgp 100
Device A(cfg-router)# peer 10.1.1.2 remote-as 100
Device A(cfg-router)# peer 10.1.1.2 bmp-active server 1
```

5. Verification

On device A, check BGP information.

```
Device A# display bgp-instance all status
For address family: IPv4 Unicast
BGP router identifier 11.1.1.1, local AS number 100
BGP table version is 1
1 BGP AS-PATH entries
0 BGP Community entries
25 BGP Prefix entries (Maximum-prefix:262144)

Neighbor        V         AS MsgRcvd MsgSent   TblVer  InQ OutQ Up/Down   State/PfxRcd
10.1.1.2        4        100       0       0        0    0    0 05:36:32        23

Device A# display bgp-instance bmp peer

Neighbor        CfgSvr#         ActSvr#
10.1.1.2        1               1

Device A# display bgp-instance bmp status
ID   Host                 Port      State        Time        NBRs
1    11.1.1.2             12345     Established  05:36:12    1
```

6. Configuration Files

- Device A configuration file

```
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 10.1.1.2 remote-as 100
 peer 10.1.1.2 bmp-active server 1
 address-family-mode ipv4
  peer 10.1.1.2 activate
  exit-address-family-mode
!
bmp server 1
 address 11.1.1.2 port 12345
 update-source GigabitEthernet 0/1
 adj-rib-in post-policy
 password Red-Giant
!
```

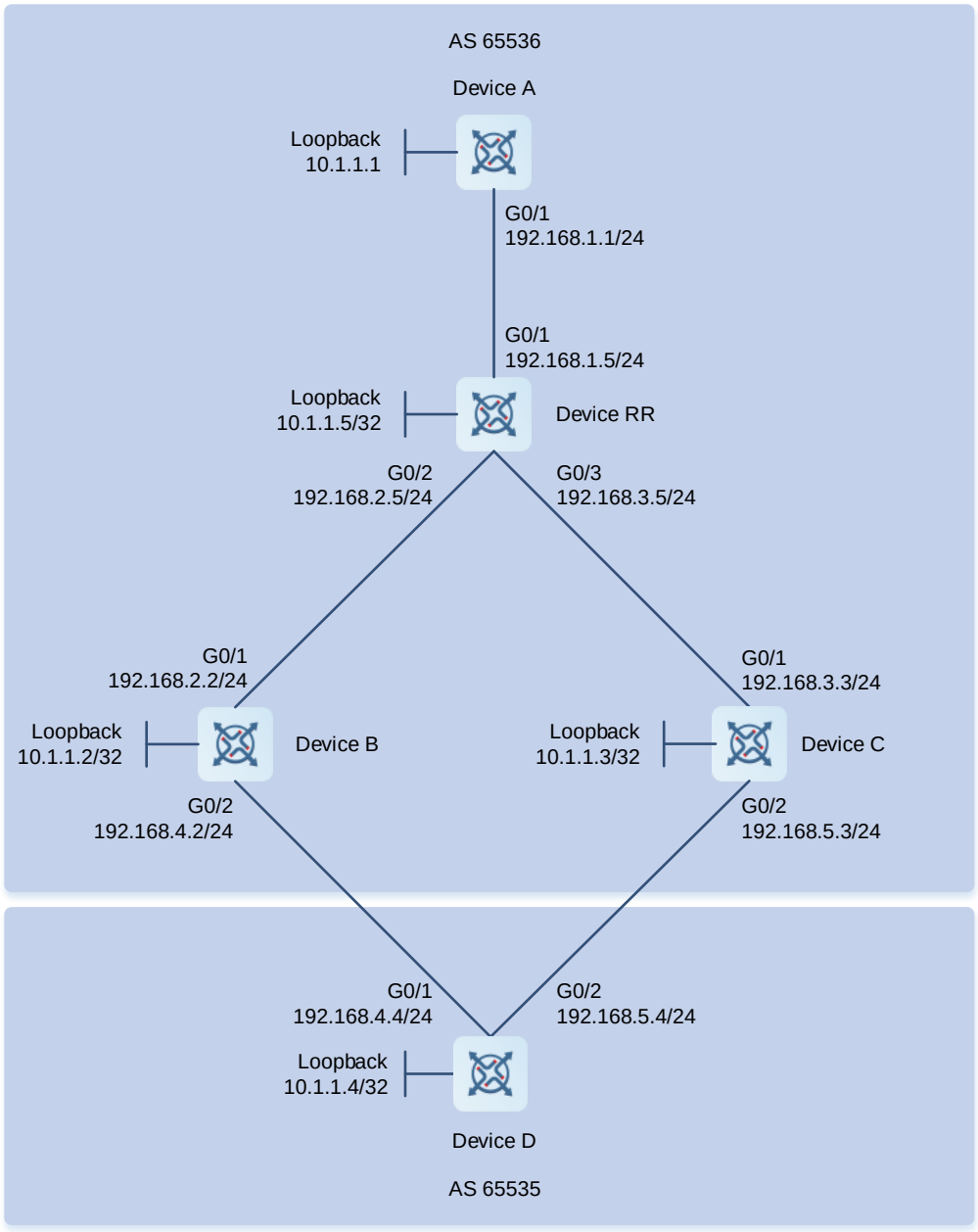
1.13.21 Configuring BGP ADD-PATH

1. Requirements

When the device is allowed to advertise multiple routes with the same prefix to a neighbor, you can define ADD-PATH to enable multiple links to implement load sharing or backup, thereby increasing the network link reliability.

2. Topology

Figure 1-22 Configuring BGP ADD-PATH



3. Notes

- As shown in Figure 1-22, configure interconnection addresses for devices A, B, C, D, and RR, configure devices A, B, C, and RR to establish IBGP neighbor relationships, configure device RR as a BGP route reflector, and configure devices B and D and devices C and D to establish EBGP neighbor relationships.
- Configure a loopback address on device D and advertise it to BGP.
- Configure the ADD-PATH advertisement capability on device RR and the ADD-PATH receiving capability on device A.

4. Procedure

(1) Configure BGP basic features and interface IP addresses on devices A, B, C, and D.

Configure device A.

```
Device A# config-system terminal
Device A(cfg)# interface GigabitEthernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.1 255.255.255.0
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# router-protocol bgp 65536
Device A(cfg-router)# peer 192.168.1.5 remote-as 65536
Device A(cfg-router)# peer 192.168.1.5 alternative-paths recv
```

Configure device B.

```
Device B# config-system terminal
Device B(cfg)# interface GigabitEthernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface GigabitEthernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# ip address 192.168.4.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/2)# exit
Device B(cfg)# router-protocol bgp 65536
Device B(cfg-router)# peer 192.168.2.5 remote-as 65536
Device B(cfg-router)# peer 192.168.4.4 remote-as 65535
```

Configure device C.

```
Device C# config-system terminal
Device C(cfg)# interface GigabitEthernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# ip address 192.168.3.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface GigabitEthernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# ip address 192.168.5.3 255.255.255.0
Device C(cfg-if-GigabitEthernet 0/2)# exit
Device C(cfg)# router-protocol bgp 65536
Device C(cfg-router)# peer 192.168.3.5 remote-as 65536
Device C(cfg-router)# peer 192.168.5.4 remote-as 65535
```

Configure device D.

```
Device D# config-system terminal
Device D(cfg)# interface loopback 0
Device D(cfg-if-Loopback 0)# ip address 10.1.1.4 255.255.255.255
Device D(cfg-if-Loopback 0)# exit
Device D(cfg)# interface GigabitEthernet 0/1
Device D(cfg-if-GigabitEthernet 0/1)# ip address 192.168.4.4 255.255.255.0
Device D(cfg-if-GigabitEthernet 0/1)# exit
Device D(cfg)# interface GigabitEthernet 0/1
Device D(cfg-if-GigabitEthernet 0/1)# ip address 192.168.5.4 255.255.255.0
Device D(cfg-if-GigabitEthernet 0/1)# exit
Device D(cfg)# router-protocol bgp 65535
```

```
Device D(cfg-router)# peer 192.168.4.2 remote-as 65536
Device D(cfg-router)# peer 192.168.5.3 remote-as 65536
Device D(cfg-router)# network-advertise 10.1.1.4 mask 255.255.255.255
```

Configure device RR.

```
Device RR# config-system terminal
Device RR(cfg)# interface loopback 0
Device RR(cfg-if-Loopback 0)# ip address 10.1.1.5 255.255.255.255
Device RR(cfg-if-Loopback 0)# exit
Device RR(cfg)# interface GigabitEthernet 0/1
Device RR(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.5 255.255.255.0
Device RR(cfg-if-GigabitEthernet 0/1)# exit
Device RR(cfg)# interface GigabitEthernet 0/2
Device RR(cfg-if-GigabitEthernet 0/2)# ip address 192.168.2.5 255.255.255.0
Device RR(cfg-if-GigabitEthernet 0/2)# exit
Device RR(cfg)# interface GigabitEthernet 0/3
Device RR(cfg-if-GigabitEthernet 0/3)# ip address 192.168.3.5 255.255.255.0
Device RR(cfg-if-GigabitEthernet 0/3)# exit
Device RR(cfg)#ip route-config 192.168.4.0 255.255.255.0 192.168.2.2
Device RR(cfg)#ip route-config 192.168.5.0 255.255.255.0 192.168.3.3
Device RR(cfg)# router-protocol bgp 65536
Device RR(cfg-router)# peer 192.168.1.1 remote-as 65536
Device RR(cfg-router)# peer 192.168.1.1 route-reflector-client
Device RR(cfg-router)# peer 192.168.2.2 remote-as 65536
Device RR(cfg-router)# peer 192.168.2.2 route-reflector-client
Device RR(cfg-router)# peer 192.168.3.3 remote-as 65536
```

(2) Configure the ADD-PATH rules for the route reflector RR.

Configure device RR.

```
Device RR# config-system terminal
Device RR(cfg)# router-protocol bgp 65536
Device RR(cfg-router)# bgp-instance alternative-paths select best 2
Device RR(cfg-router)# peer 192.168.1.1 advertise alternative-paths best 2
Device RR(cfg-router)# peer 192.168.1.1 alternative-paths send
```

5. Verification

On device RR, check the BGP neighbor status.

```
Device RR# display ip bgp-instance
BGP table version is 3, local router ID is 10.1.1.5
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                    S Stale, b - backup entry, a additional-path
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network            Next Hop              Metric     LocPrf     Weight Path
* ia10.1.1.4/32      192.168.4.4                0        100          0 65535 i
*>i                   192.168.5.4                0        100          0 65535 i
Total number of prefixes 2
```

On device A, run the **display** command to display the BGP neighbor status.

```
Device A# display ip bgp-instance
BGP table version is 2, local router ID is 192.168.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                    S Stale, b - backup entry, a additional-path
Origin codes: i - IGP, e - EGP, ? - incomplete

    Network          Next Hop            Metric     LocPrf     Weight Path
* i 10.1.1.4/32     192.168.5.4              0      100          0 65535 i
* i                  192.168.4.4              0      100          0 65535 i
Total number of prefixes 2
```

6. Configuration Files

- Device A configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.1.1 255.255.255.0
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.5 remote-as 65536
 address-family-mode ipv4
  peer 192.168.1.5 activate
  peer 192.168.1.5 alternative-paths recv
  exit-address-family-mode
!
```

- Device B configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.2.2 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 ip address 192.168.4.2 255.255.255.0
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
```

```
 peer 192.168.2.5 remote-as 65536
 peer 192.168.4.4 remote-as 65535
 address-family-mode ipv4
  peer 192.168.2.5 activate
  peer 192.168.4.4 activate
  exit-address-family-mode
!
```

- Device C configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.3.3 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 ip address 192.168.5.3 255.255.255.0
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.3.5 remote-as 65536
 peer 192.168.5.4 remote-as 65535
 address-family-mode ipv4
  peer 192.168.3.5 activate
  peer 192.168.5.4 activate
  exit-address-family-mode
!
```

- Device D configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.4.4 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 ip address 192.168.5.4 255.255.255.0
!
interface Loopback 0
 ip address 10.1.1.4 255.255.255.255
!
router-protocol bgp 65535
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
```

```
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.4.2 remote-as 65536
 peer 192.168.5.3 remote-as 65536
 address-family-mode ipv4
  network-advertise 10.1.1.4 mask 255.255.255.255
  peer 192.168.4.2 activate
  peer 192.168.5.3 activate
  exit-address-family-mode
!
```

- Device RR configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 192.168.1.5 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 ip address 192.168.2.5 255.255.255.0
!
interface GigabitEthernet 0/3
 unset switch-port
 ip address 192.168.3.5 255.255.255.0
!
interface Loopback 0
 ip address 10.1.1.5 255.255.255.255
!
router-protocol bgp 65536
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 192.168.1.1 remote-as 65536
 peer 192.168.2.2 remote-as 65536
 peer 192.168.3.3 remote-as 65536
 address-family-mode ipv4
  bgp-instance alternative-paths select best 2
  peer 192.168.1.1 activate
  peer 192.168.1.1 route-reflector-client
  peer 192.168.1.1 alternative-paths send
  peer 192.168.1.1 advertise alternative-paths best 2
  peer 192.168.2.2 activate
  peer 192.168.2.2 route-reflector-client
  peer 192.168.3.3 activate
  exit-address-family-mode
!
```

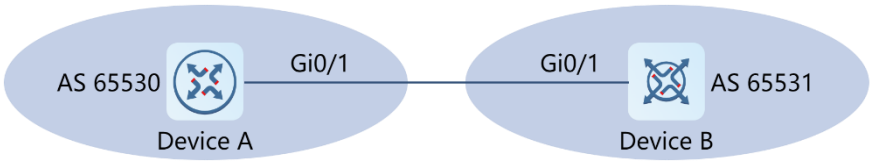
1.13.22 Configuring BGP Neighbor Auto-Discovery

1. Requirements

A BGP session is automatically established over a specific interface.

2. Topology

Figure 1-23 Configuring BGP Neighbor Auto-Discovery



3. Notes

- Enable BGP on all devices.
- Configure peer groups and enable the IPv4 and IPv6 address families on all devices.
- Enable BGP neighbor auto-discovery on all devices.
- Configure the **ipv6 enable** command on the interface between device A and device B.
- Configure BGP neighbor auto-discovery on the interface between device A and device B.

4. Procedure

(1) Configure the BGP peer groups, enable the IPv4 and IPv6 address families, and enable BGP neighbor auto-discovery.

Configure device A.

```
Device A# config-system terminal
Device A(cfg)# router-protocol bgp 65530
Device A(cfg-router)# peer ad peer-group
Device A(cfg-router)# address-family-mode ipv4
Device A(cfg-router-af)# peer ad activate
Device A(cfg-router-af)# exit-address-family-mode
Device A(cfg-router)# address-family-mode ipv6
Device A(cfg-router-af)# peer ad activate
Device A(cfg-router-af)# exit-address-family-mode
Device A(cfg-router)# auto-discovery
Device A(cfg-router-ad)# exit-auto-discovery
Device A(cfg-router)# exit
```

Configure device B.

```
Device B# config-system terminal
Device B(cfg)# router-protocol bgp 65531
Device B(cfg-router)# peer ad peer-group
Device B(cfg-router)# address-family-mode ipv4
Device B(cfg-router-af)# peer ad activate
Device B(cfg-router-af)# exit-address-family-mode
Device B(cfg-router)# address-family-mode ipv6
```

```
Device B(cfg-router-af)# peer ad activate
Device B(cfg-router-af)# exit-address-family-mode
Device B(cfg-router)# auto-discovery
Device B(cfg-router-ad)# exit-auto-discovery
Device B(cfg-router)# exit
```

(2) Configure on the **ipv6 enable** command and BGP neighbor auto-discovery on the interface.

Configure device A.

```
Device A# config-system terminal
Device A(cfg)# interface GigabitEthernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ipv6 enable
Device A(cfg-if-GigabitEthernet 0/1)# bgp-instance auto-discovery local-as 65530
Device A(cfg-if-GigabitEthernet 0/1)# bgp-instance auto-discovery source-interface source-self peer-group ad
```

Configure device B.

```
Device B# config-system terminal
Device B(cfg)# interface GigabitEthernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ipv6 enable
Device B(cfg-if-GigabitEthernet 0/1)# bgp-instance auto-discovery local-as 65531
Device B(cfg-if-GigabitEthernet 0/1)# bgp-instance auto-discovery source-interface source-self peer-group ad
```

5. Verification

On device A, check the BGP neighbor information.

```
Device A# display bgp-instance all peers


For address family: IPv4 Unicast
BGP peer is fe80::a00:27ff:fe50:d4b0%Gi0/1, remote AS 65531, local AS 65530, external link
  MyAs 65531
  Peer ID 1
  Memmory Address 0x7fcdc33d8000 (lock:1)
  Index 1
  cap 0XC7, flags 0X1000A400, sflags 68719517698, config 0X0
  Gstatus 0XE, ha sflags 0X0
 From auto-discovery for session arguments
  Use graceful restart time 120
 Member of peer-group ad (0x7fcdc3346bc0) for session arguments
  BGP version 4, remote router ID 2.2.2.2
    local router ID 1.1.1.1
    saved remote router ID 2.2.2.2
```

```
  BGP state = Established, up for 00:01:12
  Last read 00:00:42, hold time is 180, keepalive interval is 60 seconds
  TCP unackeq seq 4223169434
  ha sync end check flags synchronizing
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
    Address family IPv6 Unicast: advertised and received
    Graceful Restart Capability: advertised and received
      Remote Restart timer is 120 seconds
      Address families preserved by peer:
        IPv4 Unicast (was preserved)
        IPv6 Unicast (was preserved)
    Extended Next Hop Encoding capability
      IPv4 Unicast: advertise IPv6 and received IPv6
  Received 8 messages, 0 notifications, 0 in queue
    open message:1 update message:4 keepalive message:1
    refresh message:2 dynamic cap:0 notifications:0
  Sent 7 messages, 0 notifications, 0 in queue
    open message:1 update message:4 keepalive message:0
    refresh message:2 dynamic cap:0 notifications:0
  Route refresh request: received 2, sent 2
  PIC information valid change message: received 0, send 0
  Minimum time between advertisement runs is 30 seconds
  Minimum time between as origination advertisement runs is 1 seconds
 For address family: IPv4 Unicast
  BGP table version 1, peer version 1
  Index 1, Offset 0, Mask 0x2
  ad peer-group member
  17 update-group member
  0 accepted prefixes
  0 announced prefixes in update-group 17
  rib is empty
  Advertise statistics:
    Normal count: 0
    BoRR count: 0
```

```
    EoRR count: 0
    EoR count: 0
  Send 0 EoRR

 For address family: IPv6 Unicast
  BGP table version 1, peer version 1
  Index 1, Offset 0, Mask 0x2
  ad peer-group member
  18 update-group member
  0 accepted prefixes
  0 announced prefixes in update-group 18
  rib is empty
  Advertise statistics:
    Normal count: 0
    BoRR count: 0
    EoRR count: 0
    EoR count: 0
  Send 0 EoRR

 Connections established 1; dropped 0
 Exceeded max size dropped 0
 Treat as withdraw count 0, last withdraw prefix
 Graceful-restart Status:
  Remote restart-time is 120 sec

 BGP peer may be up to 255 hops away.
Local host: fe80::a00:27ff:fecd:d322%Gi0/1, Local port: 40767
Foreign host: fe80::a00:27ff:fe50:d4b0%Gi0/1, Foreign port: 179
Sock_fd: 76
Sock_cb buffer size: 4096
Stream sock cb (master-fd 0) state info:
  ssock-fd:76
  ha-state:1
  ssock-buf-size:4096
  ssock-ha-writesize:0
  ssock-su-local:port 40767 addr fe80::a00:27ff:fecd:d322
  ssock-su-remote:port 179 addr fe80::a00:27ff:fe50:d4b0
```

```
  curr-ibuf-size:0
  curr-obuf-size:0
  ssock-state:Connected
  status-func-code:0
  read-func-code:7
  read-func-arg:19
  ssock-flags:1
  option-flags:0
  traffic-threshold:0
  traffic-count:0
  ssock-read-slave:0
  unread seq: 0
  send_count: 0
  recv_count: 0
  sync-read-buf-data:0
  message count:0
  need_to_read:0
Nexthop interfaces index: 2
Nexthop global: ::
Nexthop local: fe80::a00:27ff:fecd:d322
BGP connection: shared network
Last Reset:              , due to BGP Notification received
Notification Error Message: (Cease/Other Configuration Change.)



For address family: IPv6 Unicast
BGP peer is fe80::a00:27ff:fe50:d4b0%Gi0/1, remote AS 65531, local AS 65530, external link
  MyAs 65531
  Peer ID 1
  Memmory Address 0x7fcdc33d8000 (lock:1)
  Index 1
  cap 0XC7, flags 0X1000A400, sflags 68719517698, config 0X0
  Gstatus 0XE, ha sflags 0X0
 From auto-discovery for session arguments
  Use graceful restart time 120
 Member of peer-group ad (0x7fcdc3346bc0) for session arguments
  BGP version 4, remote router ID 2.2.2.2
```

```
    local router ID 1.1.1.1
    saved remote router ID 2.2.2.2
  BGP state = Established, up for 00:01:12
  Last read 00:00:42, hold time is 180, keepalive interval is 60 seconds
  TCP unackeq seq 4223169434
  ha sync end check flags synchronizing
  Neighbor capabilities:
    Route refresh: advertised and received (old and new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
    Address family IPv6 Unicast: advertised and received
    Graceful Restart Capability: advertised and received
      Remote Restart timer is 120 seconds
      Address families preserved by peer:
        IPv4 Unicast (was preserved)
        IPv6 Unicast (was preserved)
    Extended Next Hop Encoding capability
      IPv4 Unicast: advertise IPv6 and received IPv6
  Received 8 messages, 0 notifications, 0 in queue
    open message:1 update message:4 keepalive message:1
    refresh message:2 dynamic cap:0 notifications:0
  Sent 7 messages, 0 notifications, 0 in queue
    open message:1 update message:4 keepalive message:0
    refresh message:2 dynamic cap:0 notifications:0
  Route refresh request: received 2, sent 2
  PIC information valid change message: received 0, send 0
  Minimum time between advertisement runs is 30 seconds
  Minimum time between as origination advertisement runs is 1 seconds
 For address family: IPv4 Unicast
  BGP table version 1, peer version 1
  Index 1, Offset 0, Mask 0x2
  ad peer-group member
  17 update-group member
  0 accepted prefixes
  0 announced prefixes in update-group 17
  rib is empty
  Advertise statistics:
```

```
    Normal count: 0
    BoRR count: 0
    EoRR count: 0
    EoR count: 0
  Send 0 EoRR

 For address family: IPv6 Unicast
  BGP table version 1, peer version 1
  Index 1, Offset 0, Mask 0x2
  ad peer-group member
  18 update-group member
  0 accepted prefixes
  0 announced prefixes in update-group 18
  rib is empty
  Advertise statistics:
    Normal count: 0
    BoRR count: 0
    EoRR count: 0
    EoR count: 0
  Send 0 EoRR

 Connections established 1; dropped 0
 Exceeded max size dropped 0
 Treat as withdraw count 0, last withdraw prefix
 Graceful-restart Status:
  Remote restart-time is 120 sec

 BGP peer may be up to 255 hops away.
Local host: fe80::a00:27ff:fecd:d322%Gi0/1, Local port: 40767
Foreign host: fe80::a00:27ff:fe50:d4b0%Gi0/1, Foreign port: 179
Sock_fd: 76
Sock_cb buffer size: 4096
Stream sock cb (master-fd 0) state info:
  ssock-fd:76
  ha-state:1
  ssock-buf-size:4096
  ssock-ha-writesize:0
```

```
  ssock-su-local:port 40767 addr fe80::a00:27ff:fecd:d322
  ssock-su-remote:port 179 addr fe80::a00:27ff:fe50:d4b0
  curr-ibuf-size:0
  curr-obuf-size:0
  ssock-state:Connected
  status-func-code:0
  read-func-code:7
  read-func-arg:19
  ssock-flags:1
  option-flags:0
  traffic-threshold:0
  traffic-count:0
  ssock-read-slave:0
  unread seq: 0
  send_count: 0
  recv_count: 0
  sync-read-buf-data:0
  message count:0
  need_to_read:0
Nexthop interfaces index: 2
Nexthop global: ::
Nexthop local: fe80::a00:27ff:fecd:d322
BGP connection: shared network
Last Reset:              , due to BGP Notification received
Notification Error Message: (Cease/Other Configuration Change.)
```

6. Configuration Files

- Device A configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 bgp-instance auto-discovery local-as 65530
 bgp-instance auto-discovery source-interface source-self peer-group ad
 ipv6 enable
!
router-protocol bgp 65530
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
```

```
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer ad peer-group
 auto-discovery
  exit-auto-discovery
 address-family-mode ipv4
  peer ad activate
  exit-address-family-mode
 address-family-mode ipv6
  peer ad activate
  exit-address-family-mode
!
```

- Device B configuration file

```
!
interface GigabitEthernet 0/1
 unset switch-port
 bgp-instance auto-discovery local-as 65531
 bgp-instance auto-discovery source-interface source-self peer-group ad
 ipv6 enable
!
router-protocol bgp 65531
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer ad peer-group
 auto-discovery
  exit-auto-discovery
 address-family-mode ipv4
  peer ad activate
  exit-address-family-mode
 address-family-mode ipv6
  peer ad activate
  exit-address-family-mode
!
```

Contents

1 Configuring VRF....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.1.3 Protocols and Standards....3
1.2 Restrictions and Guidelines....3
1.3 Configuration Task Summary....3
1.4 Configuring a Single-protocol VRF....4
1.4.1 Overview....4
1.4.2 Restrictions and Guidelines....4
1.4.3 Configuration Tasks....5
1.4.4 Creating a Single-protocol VRF....5
1.4.5 Configuring the RD for a Single-protocol VRF....5
1.4.6 Configuring the RT for a Single-protocol VRF....6
1.4.7 Configuring the Label Allocation Mode for a Single-protocol VRF....6
1.4.8 Configuring an Export Policy Associated with a Single-protocol VRF....7
1.4.9 Configuring an Import Policy Associated with a Single-protocol VRF....8
1.4.10 Adding an Interface to a Single-protocol VRF....8
1.4.11 Importing Routes of an Interface to a Single-protocol VRF....9
1.4.12 Configuring Maximum Number of Routes for a Single-protocol VRF....10
1.5 Configuring a Multi-protocol VRF....11
1.5.1 Overview....11

1.5.2 Restrictions and Guidelines 11
1.5.3 Configuration Tasks 11
1.5.4 Creating a Multi-protocol VRF 11
1.5.5 Configuring the Address Family for a Multi-protocol VRF 12
1.5.6 Configuring the RD for a Multi-protocol VRF 12
1.5.7 Configuring the RT for a Multi-protocol VRF 13
1.5.8 Configuring the Label Allocation Mode for a Multi-protocol VRF 13
1.5.9 Configuring an Export Policy Associated with a Multi-protocol VRF 14
1.5.10 Configuring an Import Policy Associated with a Multi-protocol VRF 15
1.5.11 Adding an Interface to a Multi-protocol VRF 15
1.5.12 Importing Routes of an Interface to a Multi-protocol VRF 17
1.5.13 Configuring Maximum Number of Routes for a Multi-protocol VRF 18
1.6 VRF Configuring a Global VRF 18
1.6.1 Overview 18
1.6.2 Restrictions and Guidelines 18
1.6.3 Configuration Tasks 18
1.6.4 Configuring the Address Family for a Global VRF 19
1.6.5 Configuring the RD for a Global VRF 19
1.6.6 Configuring the RT for a Global VRF 20
1.6.7 Configuring the Label Allocation Mode for a Global VRF 20
1.6.8 Configuring an Export Policy Associated with a Global VRF 21
1.6.9 Configuring an Import Policy Associated with a Global VRF 21
1.6.10 Configuring Maximum Number of Routes for a Global VRF 22
1.7 Monitoring 22

1.8 Configuration Examples.....23

1.8.1 Enabling Local Inter-VPN Access.....23

1 Configuring VRF

1.1 Introduction

1.1.1 Overview

VPN Routing and Forwarding Table (VRF) is used to resolve local routing conflicts caused by the same IP address space between different Virtual Private Networks (VPNs). One VPN corresponds to one VRF. Each VRF is equivalent to a "virtual router" that receives or advertises VPN routing information. Each VRF includes:

- An independent routing table;
- A set of interfaces belonging to this VRF;
- A set of routing protocols used for this VRF only.

Multiple routing tables on the device that provides the VPN service including a public network routing table and one or more VRFs. The public network routing table is used to forward public network packets, and the VRF is used to forward VPN packets. Routes in the public network routing table are separated from those in VRFs, and routes are separated between VRFs.

Note

- A VPN is a private dedicated network built in the public network. "Virtual" means that the VPN is logically exclusive, but not physically exclusive.
- The private dedicated network is called “private network” for short.

1.1.2 Principles

1. VPN Instance

A VPN instance is used to provide the VPN service. On a device that provides the VPN service, a VPN instance includes the VRFs, interfaces, routing protocol processes, and configurations that belong to the same VPN. Generally, the VRF represents a VPN instance.

2. VPN Route

A VPN route is only used to forward VPN packets. It comes from:

- Direct route and host route generated after an interface is added to the VRF;
- Direct route and host route imported from an interface that is not added to the VRF;
- Static and dynamic routes (RIP, RIPng, OSPFv2, OSPFv3, ISIS, and BGP) in the configured VPN instance.

Note

- For configuration of the static route in a VPN instance, see *Configuring IP Route Basics*.
- For configuration of RIP in a VPN instance, see *Configuring RIP*.
- For configuration of RIPng in a VPN instance, see *Configuring RIPng*.
- For configuration of OSPFv2 in a VPN instance, see *Configuring OSPFv2*.
- For configuration of OSPFv3 in a VPN instance, see *Configuring OSPFv3*.
- For configuration of ISIS in a VPN instance, see *Configuring ISIS*.

- For configuration of BGP in a VPN instance, see *Configuring BGP*.

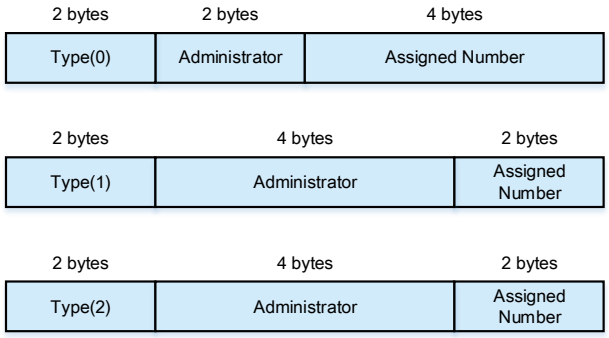
3. VPN Route Attribute

A VPN route has two attributes: Route Distinguisher (RD) and Route Target (RT).

- The RD is used to resolve routing conflicts when the Border Gateway Protocol (BGP) advertises routes. If different VPNs have the same network addresses, the BGP only selects an optimum route from these addresses and advertises this route. As a result, some VPNs cannot receive the routing information. To resolve this problem, you can add different RDs to the same addresses. Based on the different RDs carried in VPN information, the BGP can distinguish the same network addresses, and therefore each VPN can receive its own routing information.

  Generally, a unique RD value is specified for each VPN. A VPN route carries only one RD value. An RD consists of a **Type** field, an **Administrator** field, and an **Assigned Number** field. As shown in Figure 1-1, based on the value of the **Type** field, the RD is encoded in three formats:

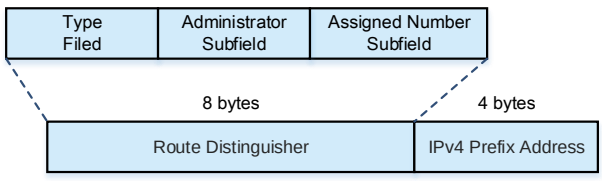
Figure 1-1 RD structure



  - When **Type** is set to **0**, the **Administrator** field has two bytes and is set to an Autonomous System (AS) number, which must be a public AS number; the **Assigned Number** field has four bytes and is managed by the service provider (SP).
  - When **Type** is set to **1**, the **Administrator** field has four bytes and is set to an IPv4, which must be a global IP address; the **Assigned Number** field has two bytes and is managed by the SP.
  - When **Type** is set to **2**, the **Administrator** field has four bytes and is set to a 4-byte AS number; the **Assigned Number** field has two bytes and is managed by the SP.

  An IPv4 address with the RD is called VPN-IPv4 address, as shown in Figure 1-2. After receiving an IPv4 route, the BGP converts it into a globally unique VPN-IPv4 route, and advertises this route in the public network.

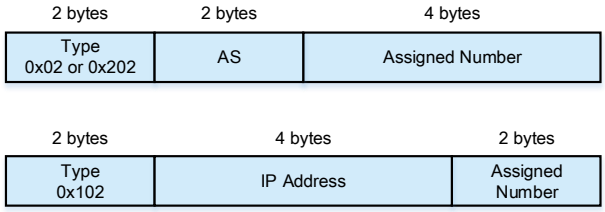
Figure 1-2 VPN-IPv4 address structure



- RT is used to control the advertising and installation policies of VPN routes. The RT attribute falls into the following two types:
  - **Import**: When a device receives the advertised VPN-IPv4 route, the device checks the **Export** attribute of this route. If the **Import** attribute of a specified VPN instance on the device matches the **Export** attribute, the device adds the route to the VPN instance.
  - **Export**: After the local device learns an IPv4 route, it converts this route into a VPN-IPv4 route. The **Export** attribute is set for this route and advertised as the BGP extended community attribute to other devices together with the route.

  The BGP extended community attribute defines the RT structure, as shown in Figure 1-3.

Figure 1-3 RT structure



The RT value and the RD value are similarly defined. When **Type** is set to **0x02** or **0x202**, the AS number must be public. When **Type** is set to **0x102**, the IPv4 address must be a global IP address.

1.1.3 Protocols and Standards

- RFC 4364: BGP/MPLS IP Virtual Private Networks (VPNs)

1.2 Restrictions and Guidelines

- You can configure only one type of VRF: single-protocol VRF or Multi-protocol VRF.
- Single-protocol VRFs only support IPv4, whereas Multi-protocol VRFs support both IPv4 and IPv6.

1.3 Configuration Task Summary

The configuration tasks below are mutually exclusive. Please configure only one task.

- Configuring a Single-protocol VRF
  - Creating a Single-protocol VRF
  - (Optional)Configuring the RD for a Single-protocol VRF
  - (Optional)Configuring the RT for a Single-protocol VRF
  - (Optional)Configuring the Label Allocation Mode for a Single-protocol VRF
  - (Optional)Configuring an Export Policy Associated with a Single-protocol VRF
  - (Optional)Configuring an Import Policy Associated with a Single-protocol VRF
  - Adding an Interface to a Single-protocol VRF
  - (Optional)Importing Routes of an Interface to a Single-protocol VRF

- (Optional)Configuring Maximum Number of Routes for a Single-protocol VRF

- Configuring a Multi-protocol VRF
  - Creating a Multi-protocol VRF
  - (Optional) Configuring the Address Family for a Multi-protocol VRF
  - (Optional)Configuring the RD for a Multi-protocol VRF
  - (Optional)Configuring the RT for a Multi-protocol VRF
  - (Optional)Configuring the Label Allocation Mode for a Multi-protocol VRF
  - (Optional)Configuring an Export Policy Associated with a Multi-protocol VRF
  - (Optional)Configuring an Import Policy Associated with a Multi-protocol VRF
  - Adding an Interface to a Multi-protocol VRF
  - (Optional)Importing Routes of an Interface to a Multi-protocol VRF
  - (Optional)Configuring Maximum Number of Routes for a Multi-protocol VRF
- VRF Configuring a Global VRF
  - Configuring the Address Family for a Global VRF
  - (Optional)Configuring the RD for a Global VRF
  - (Optional)Configuring the RT for a Global VRF
  - (Optional)Configuring the Label Allocation Mode for a Global VRF
  - (Optional)Configuring an Export Policy Associated with a Global VRF
  - (Optional)Configuring an Import Policy Associated with a Global VRF
  - (Optional)Configuring Maximum Number of Routes for a Global VRF

1.4 Configuring a Single-protocol VRF

1.4.1 Overview

A single-protocol VRF is configured to manage VPN routes.

1.4.2 Restrictions and Guidelines

- If a device forwards only packets of a specified VPN or only public network packets, you do not need to configure a single-protocol VRF.
- If the device needs to forward both the public network packets and VPN packets or forward packets from multiple VPNs, VRFs must be created to separate routes.
- You need to add static or dynamic routes (RIP, OSPF, ISIS, and BGP) to a VRF.
- Only IPv4 is supported.
- When a Layer 3 interface joins or exits a single-protocol VRF, the IPv4 address, related routing protocols and Layer 3 attributes of the interface will be deleted as well.
- If a Layer 3 interface has been added to a single-protocol VRF, the IPv4 address, related routing protocols, and Layer 3 attributes of the interface will be deleted when the VRF is deleted.

1.4.3 Configuration Tasks

The configuration of a single-protocol VRF includes the following tasks:

- Creating a Single-protocol VRF
- (Optional)Configuring the RD for a Single-protocol VRF
- (Optional)Configuring the RT for a Single-protocol VRF
- (Optional)Configuring the Label Allocation Mode for a Single-protocol VRF
- (Optional)Configuring an Export Policy Associated with a Single-protocol VRF
- (Optional)Configuring an Import Policy Associated with a Single-protocol VRF
- Adding an Interface to a Single-protocol VRF
- (Optional)Importing Routes of an Interface to a Single-protocol VRF
- (Optional)Configuring Maximum Number of Routes for a Single-protocol VRF

1.4.4 Creating a Single-protocol VRF

1. Overview

This section describes how to create a single-protocol VRF for each VPN.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a single-protocol VRF, and enter the VRF configuration mode.

ip vrf *vrf-name*

No single-protocol VRF is configured by default.

(4) (Optional) Configure a description for single-protocol VRF.

description *description*

No description is configured for single-protocol VRF by default.

1.4.5 Configuring the RD for a Single-protocol VRF

1. Overview

This section describes how to configure the RD to resolve routing conflicts when the BGP advertises routes.

2. Restrictions and Guidelines

- Configure the RD command in single-protocol VRF configuration mode.
- If you have configured a VRF value and its RD value and want to modify the RD value, you must first delete the VRF, reconfigure the VRF, and then configure the RD value.
- You can configure only one RD value for a VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the single-protocol VRF configuration mode.

ip vrf *vrf-name*

(4) Configure the RD.

rd *rd-value*

No RD value is configured by default.

1.4.6 Configuring the RT for a Single-protocol VRF

1. Overview

This section describes how to configure the RT value to control advertisement of VPN routes.

2. Restrictions and Guidelines

- Configure the RT value in single-protocol VRF configuration mode.
- You can configure multiple **Import** or **Export** values for each VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the single-protocol VRF configuration mode.

ip vrf *vrf-name*

(4) Configure the RT.

route-target { **both** | **export** | **import** } *route-value*

No RT is configured by default.

1.4.7 Configuring the Label Allocation Mode for a Single-protocol VRF

1. Overview

This section describes how to configure the label allocation mode for a single-protocol VRF.

2. Restrictions and Guidelines

- Route-based label allocation mode: One route corresponds to one label. The advantage is that the system can look up the ILM table based on the label value and quickly forward packets to the next hop. The disadvantage is that the capacity of the ILM table must be large.
- VRF-based label allocation mode: One VRF corresponds to one label. The advantage is that all routes in the VRF share this label, which reduces the capacity of the ILM table. The disadvantage is that packet

forwarding is slower than that in route-based label allocation mode. This is because the system needs to look up the table twice in the forwarding process. It first looks up the ILM table to find the VRF where the packet is located, and then forwards the packet based on the destination IP address in the routing table of the VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the single-protocol VRF configuration mode.

ip vrf *vrf-name*

(4) Configure the label allocation mode.

alloc-label { **per-route** | **per-vrf** }

No label allocation mode is configured by default.

1.4.8 Configuring an Export Policy Associated with a Single-protocol VRF

1. Overview

This section describes how to configure a policy for exporting the extended community attributes of VPN routes from the local VRF to a remote device.

2. Restrictions and Guidelines

- To more precisely control the extended community attributes of exported routes, you can define a precise rule as needed in the associated route map.
- You can run this command to add or modify the extended community attributes defined by the **route-target export** command.
- You can only run the **match ip address** and **set extcommunity** commands to configure rules for a route map associated by this command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the single-protocol VRF configuration mode.

ip vrf *vrf-name*

(4) Configure the associated export policy.

export map *route-map-name*

No export policy associated with the VRF is configured by default.

1.4.9 Configuring an Import Policy Associated with a Single-protocol VRF

1. Overview

This section describes how to configure a policy for importing VPN routes from a remote device to the local VRF.

2. Restrictions and Guidelines

- To more precisely filter routes imported to the local VRF, you can define a precise rule as needed in the associated route map.
- The rule defined by the **import map** command takes effect after the **Import** extended community attributes defined in the VRF. That is, only after the route matches the extended community attributes defined by the **route-target import** command in the VRF, can the VPN routes received from the remote device be filtered again by rules defined by the **import map** command.
- You can only run the **match ip address** and **match extcommunity** commands to configure rules for a route map associated by the **import map** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the single-protocol VRF configuration mode.

ip vrf *vrf-name*

(4) Configure the associated import policy.

import map *route-map-name*

No import policy associated with the VRF is configured by default.

1.4.10 Adding an Interface to a Single-protocol VRF

1. Overview

This section describes how to add an interface to a specified VRF. This VRF is used to forward incoming packets of the interface.

2. Restrictions and Guidelines

You must add the interface to the single-protocol VRF before configuring the IPv4 address. If you configure the IPv4 address for an interface and then add the interface to the single-protocol VRF, the configuration of the IPv4 address fails, but the configuration of the IPv6 address is retained.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Add the interface to the single-protocol VRF.

ip vrf binding *vrf-name*

By default, an interface does not belong to any VRF.

1.4.11 Importing Routes of an Interface to a Single-protocol VRF

1. Overview

This section describes how to import the host route and direct route of an interface to a specified VRF.

2. Restrictions and Guidelines

- If you need to use policy-based routing (PBR) to select the VRF, configure the **ip vrf receive** command on the interface to which PBR is applied, and import the direct route and host route on the interface to each VRF available for choice.
- If you need to import the host route and direct route of an interface to multiple VRFs, run this command multiple times.
- The **ip vrf binding** and **ip vrf receive** commands are mutually exclusive on a specified interface. Unlike the **ip vrf binding** command, the **ip vrf receive** command does not add the interface to the VRF, and the interface does not belong to any VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Import the host route and direct route of the interface to a specified VRF.

ip vrf receive *vrf-name*

The direct route and host route of an interface are not imported to any specified VRF by default.

1.4.12 Configuring Maximum Number of Routes for a Single-protocol VRF

1. Overview

You can limit the maximum number of routing entries in a single-protocol VRF by performing this configuration.

2. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the single-protocol VRF configuration mode.
ip vrf *vrf-name*

(4) Configure the maximum number of routes in a single-protocol VRF.
maximum routes-number *limit* { **warning** | *warning-threshold* }

The maximum number of routes in the single-protocol VRF is not limited by default.

1.5 Configuring a Multi-protocol VRF

1.5.1 Overview

This section describes how to configure a Multi-protocol VRF to manage VPN routes. You can configure the address family to support IPv4 and IPv6. The BGP helps you to flexibly isolate VPNs or allow inter-VPN access, or provide the VPN service in an IP/MPLS backbone network.

1.5.2 Restrictions and Guidelines

- If a device forwards only packets of a specified VPN or only public network packets, you do not need to configure a Multi-protocol VRF.
- If the device needs to forward both the public network packets and VPN packets or forward packets from multiple VPNs, VRFs must be created to separate routes.
- You need to add static or dynamic routes (RIP, OSPF, ISIS, and BGP) to a VRF.
- When a Layer 3 interface joins or exits a multi-protocol VRF, the IPv4 address, related routing protocols and Layer 3 attributes of the interface will be deleted as well.
- If a Layer 3 interface has been added to a multi-protocol VRF, the IPv4 address, related routing protocols, and Layer 3 attributes of the interface will be deleted when the VRF is deleted.

1.5.3 Configuration Tasks

The configuration of a Multi-protocol VRF includes the following tasks:

- Creating a Multi-protocol VRF
- (Optional) Configuring the Address Family for a Multi-protocol VRF
- (Optional)Configuring the RD for a Multi-protocol VRF
- (Optional)Configuring the RT for a Multi-protocol VRF
- (Optional)Configuring the Label Allocation Mode for a Multi-protocol VRF
- (Optional)Configuring an Export Policy Associated with a Multi-protocol VRF
- (Optional)Configuring an Import Policy Associated with a Multi-protocol VRF
- Adding an Interface to a Multi-protocol VRF
- (Optional)Importing Routes of an Interface to a Multi-protocol VRF
- (Optional)Configuring Maximum Number of Routes for a Multi-protocol VRF

1.5.4 Creating a Multi-protocol VRF

1. Overview

This section describes how to create a Multi-protocol VRF for each VPN.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a Multi-protocol VRF, and enter the Multi-protocol VRF configuration mode.

vrf definition *vrf-name*

No Multi-protocol VRF exists by default.

(4) (Optional) Configure a description for multi-protocol VRF.

description *description*

No description is configured for multi-protocol VRF by default.

1.5.5 Configuring the Address Family for a Multi-protocol VRF

1. Overview

This section describes how to enable the IPv4/IPv6 address family for a Multi-protocol VRF, and enter the configuration mode.

2. Restrictions and Guidelines

You can configure the address family for each created VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Multi-protocol VRF configuration mode.

vrf definition *vrf-name*

(4) Configure the address family. Configure at least one of the following tasks.

- Configure the IPv4 address family, and enter the VRF IPv4 address family submode.

  address-family ipv4

- Configure the IPv6 address family, and enter the VRF IPv6 address family submode.

  address-family ipv6

No address family is configured for a Multi-protocol VRF by default.

1.5.6 Configuring the RD for a Multi-protocol VRF

1. Overview

This section describes how to configure the RD to resolve routing conflicts when the BGP advertises routes.

2. Restrictions and Guidelines

- Configure the RD command in Multi-protocol VRF configuration mode.
- If you have configured a VRF value and its RD value and want to modify the RD value, you must first delete the VRF, reconfigure the VRF, and then configure the RD value.
- You can configure only one RD value for a VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Multi-protocol VRF configuration mode.

vrf definition *vrf-name*

(4) Configure the RD.

rd *rd-value*

No RD value is configured by default.

1.5.7 Configuring the RT for a Multi-protocol VRF

1. Overview

This section describes how to configure the RT group to control advertisement of VPN routes.

2. Restrictions and Guidelines

- Configure the RT value in Multi-protocol VRF mode or Multi-protocol VRF address family mode.
- You can configure multiple **Import** or **Export** values for each VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Multi-protocol VRF configuration mode.

vrf definition *vrf-name*

- (Optional) Enter the Multi-Protocol VRF address family configuration mode.

  address-family { **ipv4** | **ipv6** }

(4) Configure the RT.

- **route-target** { **both** | **export** | **import** } *route-value*

By default, the RT is not configured.

1.5.8 Configuring the Label Allocation Mode for a Multi-protocol VRF

1. Overview

This section describes how to configure the label allocation mode for a Multi-protocol VRF.

2. Restrictions and Guidelines

- Route-based label allocation mode: One route corresponds to one label. The advantage is that the system can look up the ILM table based on the label value and quickly forward packets to the next hop. The

disadvantage is that the capacity of the ILM table must be large.

- VRF-based label allocation mode: One VRF corresponds to one label. The advantage is that all routes in the VRF share this label, which reduces the capacity of the ILM table. The disadvantage is that packet forwarding is slower than that in route-based label allocation mode. This is because the system needs to look up the table twice in the forwarding process. It first looks up the ILM table to find the VRF where the packet is located, and then forwards the packet based on the destination IP address in the routing table of the VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the multi-protocol VRF configuration mode.

vrf definition *vrf-name*

- Enter the Multi-Protocol VRF address family configuration mode.

  address-family { **ipv4** | **ipv6** }

(4) Configure the label allocation mode.

alloc-label { **per-route** | **per-vrf** }

By default, the label allocation mode is not configured.

1.5.9 Configuring an Export Policy Associated with a Multi-protocol VRF

1. Overview

This section describes how to configure a policy for exporting the extended community attributes of VPN routes from the local VRF to the remote device

2. Restrictions and Guidelines

- To more precisely control the extended community attributes of exported routes, you can define a precise rule as needed in the associated route map.
- You can run this command to add or modify the extended community attributes defined by the **route-target export** command.
- You can only run the **match ip address** and **set extcommunity** commands to configure rules for a route map associated by this command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Multi-protocol VRF configuration mode.

vrf definition *vrf-name*

- Enter the Multi-protocol VRF address family configuration mode.
  - **address-family { ipv4 | ipv6 }**

(4) Configure the associated export policy.

export map *route-map-name*

No export policy associated with the VRF is configured by default.

1.5.10 Configuring an Import Policy Associated with a Multi-protocol VRF

1. Overview

This section describes how to configure a policy for importing VPN routes from a remote device to the local VRF.

2. Restrictions and Guidelines

- To more precisely filter routes imported to the local VRF, you can define a precise rule as needed in the associated route map.
- The rule defined by the **import map** command takes effect after the **Import** extended community attributes defined in the VRF. That is, only after these routes match the extended community attributes defined by the **route-target import** command in the VRF, can the VPN routes received from the remote device be filtered again by rules defined the **import map** command.
- You can only run the **match ip address** and **match extcommunity** commands to configure rules for a route map associated by the import map command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Multi-protocol VRF configuration mode.

vrf definition *vrf-name*

- Enter the Multi-Protocol VRF address family configuration mode.

  address-family { **ipv4** | **ipv6** }

(4) Configure the associated import policy.

import map *route-map-name*

No import policy associated with the VRF is configured by default.

1.5.11 Adding an Interface to a Multi-protocol VRF

1. Overview

This section describes how to add an interface to a specified VRF. This VRF is used to forward incoming packets of the interface.

2. Restrictions and Guidelines

- After this command is executed, the direct route and host route on the interface are automatically imported

to the VRF.

- Before you add an interface to a Multi-protocol VRF, enable the address family for the table. If the IPv4 address family is disabled when you add an interface to a Multi-protocol VRF, you are not allowed to configure the IPv4 and VRRP IPv4 addresses on this interface. If the IPv6 address family is disabled when you add an interface to a Multi-protocol VRF, you are not allowed to configure the IPv6 and VRRP IPv6 addresses on this interface.
- You must add an interface to the Multi-protocol VRF before configuring the IPv4/IPv6 or VRRP IPv4/IPv6 addresses on the interface. When you add an interface to a Multi-protocol VRF, existing IPv4/IPv6 or VRRP IPv4/IPv6 addresses on the interface are deleted, and the IPv6 protocol is disabled on the interface.
- If the IPv4 address family is deleted from the Multi-protocol VRF, the IPv4 and VRRP IPv4 addresses of all interfaces added to this VRF are deleted, and the IPv4 static routes in the VRF or the IPv4 static routes with this VRF as their next-hop VRF are also deleted.
- If the IPv6 address family is deleted from the Multi-protocol VRF, the IPv6 and VRRP IPv6 addresses of all interfaces added to the VRF are deleted, the IPv6 protocol is disabled on the interface, and the IPv6 static routes in the VRF or the IPv6 static routes with this VRF as their next-hop VRF are also deleted.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.
  - **interface overlayrouter** *interface-number*

(4) Add the interface to the Multi-protocol VRF.

vrf binding *vrf-name*

By default, an interface does not belong to any VRF.

1.5.12 Importing Routes of an Interface to a Multi-protocol VRF

1. Overview

This section describes how to import the host route and direct route of an interface to a specified VRF.

2. Restrictions and Guidelines

- If you need to use PBR to select the VRF, configure the **vrf receive** command on the interface to which PBR is applied to import the direct route and host route on the interface to each VRF available for choice.
- If you need to import the host route and direct route of an interface to multiple VRFs, run this command multiple times.
- The **vrf binding** and **vrf receive** commands are mutually exclusive on a specified interface. Unlike the **vrf binding** command, the **vrf receive** command does not add an interface to the specified VRF, and the interface does not belong to any VRF.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 aggregate interface configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the Layer 3 Ethernet sub-interface configuration mode.
  - **interface** *ethernet-type interface-number.subnumber*
- Enter the Layer 3 aggregate sub-interface configuration mode.
  - **interface aggregateport** *interface-number.subnumber*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the tunnel interface configuration mode.
  - **interface tunnel** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*
- Enter the VXLAN routed interface configuration mode.

- **interface overlayrouter** *interface-number*

(4) Import the host route and direct route of the interface to a specified VRF.

vrf receive *vrf-name*

By default, the host route and direct route of the interface are not imported to a specified VRF.

1.5.13 Configuring Maximum Number of Routes for a Multi-protocol VRF

1. Overview

You can limit the maximum number of routing entries in a multi-protocol VRF by performing this configuration.

2. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the multi-protocol VRF configuration mode.
vrf definition *vrf-name*

(4) Enter the multi-protocol VRF address family configuration mode.
address-family { **ipv4** | **ipv6** }

(5) Configure the maximum number of routes in a multi-protocol VRF.
maximum routes *limit* { **warning** | *warning-threshold* }

The maximum number of routes in the multi-protocol VRF is not limited by default.

1.6 VRF Configuring a Global VRF

1.6.1 Overview

This section describes how to configure a global VRF to manage VPN routes. Global VRF exists by default.

1.6.2 Restrictions and Guidelines

- If a device forwards only packets of a specified VPN or only public network packets, you do not need to configure a VRF.
- If the device needs to forward both the public network packets and VPN packets or forward packets from multiple VPNs, VRFs must be created to separate routes.
- You need to add static or dynamic routes (RIP, OSPF, ISIS, and BGP) to a VRF.

1.6.3 Configuration Tasks

The configuration of a Multi-protocol VRF includes the following tasks:

- Configuring the Address Family for a Global VRF
- (Optional)Configuring the RD for a Global VRF
- (Optional)Configuring the RT for a Global VRF
- (Optional)Configuring the Label Allocation Mode for a Global VRF
- (Optional)Configuring an Export Policy Associated with a Global VRF
- (Optional)Configuring an Import Policy Associated with a Global VRF
- (Optional)Configuring Maximum Number of Routes for a Global VRF

1.6.4 Configuring the Address Family for a Global VRF

1. Overview

This section describes how to enable the IPv4/IPv6 address family for a Multi-protocol VRF, and enter the configuration mode.

2. Restrictions and Guidelines

You can configure the address family for each created VRF.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the global VRF configuration mode.
vrf global-vrf

(4) (Optional) Configure a description for global VRF.
description *description*

No description is configured for global VRF by default.

(5) Configure the address family. Configure at least one of the following tasks.

- Configure the IPv4 address family, and enter the VRF IPv4 address family submode.
  address-family ipv4
- Configure the IPv6 address family, and enter the VRF IPv6 address family submode.
  address-family ipv6

By default, IPv4 and IPv6 address families are enabled.

1.6.5 Configuring the RD for a Global VRF

1. Overview

This section describes how to configure the RD to resolve routing conflicts when the BGP advertises routes.

2. Restrictions and Guidelines

- Configure the RD command in global VRF configuration mode.
- If you have configured a VRF value and its RD value and want to modify the RD value, you must first delete the VRF, reconfigure the VRF, and then configure the RD value.
- You can configure only one RD value for a VRF.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the global VRF configuration mode.
vrf global-vrf

(4) Configure the RD.
rd *rd-value*

No RD value is configured by default.

1.6.6 Configuring the RT for a Global VRF

1. Overview

This section describes how to configure the RT group to control advertisement of VPN routes.

2. Restrictions and Guidelines

- Configure the RT value in global VRF mode or global VRF address family mode.
- You can configure multiple **Import** or **Export** values for each VRF.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the global VRF configuration mode.
vrf global-vrf

(4) (Optional) Enter the global VRF address family configuration mode.
address-family { **ipv4** | **ipv6** }

(5) Configure the RT.
route-target { **both** | **export** | **import** } *route-value*

By default, the RT is not configured.

1.6.7 Configuring the Label Allocation Mode for a Global VRF

1. Overview

This section describes how to configure the label allocation mode for a global VRF.

2. Restrictions and Guidelines

- Route-based label allocation mode: One route corresponds to one label. The advantage is that the system can look up the ILM table based on the label value and quickly forward packets to the next hop. The disadvantage is that the capacity of the ILM table must be large.
- VRF-based label allocation mode: One VRF corresponds to one label. The advantage is that all routes in the VRF share this label, which reduces the capacity of the ILM table. The disadvantage is that packet forwarding is slower than that in route-based label allocation mode. This is because the system needs to look up the table twice in the forwarding process. It first looks up the ILM table to find the VRF where the packet is located, and then forwards the packet based on the destination IP address in the routing table of the VRF.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the global VRF configuration mode.
vrf global-vrf

(4) Enter the global VRF address family configuration mode.
address-family { **ipv4** | **ipv6** }

(5) Configure the label allocation mode.
alloc-label { **per-route** | **per-vrf** }

By default, the label allocation mode is not configured.

1.6.8 Configuring an Export Policy Associated with a Global VRF

1. Overview

This section describes how to configure a policy for exporting the extended community attributes of VPN routes from the local VRF to the remote device

2. Restrictions and Guidelines

- To more precisely control the extended community attributes of exported routes, you can define a precise rule as needed in the associated route map.
- You can run this command to add or modify the extended community attributes defined by the **route-target export** command.
- You can only run the **match ip address** and **set extcommunity** commands to configure rules for a route map associated by this command.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the global VRF configuration mode.
vrf global-vrf

(4) Enter the global VRF address family configuration mode.
address-family { **ipv4** | **ipv6** }

(5) Configure the associated export policy.
export map *route-map-name*

No export policy associated with the VRF is configured by default.

1.6.9 Configuring an Import Policy Associated with a Global VRF

1. Overview

This section describes how to configure a policy for importing VPN routes from a remote device to the local VRF.

2. Restrictions and Guidelines

- To more precisely filter routes imported to the local VRF, you can define a precise rule as needed in the associated route map.
- The rule defined by the **import map** command takes effect after the import extended community attributes defined in the VRF. That is, only after these routes match the extended community attributes defined by the **route-target import** command in the VRF, can the VPN routes received from the remote device be filtered again by rules defined the **import map** command.
- You can only run the **match ip address** and **match extcommunity** commands to configure rules for a route map associated by the import map command.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the global VRF configuration mode.
vrf global-vrf

(4) Enter the global VRF address family configuration mode.
address-family { **ipv4** | **ipv6** }

(5) Configure the associated import policy.
import map *route-map-name*

No import policy associated with the VRF is configured by default.

1.6.10 Configuring Maximum Number of Routes for a Global VRF

1. Overview

You can limit the maximum number of routing entries in a global VRF by performing this configuration.

2. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the global VRF configuration mode.
vrf global-vrf

(4) Enter the global VRF address family configuration mode.
address-family { **ipv4** | **ipv6** }

(5) Configure the maximum number of routes in a global VRF.
maximum routes *limit* { **warning** | *warning-threshold* }

The maximum number of routes in the global VRF is not limited by default.

1.7 Monitoring

Run the **display** command to check the running status of a configured function and verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may interrupt services due to loss of vital information.

Run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 VRF Monitoring

Command	Purpose
display ip vrf-info [**brief | count | detail | interfaces**] [*vrf-name*]	Displays a single-protocol VRF.
display global-vrf [**brief** | **detail** | **ipv4** | **ipv6**]	Displays a global VRF.
display vrf [**brief | count | detail | ipv4 | ipv6**] [*vrf-name*]	Displays a Multi-protocol VRF.
cleanup ip mrouting-table [**vrf** *vid*] **statistics** { * | *group-address* [*source-address*] }	Clears the multicast route statistics.
cleanup ip pim dense-mode [**vrf** *vid*] **track**	Clears the PIM-DM packet statistics.
cleanup ip route-config vrf *vrf-name*	Clears the routes in a specified VRF.
debugging vrf	Displays the debugging information when a VRF is created, an address family is enabled, and an interface is added to the VRF.
debugging vrf interface	Displays the debugging information of interface-related VRF operations.

1.8 Configuration Examples

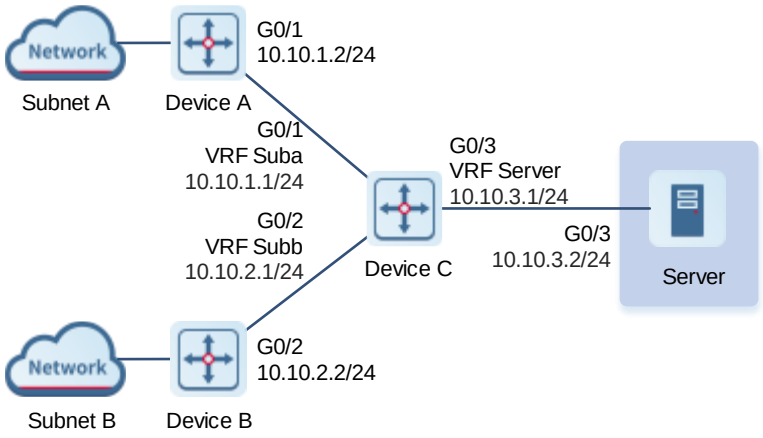
1.8.1 Enabling Local Inter-VPN Access

1. Requirements

Subnet A, Subnet B, and Server are three VPNs that have separate address spaces. Subnet A runs RIP, Subnet B runs OSPF, and Servers is a network segment directly connected to Device C. Routes in Subnet A are separated from those in Subnet B, but both Subnet A and Subnet B can access Server.

2. Topology

Figure 1-4 Configuring Local Inter-VPN Access



3. Notes

- On Device C, create a VRF named **Suba**, add the interface directly connected to Device A to the VRF **Suba**. Associate the VRF **Suba** of Device C with Device A by using RIP.
- On Device C, create a VRF named **Subb**, add the interface directly connected to Device B to the VRF **Subb**. Associate the VRF **Subb** of Device C with Device B by using OSPF.
- On Device C, create a VRF named **Server**, add the interface directly connected to Server to the VRF **Server**.
- On Device C, configure RTs for three VRFs, including **Suba**, **Subb**, and **Server**. Import the routes of **Suba** and **Subb** to **Server**, and the routes of **Server** to **Suba** and **Subb**.
- On Device C, configure the BGP, and import the RIP routes to the IPv4 VRF **Suba**, the OSPF routes to the IPv4 VRF **Subb**, and the direct routes to the IPv4 VRF **Server**.

4. Procedure

(1) (Omitted) Configure the IP address for the device.

(2) Create VRFs on Device C.

Create a VRF named **Suba**, add the interface G0/1 directly connected to Device A to the VRF **Suba**.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ip vrf Suba
DeviceC(cfg-vrf)# rd 100:1
DeviceC(cfg-vrf)# route-target import 100:3
DeviceC(cfg-vrf)# route-target export 100:1
DeviceC(cfg-vrf)# exit
DeviceC(cfg)# interface gigabitethernet 0/1
DeviceC(cfg-GigabitEthernet 0/1)# ip vrf binding Suba
DeviceC(cfg-GigabitEthernet 0/1)# ip address 10.10.1.1 255.255.255.0
DeviceC(cfg-GigabitEthernet 0/1)# exit
```

Create a VRF named **Subb**, add the interface G0/2 directly connected to Device B to the VRF **Subb**.

```
DeviceC(cfg)# ip vrf Subb
DeviceC(cfg-vrf)# rd 100:2
DeviceC(cfg-vrf)# route-target import 100:3
DeviceC(cfg-vrf)# route-target export 100:2
DeviceC(cfg-vrf)# exit
DeviceC(cfg)# interface gigabitethernet 0/2
DeviceC(cfg-GigabitEthernet 0/2)# ip vrf binding Subb
DeviceC(cfg-GigabitEthernet 0/2)# ip address 10.10.2.1 255.255.255.0
DeviceC(cfg-GigabitEthernet 0/2)# exit
```

Create a VRF named **Server**, add the interface G0/3 directly connected to Server to the VRF **Server**.

```
DeviceC(cfg)# ip vrf Server
DeviceC(cfg-vrf)# rd 100:3
DeviceC(cfg-vrf)# route-target import 100:1
DeviceC(cfg-vrf)# route-target import 100:2
DeviceC(cfg-vrf)# route-target export 100:3
DeviceC(cfg-vrf)# exit
DeviceC(cfg)# interface gigabitethernet 0/3
DeviceC(cfg-GigabitEthernet 0/3)# ip vrf binding Server
DeviceC(cfg-GigabitEthernet 0/3)# ip address 10.10.3.1 255.255.255.0
DeviceC(cfg-GigabitEthernet 0/3)# exit
```

(3) Enable the routing protocol.

Enable the routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# router-protocol rip
DeviceA(cfg-router)# version 2
DeviceA(cfg-router)# unset auto-summary
DeviceA(cfg-router)# network-advertise 10.10.1.0 0.0.0.255
```

Enable the routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# router-protocol ospf 1
DeviceB(cfg-router)# network-advertise 10.10.2.0 0.0.0.255 area 0
```

Enable the routing protocol on Device C.

```
DeviceC(cfg)# router-protocol rip
DeviceC(cfg-router)# address-family-mode ipv4 vrf Suba
DeviceC(cfg-router-af)# version 2
DeviceC(cfg-router-af)# unset auto-summary
DeviceC(cfg-router-af)# network-advertise 10.10.1.0 0.0.0.255
DeviceC(cfg-router-af)# exit-address-family-mode
DeviceC(cfg-router)# exit
DeviceC(cfg)# router-protocol ospf 2 vrf Subb
DeviceC(cfg-router)# network-advertise 10.10.2.0 0.0.0.255 area 0
```

```
DeviceC(cfg-router)# exit
```

(4) Import routes to VRFs.

Import routes to VRFs on Device C.

```
DeviceC(cfg)# router-protocol bgp 200
DeviceC(cfg-router)# address-family-mode ipv4 vrf Suba
DeviceC(cfg-router-af)# redistribute-route rip
DeviceC(cfg-router-af)# exit-address-family-mode
DeviceC(cfg-router)# address-family-mode ipv4 vrf Subb
DeviceC(cfg-router-af)# redistribute-route ospf 2
DeviceC(cfg-router-af)# exit-address-family-mode
DeviceC(cfg-router)# address-family-mode ipv4 vrf Server
DeviceC(cfg-router-af)# redistribute-route connected
DeviceC(cfg-router-af)# exit-address-family-mode
```

5. Verification

Verify the information of interfaces added to VRFs on Device C.

```
DeviceC# display ip vrf interfaces
Interface                IP-Address    VRF         Protocol
GigabitEthernet 0/1      10.10.1.1     Suba        up
GigabitEthernet 0/2      10.10.2.1     Subb        up
GigabitEthernet 0/3      10.10.3.1     Server      up
```

Verify that the routes in **Suba** are separated from those in **Subb**, and routes of **Server** are shared on Device C.

```
DeviceC# display ip route-config vrf Suba
Routing Table: Suba

Codes:   C - connected, S - static, R - RIP, B - BGP
         O - OSPF, IA - OSPF inter area
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
         ia - IS-IS inter area, * - candidate default

Gateway of last resort is no set
C       10.10.1.0/24 is directly connected, GigabitEthernet 0/1
C       10.10.1.1/32 is local host.
B       10.10.3.0/24 [20/0] via 0.0.0.0, 00:01:21, GigabitEthernet 0/3

DeviceC# display ip route-config vrf Subb
Routing Table: Subb

Codes:   C - connected, S - static, R - RIP, B - BGP
         O - OSPF, IA - OSPF inter area
         N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
         E1 - OSPF external type 1, E2 - OSPF external type 2
         i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
```

```
            ia - IS-IS inter area, * - candidate default

Gateway of last resort is no set
C       10.10.2.0/24 is directly connected, GigabitEthernet 0/2
C       10.10.2.1/32 is local host.
B       10.10.3.0/24 [20/0] via 0.0.0.0, 00:04:03, GigabitEthernet 0/3

DeviceC# display ip route-config vrf Server
Routing Table: Server

Codes:    C - connected, S - static, R - RIP, B - BGP
            O - OSPF, IA - OSPF inter area
            N1 - OSPF NSSA external type 1, N2 - OSPF NSSA external type 2
            E1 - OSPF external type 1, E2 - OSPF external type 2
            i - IS-IS, su - IS-IS summary, L1 - IS-IS level-1, L2 - IS-IS level-2
            ia - IS-IS inter area, * - candidate default

Gateway of last resort is no set
B       10.10.1.0/24 [20/1] via 0.0.0.0, 03:00:15, GigabitEthernet 0/1
B       10.10.2.0/24 [20/1] via 0.0.0.0, 00:00:14, GigabitEthernet 0/2
C       10.10.3.0/24 is directly connected, GigabitEthernet 0/3
C       10.10.3.1/32 is local host.
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 10.10.1.2 255.255.255.0
!
router-protocol rip
 version 2
 network-advertise 10.10.1.0 0.0.0.255
 unset auto-aggregation
!
```

Device B configuration file

```
hostname DeviceB
!
interface GigabitEthernet 0/2
 unset switch-port
 ip address 10.10.2.2 255.255.255.0
!
router-protocol ospf 2
 network-advertise 10.10.2.0 0.0.0.255 area 0
```

```
!
```

Device C configuration file

```
hostname DeviceC
!
ip vrf Server
 rd 100:3
 route-target export 100:3
 route-target import 100:2
 route-target import 100:1
!
ip vrf Suba
 rd 100:1
 route-target export 100:1
 route-target import 100:3
!
ip vrf Subb
 rd 100:2
 route-target export 100:2
 route-target import 100:3
!
interface GigabitEthernet 0/1
 unset switch-port
 ip vrf binding Suba
 ip address 10.10.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 ip vrf binding Subb
 ip address 10.10.2.1 255.255.255.0
!
interface GigabitEthernet 0/3
 unset switch-port
 ip vrf binding Server
 ip address 10.10.3.1 255.255.255.0
!
router-protocol bgp 200
 address-family-mode ipv4 vrf Suba
  redistribute-route rip
  exit-address-family-mode
 !
 address-family-mode ipv4 vrf Subb
  redistribute-route ospf 2 match-rule internal-route
  exit-address-family-mode
 !
 address-family-mode ipv4 vrf Server
  redistribute-route connected
```

```
   exit-address-family-mode
!
router-protocol ospf 2 vrf Subb
 network-advertise 10.10.2.0 0.0.0.255 area 0
!
Router-protocol rip
 address-family-mode ipv4 vrf Suba
   version 2
   network-advertise 10.10.1.0 0.0.0.255
   unset auto-summary
   exit-address-family-mode
!
```

Server configuration file

```
hostname Server
!
interface GigabitEthernet 0/3
 unset switch-port
 ip address 10.10.3.2 255.255.255.0
!
```

7. Common Errors

- An interface is added to a VRF after the IP address of the interface is configured. In this case, the IP address of the interface is deleted.
- When a physical link is shared by multiple VPNs, the physical interface is added to a VRF.
- VPN routes are not introduced to BGP.

Contents

1 Configuring Routing Policies 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 3
1.3 Configuring a Filter List 4
1.3.1 Overview 4
1.3.2 Restrictions and Guidelines 4
1.3.3 Configuration Tasks 4
1.3.4 Defining AS Path Filtering Rules 4
1.3.5 Defining a Community List 5
1.3.6 Defining an Extcommunity List 5
1.3.7 Creating a Prefix List 6
1.3.8 Adding a Description for a Prefix List 6
1.3.9 Enabling the Sorting Function for a Prefix List 7
1.3.10 Creating an IPv6 Prefix List 7
1.3.11 Adding a Description for an IPv6 Prefix List 7
1.3.12 Enabling the Sorting Function for an IPv6 Prefix List 8
1.4 Configuring a Route Map 8
1.4.1 Overview 8
1.4.2 Configuration Tasks 8

1.4.3 Creating a Policy 9
1.4.4 Configuring the Matching Conditions of a Rule 9
1.4.5 Configuring the Processing Actions of a Policy 11
1.4.6 Configuring the Policy Statement 16
1.5 Monitoring 17
1.6 Configuration Examples 18
1.6.1 Applying Rules Configured in a Route Map During Route Redistribution 18
1.6.2 Applying a Route Map in PBR 20
1.6.3 Configuring a Prefix List 23
1.6.4 Configuring an AS-Path List 25
1.6.5 Configuring a Community List 27

1 Configuring Routing Policies

1.1 Introduction

1.1.1 Overview

Routing policies are a policy set for changing the packet forwarding path or routing information, and are often implemented by a filter list and a route map. Routing policies provide services for different routing protocols, and completes the policy-based control of routing information based on a specified policy set. Policy-based routing (PBR) refers to policy-based routing and forwarding. It can forward packets based on the source address, destination address, packet length, and port, and these matching policies can be provided by routing policies.

Routing policies are flexibly and widely applied in the following ways:

- A filter list is used in a routing protocol to filter or modify routing information.
- A route map is used in a routing protocol to filter or modify routing information. The route map can further use a filter list.
- A route map is used in PBR to control packet forwarding or modify packet fields.

Note

In this chapter, the term "router" refers to any network device that supports the routing function. These network devices can be Layer-3 switches, routers, or firewalls.

1.1.2 Principles

Filter list: It is a group of lists defined based on a routing attribute and can be used by a routing protocol for route filtering.

Route map: It defines a policy that "if certain conditions are matched, processing actions are performed accordingly."

1. Filter list

A filter list is a group of lists defined based on a routing attribute and is a filtering tool of routing policies. To achieve the effect of route filtering, filter lists must be used in a routing protocol or route map. Five types of filter lists are available currently: access control list (ACL), prefix list, autonomous system (AS) path list, community list, and extended community (extcommunity) list.

- ACL

  ACLs include IPv4 and IPv6 ACLs. When defining an ACL, you can specify IPv4/IPv6 addresses and masks to match the destination network segment or next-hop addresses of routing information.

- Prefix list

Similar to ACLs, prefix lists, including IPv4 prefix lists and IPv6 prefix lists, are used to match destination network segments or next hops of routing information.

When configuring a prefix list as the filter list, you can define the prefix address, mask length, and mask range. The prefix address and mask length are used to specify the prefix of the routing information. The mask range specifies the mask range of the routing information to be filtered. That is, a prefix list can be defined to directly filter the subnet segment under a specified prefix network segment without calculating the prefix of the subnet segment in advance. Therefore, prefix lists are more flexible than ACLs.

- AS-path list

  AS-path lists are applicable only to Border Gateway Protocol (BGP). They define filtering rules based on the AS path information. When configuring an AS-path list, you can specify an AS path by using a regular expression to match the AS path.

- Community list

Community lists are applicable only to BGP. They define filtering rules based on the community attributes of routes. When configuring a community list, you can specify a community ID or a regular expression to match the community attribute.

- Extcommunity list

Extcommunity lists are applicable only to BGP. They define filtering rules based on the extended community attributes of routes. When configuring an extcommunity list, you can specify an extended community ID or a regular expression to match the extended community attribute.

2. Route Map

A policy in a route map is a match and set statement, which indicates that if certain conditions are matched, some processing actions will be performed accordingly.

- Main application scenarios of route maps
  - Route filtering: A filter list is used in a routing protocol to filter the routing information sent or received by the protocol.
  - Route redistribution: A route map is used in a routing protocol to filter or modify routing information and redistribute Routing Information Protocol (RIP) routes to Open Shortest Path First (OSPF). Only RIP routes with the hop count of 4 can be redistributed.
  - PBR: A route map is used in PBR to control packet forwarding or modify packet fields and specify optimum outbound interfaces for packets from different subnets.
- Policy execution

  A route map contains multiple policies. Each policy has a corresponding sequence number. A smaller sequence number means a higher priority. You can execute policies by sequence number. Once the matching condition of a policy is met, the processing action of this policy is performed and the route map exits. If no matching condition of any policy is met, no processing action is performed.

- Working modes of policies

  Policies support two working modes:

- **permit**: When the matching condition of a policy is met, the processing action for this policy is performed and the system exits the route map.
- **deny**: When the matching condition of a policy is met, the processing action for this policy is not performed and the system exits the route map.

- Matching conditions of policies

  The matching condition of a policy may contain zero, one, or more match rules.

  - If the matching condition contain zero match rules, all packets match the policy.
  - If the matching condition contains one or more match rules, the matching condition is met only when all the rules are matched.

- Processing action for a policy

  The processing action of a policy may contain zero, one, or more set rules.

  - If the processing action contains zero set rules, no processing action is performed and the system directly exits the route map.
  - If the processing action contains one or more set rules, all processing actions are performed and then the system exits the route map.

1.2 Configuration Task Summary

The configuration of a routing policy includes the following tasks.

(1) Configuring a Filter List
- Defining AS Path Filtering Rules
- Defining a Community List
- Defining an Extcommunity List
- Creating a Prefix List
- Adding a Description for a Prefix List
- Enabling the Sorting Function for a Prefix List
- Creating an IPv6 Prefix List
- Adding a Description for an IPv6 Prefix List
- Enabling the Sorting Function for an IPv6 Prefix List

(2) Configuring a Route Map
- Creating a Policy
- Configuring the Matching Conditions of a Rule
- Configuring the Processing Actions of a Policy
- (Optional) Configuring the Policy Statement

1.3 Configuring a Filter List

1.3.1 Overview

This section describes how to define a set of route filtering rules for a routing protocol.

1.3.2 Restrictions and Guidelines

A configured filter list can take effect only after it is associated with a routing protocol.

1.3.3 Configuration Tasks

The filter list configuration includes the following tasks, which are optional. Select the configuration tasks as required.

- Defining AS Path Filtering Rules
- Defining a Community List
- Defining an Extcommunity List
- Creating a Prefix List
- Adding a Description for a Prefix List
- Enabling the Sorting Function for a Prefix List
- Creating an IPv6 Prefix List
- Adding a Description for an IPv6 Prefix List
- Enabling the Sorting Function for an IPv6 Prefix List

1.3.4 Defining AS Path Filtering Rules

1. Overview

An AS path filter list defines AS path filtering rules by using regular expressions. AS-path is a BGP attribute. Therefore, AS path filtering rules are only applied to a BGP network.

2. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Define AS path filtering rules.
ip as-path access-control-list *path-list-number* { **permit** | **deny** } *regular-expression*

By default, no AS path filtering rule is configured.

1.3.5 Defining a Community List

1. Overview

A community list is a set of routes with the same attributes. Community list is a BGP attribute, and can help reduce the difficulties in BGP route management and maintenance. Similar to an ACL, you can define multiple entries in a community list, and entries are matched based on their sequence numbers. As long as a route matches one entry, the route is permitted by the community list and no further matching is performed.

2. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Define a standard community list.
ip comm-type-list { *community-list-number* | **standard-type** *community-list-name* } { **deny** | **permit** } [*community-list-number* | **gshut** | **internet** | **llgr-stale** | **local-as** | **no-advertise** | **no-export** | **no-llgr**] *

By default, no standard community list is defined.

(4) Define an expanded community list.
ip comm-type-list { *community-list-number* | **expanded** *community-list-name* } { **deny** | **permit** } [*regular-expression*]

By default, no expanded community list is defined.

1.3.6 Defining an Extcommunity List

1. Overview

An extcommunity list is a set of routes with the same attributes. Extcommunity list is a BGP attribute, and can help reduce the difficulties in BGP route management and maintenance. Similar to an ACL, you can define multiple entries in an extcommunity list, and entries are matched based on their sequence numbers. As long as a route matches one entry, the route is permitted by the extcommunity list and no further matching is performed.

2. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Define a standard extcommunity list.
ip extend-comm-list { *standard-list* | **standard-type** *list-name* } { **deny** | **permit** } [**rt** *rt-value* | **soo** *soo-value*] [**color** *color-value* [**only** { 00 | 01 | 10 | 11 }]]

By default, no standard extcommunity list is defined.

(4) Define an expanded extcommunity list.

ip extend-comm-list { *expanded-list* | **expanded** *list-name* } { **deny** | **permit** } [*regular-expression*]

By default, no expanded extcommunity list is defined.

1.3.7 Creating a Prefix List

1. Overview

A prefix list can be directly used by protocols, or used together with a route map as a matching condition of the route map. Similar to an ACL, you can define multiple entries in a prefix list, and entries are matched based on their sequence numbers. As long as a route matches one entry, the route is permitted by the prefix list and no further matching is performed. If a route does not match any entry, it is considered that the route is denied by the prefix list.

2. Restrictions and Guidelines

This command is not a routing policy command, and will not be applied to any routing protocol.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Create a prefix list.
ip prefix-list-info *prefix-list-name* [**seq** *seq-number*] { **deny** | **permit** } *ipv4-prefix* [**ge** *minimum-prefix-length*] [**le** *maximum-prefix-length*]

By default, no prefix list is configured.

1.3.8 Adding a Description for a Prefix List

1. Overview

You can add a text description for a prefix list, for example, to describe the purpose of the prefix list. Adding descriptions for prefix lists can facilitate maintenance and tracking.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Add a text description for the IPv4 prefix list.

ip prefix-list-info *prefix-list-name* **description** *description-text*

By default, no text description is configured for the prefix list.

1.3.9 Enabling the Sorting Function for a Prefix List

1. Overview

When you run the **display ip prefix-list-info** command, the sequence numbers of entries in a prefix list are not displayed by default. If these sequence numbers are displayed, you can conveniently add or remove an entry to or from the prefix list.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of displaying sequence numbers in the prefix list.

ip prefix-list-info sequence-number

The function of displaying sequence numbers in a prefix list is disabled by default.

1.3.10 Creating an IPv6 Prefix List

1. Overview

An IPv6 prefix list can be directly used by protocols, or used together with a route map as a matching condition of the route map. Similar to an ACL, you can define multiple entries in an IPv6 prefix list, and entries are matched based on their sequence numbers. As long as a route matches one entry, the route is permitted by the prefix list and no further matching is performed. If a route does not match any entry, it is considered that the route is denied by the IPv6 prefix list.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an IPv6 prefix list.

ipv6 prefix-list-info *prefix-list-name* [**seq** *seq-number*] { **deny** | **permit** } *ipv6-prefix* [**ge** *minimum-prefix-length*] [**le** *maximum-prefix-length*]

By default, no IPv6 prefix list is configured.

1.3.11 Adding a Description for an IPv6 Prefix List

1. Overview

You can add a text description for a prefix list, for example, to describe the purpose of the prefix list. Adding descriptions for prefix lists can facilitate maintenance and tracking.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Add a description for the IPv6 prefix list.

ipv6 prefix-list-info *prefix-list-name* **description** *description-text*

By default, no description is configured for an IPv6 prefix list.

1.3.12 Enabling the Sorting Function for an IPv6 Prefix List

1. Overview

When you run the **display ipv6 prefix-list-info** command, the sequence numbers of entries in a prefix list are not displayed by default. If these sequence numbers are displayed, you can conveniently add or remove an entry to or from the prefix list.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of displaying sequence numbers in the IPv6 prefix list.

ipv6 prefix-list-info sequence-number

The function of displaying sequence numbers in an IPv6 prefix list is displayed by default.

1.4 Configuring a Route Map

1.4.1 Overview

This section describes how to define a set of routing policies to be used by routing protocols or PBR.

1.4.2 Configuration Tasks

The route map configuration includes the following tasks:

(1) Creating a Policy

(2) Configuring the Matching Conditions of a Rule

(3) Configuring the Processing Actions of a Policy

(4) (Optional) Configuring the Policy Statement

1.4.3 Creating a Policy

1. Overview

You can create a route map to control route attributes.

2. Restrictions and Guidelines

- If the route map is unavailable, this command will create a route map and add a policy to the route map.
- If the route map is available, this command will add a policy to the route map.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a route map and enter the route map configuration mode.

route-map *route-map-name* [{ **permit** | **deny** } *sequence-number*]

1.4.4 Configuring the Matching Conditions of a Rule

1. Overview

In a route map, you can configure match rules based on the route attributes. If multiple match rules are configured, all the rules must be matched.

2. Restrictions and Guidelines

- If you use an ACL in a **match** command to define packet matching conditions, you must configure this ACL.
- The following table lists the **match** commands that cannot be configured at the same time:

Table 1-1 Match Commands That Cannot Be Configured at the Same Time

match Command	Exclusive match Command
match-rule ip address	**match-rule ip prefix-list-info**
match-rule ipv6 address	**match-rule ipv6 prefix-list-info**
match-rule ip next-hop	**match-rule ip next-hop prefix-list-info**
match-rule ipv6 next-hop	**match-rule ipv6 next-hop prefix-list-info**
match-rule ip route-source	**match-rule ip route-source prefix-list-info**
match-rule ipv6 route-source	**match-rule ipv6 route-source prefix-list-info**

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the route map configuration mode.

route-map *route-map-name* [{ **permit** | **deny** } *sequence-number*]

(4) Configure a next-hop outbound interface matching rule.

match-rule interface { *interface-type interface-number* }&<1-4>

No next-hop outbound interface matching rule is configured by default.

You can use an IPv4 filter list in the route map to match attributes. The following configurations are optional. You can configure multiple match rules, and all of which must be matched if multiple match rules are configured.

- Configure an AS path filtering rule for path matching.

  match-rule as-path *as-path-acl-list-number*&<1-10>

  By default, no AS-path list is configured for packet matching.

- Configure a community list for packet matching.

  match-rule comm { *community-list-number* [**exact-match**] | *community-list-name* [**exact-match**] } &<1-6>

  By default, no community list is configured for packet matching.

- Configure an extcommunity list for packet matching.

  match-rule extend-comm *extcommunity-list*

  By default, no extcommunity list is configured for packet matching.

- Configure a rule to match the target network routes that are permitted in the ACL or prefix list.
  match-rule ip address { { *acl-name* | *acl-number* }&<1-6> | **prefix-list-info** *prefix-list-name*&<1-6> }

  By default, no ACL or prefix list used for packet matching is configured.

- Configure a rule to match the target network routes whose next-hop IPv4 addresses meet rules in the ACL or prefix list.

  match-rule ip next-hop { { *acl-name* | *acl-number* }&<1-6> | **prefix-list-info** *prefix-list-name*&<1-6> }

  No next-hop IPv4 address matching rule is configured by default.

- Configure a rule to match the target network routes whose source IPv4 addresses meet rules in the ACL or prefix list.

  match-rule ip route-source { { *acl-name* | *acl-number* }&<1-6> | **prefix-list-info** *prefix-list-name*&<1-6> }

  No IPv4 source address matching rule is configured by default.

(5) You can use an IPv6 filter list in the route map to match attributes.

- Configure a rule to match the target IPv6 network routes that are permitted in the ACL or prefix list.
  match-rule ipv6 address { *acl-name* | **prefix-list-info** *prefix-list-name* }

  By default, no IPv6 ACL or IPv6 prefix list is configured.
- Configure a rule to match the target network routes whose next-hop IPv6 addresses meet rules in the ACL or prefix list.

  match ipv6 address { *acl-name* | **prefix-list-info** *prefix-list-name* }

  No next-hop IPv6 address matching rule is configured by default.
- Configure a rule to match the target network routes whose source IPv6 addresses meet rules in the ACL or prefix list.

  match ipv6 route-source { *acl-name* | **prefix-list-info** *prefix-list-name* }

  No IPv6 source address matching rule is configured by default.

(6) Configure a MAC address rule for route matching.

match-rule mac-address *mac-address*

No MAC address-based route matching rule is configured by default.

(7) Configure a rule to match the routes with a fixed metric.

match-rule metric *metric*

By default, the metric values of routes are not matched.

(8) Configure a rule to match the route type.

- Configure a rule for matching the source type of BGP routes.

  match-rule origin { **egp** | **igp** | **incomplete** }

  By default, no rule is configured to match the source type of BGP routes.
- Configure a rule to match the route type.
  match-rule route-type { **connect** | **evpn-type-1** | **evpn-type-2** | **evpn-type-3** | **evpn-type-4** | **evpn-type-5** | **external-route** [**type-1** | **type-2**] | **internal-route** | **level-1** | **level-2** | **local** | **nssa-external-type-route** [**type-1** | **type-2**] | **rip** | **static** } *

  By default, no rule is configured to match any route type.

(9) Configure a rule to match the routes with a specified tag.

match-rule tag *tag*&<1-4>

By default, no rule to match the routes with a specified tag is configured.

1.4.5 Configuring the Processing Actions of a Policy

1. Overview

Processing actions are configured for traffic that matches the rules. The device provides abundant **set** commands for configuring flexible processing actions.

2. Restrictions and Guidelines

- If no set rule is configured, no processing action will be performed.
- If multiple set rules are configured, all set rules must be executed. If the set rules have different priorities, only the set rule with the highest priority takes effect.
- The following table lists the **set-rule** commands that cannot be configured at the same time.

Table 1-2 set-rule Commands That Cannot Be Configured at the Same Time

set Command	Exclusive set Command
set-rule ip-address next-hop	**set-rule ip-address next-hop verify-availability**
set-rule ip-address dscp	**set-rule ip-address tos**
set-rule ip-address dscp	**set-rule ip-address precedence**

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the route map configuration mode.

route-map *route-map-name* [{ **permit** | **deny** } *sequence-number*]

(4) Configure a rule to adjust the BGP attributes.

- Configure the AS number for BGP.

  set-rule aggregator as *as-number ipv4-address*

  By default, no AS number of the aggregator is specified.

- Configure the Accumulated Interior Gateway Protocol (AIGP) metric attribute for Exterior Gateway Protocol (EGP).

  set-rule aigp-metric { *metric-number* | **igp-metric** }

  By default, no AIGP metric attribute is configured for routes.

- Configure a rule to add the specified AS_PATH values.

  set-rule as-path prepend *as-number*&<1-10>

  By default, no rule is configured to configure AS_PATH values for routes.

- Configure a rule to replace the AS_PATH attributes of routes.

  set-rule as-path replace *as-number*&<1-10>

By default, the AS_PATH values of matched routes will not be replaced.

- Configure the atomic-aggregate attribute for routes.

  set-rule atomic-aggregate

  By default, no atomic-aggregate attribute is configured for routes.

- Configure a rule to delete all the community values from BGP routes matching the rules according to the community list.

  set-rule comm-list { *community-list-number* | *community-list-name* } **delete**

  By default, no rule is configured to delete all the community values from routes matching the rules according to the community list.

- Configure the community value for BGP routes.

  set-rule comm { { *community-number*&<1-32> | **internet** | **llgr-stale** | **local-AS** | **no-advertise** | **no-export** | **no-llgr** | **none** } * [**additive**] }

  By default, no community value is specified for a route.

- Configure the flapping parameters for BGP routes.

  set-rule dampening *half-life reuse suppress max-suppress-time*

  By default, when the half-life period is 15 minutes, and the penalty value of a route is lower than 750, route suppression is canceled. When the penalty value of a route exceeds 2000, the route is suppressed. A route can be suppressed for a maximum of 60 minutes.

- Configure a rule to delete the extended community attribute of BGP routes.

  set-rule extcomm-list { *extcommunity-list-number* | *extcommunity-list-name* } **delete**

- Configure the extended community attribute of BGP routes.

  set-rule extend-comm { **rt** *extend-community-value* [**additive**] | **soo** *extend-community-value* }

- Specify the extended community value of **color** for routes that match rules.
  set-rule extend-comm color *color-value* [**only** { 00 | 01 | 10 | 11 }] [**additive**]

(5) Configure the backup outbound interface and backup next hop for fast reroute (FRR).

- Configure the backup outbound interface and backup next hop for FRR.

  set-rule frr backup-interface *interface-type interface-number* **backup-nexthop** *ipv4-address*

  By default, no backup outbound interface and backup next hop of FRR are specified for matched routes.

(6) Configure the next hop.

- Configure the default next-hop IPv4 address for routes.

  set-rule ip-address [**global** | **vrf** *vrf-name*] **default next-hop** { *ipv4-address* [*weight*] }&<1-32>

  By default, no default next-hop IPv4 address is specified for matched routes.

  Note
  This command is not supported by the S6980-64QC, S6920-4C, S6580-48CQ8QC, or S6580-24DC8QC.

- Configure the next-hop IPv4 address for IPv4 packets.

set-rule ip-address [**global vrf** | *vrf-name*] **next-hop** { *ipv4-address* [*weight*] }&<1-32>

By default, no next-hop IPv4 address is specified for matched packets.

- Configure the next-hop IPv4 address for routes.
  set-rule next-hop *ipv4-address*

  By default, no next-hop IPv4 address is specified for matched routes.

- Configure the default next-hop IPv6 address for an IPv6 packet.
  set-rule ipv6-address [**global** | **vrf** *vrf-name*] **default next-hop** { *ipv6-address* [*weight*] }&<1-32>

  By default, no default next-hop IPv6 address is specified for matched IPv6 routes.

  Note

  This command is not supported by the S6980-64QC, S6920-4C, S6580-48CQ8QC, or S6580-24DC8QC.

- Configure the next-hop IPv6 address for an IPv6 packet.
  set-rule ipv6-address [**global** | **vrf** *vrf-name*] **next-hop** { *ipv6-address* [*weight*] }&<1-32>

- Set the next hops of routes to the device itself.
  set-rule ip-address next-hop self

  By default, the next hop is not set to the device itself for matched packets.

- Configure the next hop of a BGP IPv4 route to keep unchanged.
  set-rule ip-address next-hop unchanged

  By default, the next hop is not set to keep unchanged for matched packets.

- Configure the recursive next-hop IP address for packets.
  set-rule ip-address [**vrf** *vrf-name* | **global**] **next-hop recursive** *ipv4-address*

  By default, no recursive next-hop IPv4 address is specified for matched packets.

- Configure the next hop of BGP IPv4 and IPv6 routes to the device itself.
  set-rule next-hop self

  By default, the next hop is not set to the device itself for matched routes.

- Configure the next hop of the routes to keep unchanged.
  set-rule next-hop unchanged

  By default, the next hop is not set to keep unchanged for matched routes.

- Configure the recursive next-hop IPv6 address.
  set-rule ipv6-address [**vrf** *vrf-name* | **global**] **next-hop recursive** *ipv6-address*

  By default, no recursive next-hop IPv6 address is specified for matched IPv6 packets.

- Configure the next hop of the route to the device itself.
  set-rule ipv6-address next-hop self

  By default, the next hop address is not set to the device itself for matched IPv6 packets.

- Set the next hop of the specified IPv6 route to keep unchanged.
  set-rule ipv6-address next-hop unchanged

  By default, the next hop is not set to keep unchanged for matched IPv6 packets.

- Set the L3 VPN next hop to the local virtual routing and forwarding (VRF) instance.

set-rule l3vpn nexthop local-vrf

By default, the L3 VPN next hop is not set to the local VRF instance for routes matching the rules.

- Verify availability of the next-hop IPv4 address.
  set-rule ip-address [**vrf** *vrf-name* | **global**] **next-hop verify-accessibility** *ipv4-address* [*weight*] {**track** *track-obj-number* | **bfd** *interface-type interface-number gateway* }

  By default, availability of the next-hop IPv4 address is not verified for matched packets.

- Verify availability of the next-hop IPv6 address.
  set-rule ipv6-address [**vrf** *vrf-name* | **global**] **next-hop verify-accessibility** *ipv6-address* [*weight*] **bfd** *interface-type interface-number gateway*

  By default, availability of the next-hop IPv6 address is not verified for matched packets.

(7) Configure the flow control parameters.

- Configure the IPv4 header priority.
  set-rule ip-address precedence { *precedence* | **critical** | **flash** | **flash-override** | **immediate** | **internet** | **network** | **priority** | **routine** }

  By default, no IPv4 header priority is configured for matched packets.

- Configure the IPv6 header priority.
  set-rule ipv6-address precedence { *precedence-number* | **critical** | **flash** | **flash-override** | **immediate** | **internet** | **network** | **priority** | **routine** }

  By default, no IPv6 header priority is configured for matched packets.

- Configure the type of service (ToS) for the IPv4 header of a packet.
  set-rule ip-address tos { *tos-number* | **max-reliability** | **max-throughput** | **min-delay** | **min -cost** | **normal** }

  By default, no ToS is configured for the IP header of a matched packet.

- Configure the Differentiated Services Code Point (DSCP) in an IPv4 packet.
  set-rule ip-address dscp *dscp_value*

  By default, no DSCP is configured in an IPv4 packet for matched routes.

- Configure the quality of service (QoS) ID for a route.
  set-rule qos-id *qos-id*

  By default, no QoS ID is specified for matched routes.

- Configure the LOCAL_PREFERENCE parameter.
  set-rule local-preference *precedence-number*

  By default, no LOCAL_PREFERENCE value is configured for matched routes.

(8) Configure redistribution-related information.

- Configure the type of the destination area to which redistributed routes are to be advertised.
  set-rule level { backbone | **level-1 | level-1-2 | level-2** | **stub-area** }

  By default, the type of the destination area is not specified for matched routes.

- Configure the redistributed route tag.
  set-rule tag *tag*

By default, no route tag is configured for matched routes.

(9) Configure the weight, metric, and metric type.

- Configure the metric for routes.
  set-rule metric { **+** *metric-value* | **-** *metric-value* | *metric-value* }

  By default, the metric of a matched route is not modified.

- Configure the metric type for routes.
  set-rule metric-type { **external** | **internal** | **type-1** | **type-2** }

- Configure the weight for BGP routes.
  set-rule weight *weight-number*

  By default, no weight is configured for matched routes.

(10) Configure the route source information.

- Configure the route source attribute.
  set-rule origin { **egp** | **igp** | **incomplete** }

  By default, no route source is specified for matched routes.

- Configure the route originator address.
  set-rule originator-inf *ipv4-address*

  By default, no route originator address is configured for matched routes.

(11) Configure the route management distance.

set-rule distance distance-*number*

By default, no management distance is modified for matched routes.

1.4.6 Configuring the Policy Statement

1. Overview

Similar to the description of a prefix list, you can configure a statement to describe a route map.

2. Restrictions and Guidelines

- This command is not a routing policy command, and will not be applied to any routing protocol.
- This command takes effect only in OPENCONF YANG, and is used to record the statement name of a route map.
- If this command is not configured, the sequence number of a route map is used as the default statement name of the route map. Therefore, the range from 0 to 65535 is reserved for the **statement-name** command as the default configuration range. Within this range, all the sequence numbers except the sequence number of a route map configured with **statement-name** cannot be used because they are reserved for the sub-route maps with the same name as the route map and used as the default statement content.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the route map configuration mode.

route-map *route-map-name* [{ **permit** | **deny** } *sequence-number*]

(4) Configure the statement name for the route map.

statement-name [*name*]

By default, the statement name is the sequence number of the route map.

1.5 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Table 1-3 Routing Policy Monitoring

Command	Purpose
display route-map [*route-map-name*]	Displays the configurations of a route map.
display access-control-lists [*acl-id* | *name*]	Displays the configurations of an ACL.
display ip prefix-list-info [*prefix-name*]	Displays the configurations of an IPv4 prefix list.
display ip protocols [**vrf** *vrf-name*] [**bgp** | **isis** | **ospf** | **rip**]	Displays the status of the IPv4 routing protocols that are currently running.
display ipv6 prefix-list-info [*prefix-name*]	Displays the configurations of an IPv6 prefix list.
display ip as-access-control-list [*as-path-access-list-number*]	Displays the configurations of an AS-path list.
display ip comm-type-list [*community-list-number* | *community-list-name*]	Displays the configurations of a community list.
display ip extend-comm-list [*extcommunity-list-number* | *extcommunity-list-name*]	Displays the configurations of an extcommunity list.

1.6 Configuration Examples

1.6.1 Applying Rules Configured in a Route Map During Route Redistribution

1. Requirements

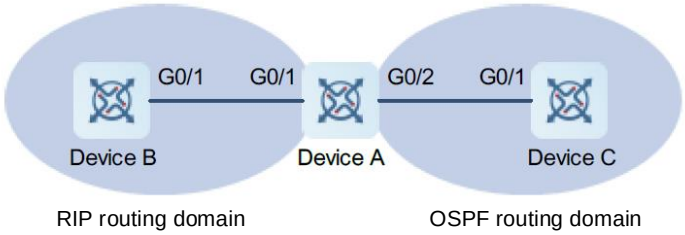
The RIP routing domain needs to exchange routes with the OSPF routing domain to realize the network-wide interworking.

When RIP routes are redistributed to OSPF, the routes with too many hops are considered unreliable. Therefore, it is required that only RIP routes with the hop count of 4 be redistributed. In the OSPF routing domain, if a route is a Type-1 external route, the tag value of the route is set to 40.

When OSPF routes are redistributed to RIP, only OSPF routes with the tag value of 10 are redistributed. In the RIP routing domain, the initial metric of the routes redistributed to RIP is set to 10.

2. Topology

Figure 1-1 Applying Rules Configured in a Route Map During Route Redistribution



3. Notes

- Configure the route map **redrip** as follows: Configure a rule to match routes with the hop count of 4, set the initial metric of the routes to **40**, set the route type to the external route **type-1**, and set the tag of the routes to **40**.
- Configure the route map **redospf** as follows: Configure a rule to match routes with the tag of **10** and set the initial metric of the routes to **10**.
- Apply the route map **redrip** when RIP routes are redistributed to OSPF.
- Apply the route map **redospf** when OSPF routes are redistributed to RIP.

4. Procedure

(1) Configure IP addresses of the interfaces, and complete the basic OSPF and RIP configuration. (Omitted)

(2) Configure a route redistribution policy for OSPF. In the policy, configure a rule to match routes with the hop count of 4, set the route type to **Type-1** and the tag to **40**.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# route-map redrip permit 10
```

```
Device A(cfg-routemap)# match-rule metric 4
Device A(cfg-routemap)# set-rule metric-type type-1
Device A(cfg-routemap)# set-rule tag 40
Device A(cfg-routemap)# exit
```

(3) Configure a route redistribution policy for RIP. In the policy, configure a rule to match routes with the route tag of **10** and set the metric to **10**.

```
Device A(cfg)# route-map redospf permit 10
Device A(cfg-routemap)# match-rule tag 10
Device A(cfg-routemap)# set-rule metric 10
Device A(cfg-routemap)# exit
```

(4) Configure redistribution of RIP routes to OSPF, and use the route map **redrip** for routing control.

```
Device A(cfg)# router-protocol ospf 1
Device A(cfg-router)# redistribute-route rip subnets route-map redrip
Device A(cfg-router)# exit
```

(5) Configure the redistribution of OSPF routes to RIP, and use the route map **redospf** for routing control.

```
Device A(cfg)# router-protocol rip
Device A(cfg-router)# redistribute-route ospf 1 route-map redospf
Device A(cfg-router)# exit
```

5. Verification

Check the route map configurations on device A to verify the policy rules.

```
Device A# display route-map
route-map redrip, permit, sequence 10
  Match clauses:
    metric 4
  Set clauses:
    metric 40
    metric-type type-1
    tag 40
route-map redospf, permit, sequence 10
  Match clauses:
    tag 10
  Set clauses:
metric 10
```

Check the OSPF routing information library on device A to verify that the routes matching the policy rules are redistributed.

```
Device A# display ip ospf database-info external-route

                OSPF Router with ID (192.100.1.9) (Process ID 1)
```

```
                AS External Link States

  LS age: 5
  Options: 0x2 (-|-|-|-|-|-|E|-)
  LS Type: AS-external-LSA
  Link State ID: 192.168.199.0 (External Network Number)
  Advertising Router: 192.100.1.9
  LS Seq Number: 80000001
  Checksum: 0x554d
  Length: 36
  Network Mask: /24
        Metric Type: 1
        TOS: 0
        Metric: 4
        Forward Address: 0.0.0.0
        External Route Tag: 40
```

6. Configuration Files

Device A configuration file

```
!
route-map redrip permit 10
 match-rule metric 4
 set-rule metric-type type-1
 set-rule tag 40
!
route-map redospf permit 10
 match-rule tag 10
 set-rule metric 10
!
router-protocol ospf 1
redistribute-route rip route-map redrip subnets
!
router-protocol rip
redistribute-route ospf 1 route-map redospf
!
```

1.6.2 Applying a Route Map in PBR

1. Requirements

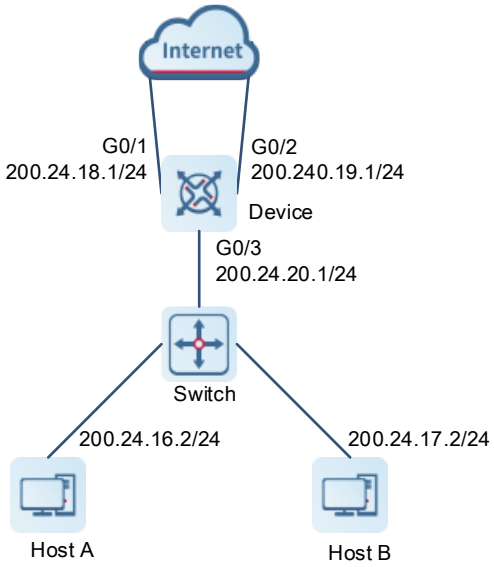
Configure PBR on the devices to achieve the following purposes:

- Packets from subnet 1 (200.24.16.0/24) are sent from GE0/1 first.
- Packets from subnet 2 (200.24.17.0/24) are sent from GE0/2 first.

The two egress links work in mutual backup mode.

2. Topology

Figure 1-2 Applying a Route Map in PBR



3. Notes

- Configure two different ACLs to match packets from subnets 1 and 2 respectively.
- Configure the route map RM_FOR_PBR: Configure policy 10 to ensure that "packets from host A are sent from GigabitEthernet 0/1 first"; configure policy 20 to ensure that "packets from host B are sent from GigabitEthernet 0/2 first".
- Configure PBR and apply the route map RM_FOR_PBR to packets received from GigabitEthernet 0/3.
- Configure a PBR to implement redundant backup among multiple next hops.

4. Procedure

(1) Configure the IP addresses of interfaces to implement interworking between networks. (Omitted)

(2) Configure ACLs for subnets 200.24.16.0/24 and 200.24.17.0/24, respectively.

```
Device> privilege
Device# config-system terminal
Device(cfg)# access-control-list 1 permit 200.24.16.0 0.0.0.255
Device(cfg)# access-control-list 2 permit 200.24.17.0 0.0.0.255
```

(3) Configure the routing map **RM_FOR_PBR**, set the next hop of 200.24.16.0 to 200.24.18.1 and its backup next hop to 200.24.19.1, and set the next hop of 200.24.17.0 to 200.24.19.1 and its backup next hop to 200.24.18.1.

```
Device(cfg)# route-map RM_FOR_PBR 10
```

```
Device(cfg-routemap)# match-rule ip address 1
Device(cfg-routemap)# set-rule ip-address next-hop 200.24.18.1
Device(cfg-routemap)# set-rule ip-address next-hop 200.24.19.1
Device(cfg-routemap)# exit
Device(cfg)# route-map RM_FOR_PBR 20
Device(cfg-routemap)# match-rule ip address 2
Device(cfg-routemap)# set-rule ip-address next-hop 200.24.19.1
Device(cfg-routemap)# set-rule ip-address next-hop 200.24.18.1
Device(cfg-routemap)# exit
```

(4) Apply the route map **RM_FOR_PBR** to the interface GigabitEthernet 0/3, and enable load balancing.

```
Device(cfg)# interface GigabitEthernet 0/3
Device(cfg-if-GigabitEthernet 0/3)# ip policy-based route-map RM_FOR_PBR
Device(cfg-if-GigabitEthernet 0/3)# exit
Device(cfg)# ip policy-based redundance
```

5. Verification

Check the PBR configurations to verify that the route map is applied to the interface.

```
Device# display ip policy-based
Balance mode: redundance
Interface                          Route map
GigabitEthernet 0/3                RM_FOR_PBR
```

Check the configurations of the route map to verify the policy rules.

```
Device# display route-map
route-map RM_FOR_PBR, permit, sequence 10
   Match clauses:
      ip address 1
   Set clauses:
      ip next-hop 200.24.18.1
      ip next-hop 200.24.19.1
route-map RM_FOR_PBR, permit, sequence 20
   Match clauses:
      ip address 2
   Set clauses:
      ip next-hop 200.24.19.1
      ip next-hop 200.24.18.1
```

Check the ACL configurations to verify the packet filtering rules.

```
Device# display access-control-lists
ip access-control-list standard-type 1
  10 permit 200.24.16.0 0.0.0.255
  10 permit 200.24.16.0 0.0.0.255
ip access-control-list standard-type 2
```

```
10 permit 200.24.17.0 0.0.0.255
```

6. Configuration Files

Device configuration file

```
!
ip policy-based redundance
!
access-control-list 1 permit 200.24.16.0 0.0.0.255
access-control-list 2 permit 200.24.17.0 0.0.0.255
!
route-map RM_FOR_PBR 10
 match-rule ip address 1
 set-rule ip-address next-hop 200.24.18.1
 set-rule ip-address next-hop 200.24.19.1
!
route-map RM_FOR_PBR 20
 match-rule ip address 2
 set-rule ip-address next-hop 200.24.19.1
 set-rule ip-address next-hop 200.24.18.1
!
interface GigabitEthernet 0/3
 ip policy-based route-map RM_FOR_PBR
!
```

7. Common Errors

- After a policy that uses ACLs and prefix-lists for matching is configured, the corresponding ACLs and prefix lists are not defined.

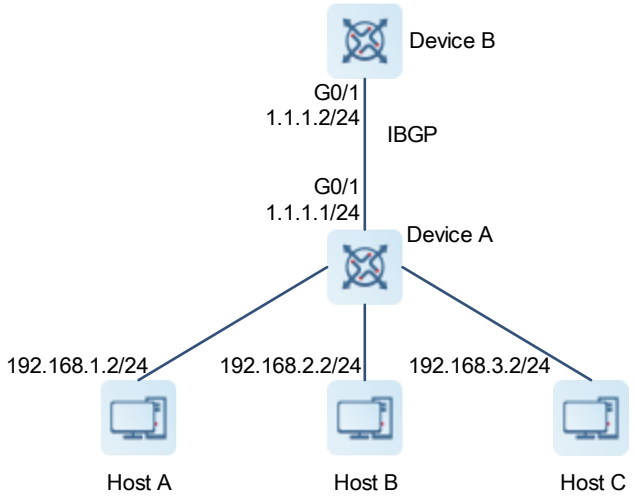
1.6.3 Configuring a Prefix List

1. Requirements

Device A and device B are Interior Border Gateway Protocol (IBGP) neighbors. Device A performs routing control, and only advertises 192.168.1.0/24 network segment of the route to neighbor Device B.

2. Topology

Figure 1-3 Configuring a Prefix List



3. Notes

- Configure Device A and Device B as IBGP neighbors and advertise routes to the connected three subnets.
- Configure a prefix list.
- Associate the prefix list with device A to filter sent routes.

4. Procedure

(1) Configure the IP addresses of interfaces, and configure device A and device B to advertise routes and become IBGP neighbors. (Omitted)

(2) Configure a prefix list to match the traffic of 192.168.1.0/24.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ip prefix-list-info pre1 permit 192.168.1.0/24
```

(3) Specify the neighbor 1.1.1.2 in BGP, and apply the prefix list **pre1** to filter sent routes.

```
Device A(cfg)# router-protocol bgp 100
Device A(cfg-router)# peer 1.1.1.2 prefix-list-info pre1 out
```

5. Verification

Check the prefix list.

```
Device A> privilege
Device A# display ip prefix-list-info
ip prefix-list-info pre1: 1 entries
```

```
        seq 5 permit 192.168.1.0/24

Device A# display ip bgp-instance
BGP table version is 2, local router ID is 1.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                    S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

    Network          Next Hop            Metric     LocPrf     Weight Path
*> 192.168.1.0      0.0.0.0                  0                    32768      i
*> 192.168.2.0      0.0.0.0                  0                    32768      i
*> 192.168.3.0      0.0.0.0                  0                    32768      i

Total number of prefixes 3
```

Check the BGP routing table to verify that the routes are filtered correctly.

```
Device B> privilege
Device B# display ip bgp-instance
BGP table version is 4, local router ID is 1.1.1.2
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                    S Stale
Origin codes: i - IGP, e - EGP, ? - incomplete

    Network          Next Hop            Metric     LocPrf     Weight Path
*>i192.168.1.0      1.1.1.1                  0          100          0      i

Total number of prefixes 1
```

6. Configuration Files

- Device A configuration file

```
!
ip prefix-list-info pre1 permit 192.168.1.0/24
!
router-protocol bgp 100
peer 1.1.1.2 prefix-list-info pre1 out
!
```

1.6.4 Configuring an AS-Path List

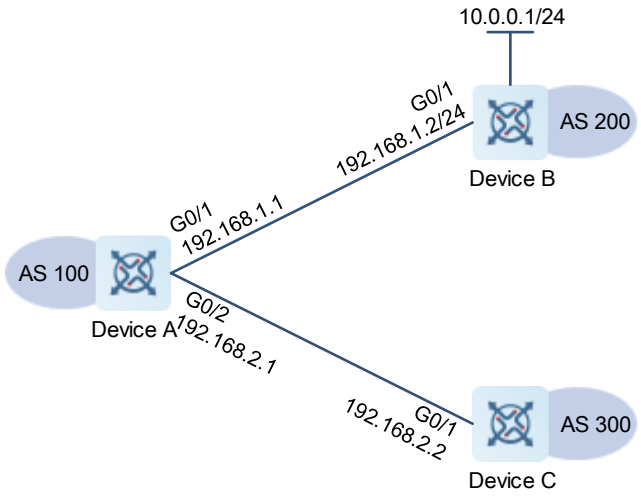
1. Requirements

Device A, Device B and Device C are EBGP neighbors.

Device A filters only the route information containing AS 200 from Device B and Device C.

2. Topology

Figure 1-4 Configuring an AS-Path List



3. Notes

- Create an AS-path filtering rule to match path information containing only the AS number of 200.
- On device A, establish the EBGP neighborship with device B and device C.
- Associate an AS-path list with device A to filter the routes received from device B and device C.

4. Procedure

(1) Configure IP addresses of the interfaces, and complete the basic EBGP configuration. (Omitted)

(2) Configure an AS-path list to match routes with the AS number of 200.

```
Device A(cfg)# ip as-path access-control-list 123 permit ^200$
```

(3) On device A, specify the BGP neighbors, and use the AS path list **123** to filter routes.

```
Device A(cfg)# router-protocol bgp 100
Device A(cfg-router)# peer 192.168.1.2 filter-rule-list 123 in
Device A(cfg-router)# peer 192.168.2.2 filter-rule-list 123 in
```

5. Verification

Check the AS-path list.

```
Device A> privilege
Device A# display ip as-access-control-list
AS path access list 123
permit ^200$
```

Before associating the AS-path list, check the BGP routing table to verify that all routes are learned.

```
Device A# display ip bgp-instance
BGP table version is 1, local router ID is 1.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop        Metric     LocPrf     Weight Path
*> 10.0.0.0/24      192.168.1.2          0                 0 200 i
*> 20.0.0.0/24      192.168.2.2          0                 0 300 i

Total number of prefixes 2
```

After associating the AS path list with device A, check the BGP routing table to verify that routes are filtered correctly.

```
Device A# display ip bgp-instance
BGP table version is 1, local router ID is 1.1.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
              S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop        Metric     LocPrf     Weight Path
*> 10.0.0.0/24      192.168.1.2          0                 0 200 i
Total number of prefixes 1
```

6. Configuration Files

- Device A configuration file

```
!
ip as-path access-control-list 123 permit ^200$
!
router-protocol bgp 100
peer 192.168.1.2 filter-rule-list 123 in
peer 192.168.2.2 filter-rule-list 123 in
!
```

1.6.5 Configuring a Community List

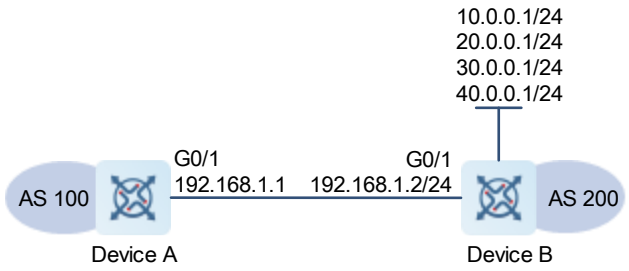
1. Requirements

Device A and Device B, are EBGP neighbors.

The routes received from Device B on Device A will be filtered, and only routes with the community attribute of 100:20 will be received.

2. Topology

Figure 1-5 Configuring a Community List



3. Notes

- Define a standard community list to match routes with the community attribute of **100:20**.
- Establish the EBGP neighborship between device A and device B.
- Configure device B to advertise routes with the community attribute.
- Associate the community list on device A (BGP can be applied only through a route map) to filter routes received from device B.

4. Procedure

(1) Configure IP addresses of the interfaces, and complete the related basic configuration. (Omitted)

(2) Configure a route map on device B, and configure the community attribute.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# route-map comm1
Device B(cfg-routemap)# set-rule comm 100:20 200:20
Device B(cfg-routemap)# exit
Device B(cfg)# route-map comm2
Device B(cfg-routemap)# set-rule comm 100:20
Device B(cfg-routemap)# exit
Device B(cfg)# route-map comm3
Device B(cfg-routemap)# set-rule comm 200:20
Device B(cfg-routemap)# exit
```

(3) Configure device B to advertise routes with the community attribute.

```
Device B(cfg)# router-protocol bgp 200
Device B(cfg-router)# peer 192.168.1.1 send-community
Device B(cfg-router)# network-advertise 10.0.0.0 mask 255.255.255.0 route-map comm1
Device B(cfg-router)# network-advertise 20.0.0.0 mask 255.255.255.0 route-map comm2
Device B(cfg-router)# network-advertise 30.0.0.0 mask 255.255.255.0 route-map comm3
```

```
Device B(cfg-router)# network-advertise 40.0.0.0 mask 255.255.255.0
```

(4) On device A, configure a community list and route map to match routes with the community attribute of **100:20**.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ip comm-type-list standard-type test permit 100:20
Device A(cfg)# route-map COM
Device A(cfg-routemap)# match-rule comm test
Device A(cfg-routemap)# exit
```

(5) On device A, specify the neighbor 192.168.1.2, and use the route map **COM** to filter routes received from the neighbor.

```
Device A(cfg)# router-protocol bgp 100
Device A(cfg-router)# peer 192.168.1.2 route-map COM in
```

5. Verification

Check the community list of device A.

```
Device A> privilege
Device A# display ip comm-type-list
Named Community standard-type list test
permit 100:20
```

Before associating the community list with device A, check the BGP routing table.

```
Device A# display ip bgp-instance
BGP table version is 1, local router ID is 192.168.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                    S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric     LocPrf     Weight Path
*> 10.0.0.0/24      192.168.1.2              0                    0 200 i
*> 20.0.0.0/24      192.168.1.2              0                    0 200 i
*> 30.0.0.0/24      192.168.1.2              0                    0 200 i
*> 40.0.0.0/24      192.168.1.2              0                    0 200 i

Total number of prefixes 4
```

Check the BGP routes on device A.

```
Device A# display ip bgp-instance 10.0.0.0
BGP routing table entry for 10.0.0.0/24
Paths: (1 available, best #1, table Default-IP-Routing-Table)
  Not advertised to any peer
  200
    192.168.1.2 from 192.168.1.2 (192.168.1.2)
```

```
      Origin IGP, metric 0, localpref 100, valid, external, best
      Community: 100:20 200:20
      Last update: Wed Nov  6 18:58:18 2013
```

After configuring the community list on device A, check the BGP routes to verify that the routes are filtered correctly.

```
Device A# display ip bgp-instance
BGP table version is 1, local router ID is 192.168.1.1
Status codes: s suppressed, d damped, h history, * valid, > best, i - internal,
                S Stale, b - backup entry
Origin codes: i - IGP, e - EGP, ? - incomplete

   Network          Next Hop            Metric     LocPrf     Weight Path
*> 10.0.0.0/24      192.168.1.2              0                    0 200 i
*> 20.0.0.0/24      192.168.1.2              0                    0 200 i

Total number of prefixes 2
```

6. Configuration Files

- Device A configuration file

```
!
ip comm-type-list standard-type test permit 100:20
!
route-map COM
 match-rule comm test
!
router-protocol bgp 100
 peer 192.168.1.2 route-map COM in
!
```

- Device B configuration file

```
!
route-map comm1
 set-rule comm 100:20 200:20
!
route-map comm2
 set-rule comm 100:20
!
route-map comm3
 setrule comm 200:20
!
router-protocol bgp 200
 peer 192.168.1.1 send-community
 network-advertise 10.0.0.0 mask 255.255.255.0 route-map comm1
```

```
 network-advertise 20.0.0.0 mask 255.255.255.0 route-map comm2
 network-advertise 30.0.0.0 mask 255.255.255.0 route-map comm3
 network-advertise 40.0.0.0 mask 255.255.255.0
!
```

Contents

1 Configuring PBR....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.2 Configuration Task Summary....4
1.3 Configuring Basic Functions of PBR....5
1.3.1 Overview....5
1.3.2 Restrictions and Guidelines....5
1.3.3 Configuration Tasks....10
1.3.4 Applying PBR to an Interface....11
1.3.5 Applying PBR to a Local Device....11
1.4 Configuring Redundant Backup or Load Balancing....12
1.4.1 Overview....12
1.4.2 Restrictions and Guidelines....12
1.4.3 Procedure....13
1.5 Configuring Source-Address-Based PBR....14
1.5.1 Overview....14
1.5.2 Restrictions and Guidelines....14
1.5.3 Procedure....14
1.6 Monitoring....15
1.7 Configuration Examples....16

1.7.1 Configuring PBR to Implement Multi-Exit High Availability 16

1.7.2 Selecting an Output Link based on Source Addresses of Packets 20

1 Configuring PBR

1.1 Introduction

1.1.1 Overview

You can apply a route map containing policies to interfaces and devices to implement Policy-Based Routing (PBR).

Similar to static routing, PBR is manually configured and does not automatically update with network changes. It is effective only for packets from local interfaces and devices. PBR is more flexible than static and dynamic routing. Static and dynamic routing forwards packets only based on destination addresses. PBR forwards packets based on the source IP address, destination IP address, packet length, and ingress port.

1.1.2 Principles

1. PBR-Procedure

To implement PBR, perform the following steps:

(1) Define a policy: Use the route map to define matching rules for traffic characteristics and set processing actions.

(2) Apply the policy: Apply the policy to interfaces or local devices.

2. Packet Processing Flow

After the device identifies packets, it processes packets as follows:

(1) Identify the source of traffic to determine the traffic is forwarded or is local, and check whether any policy is applied.

(2) Forward the packets based on the defined policy.

(3) If no next hop is matched or the next hop is unreachable, forward the packets based on the routing table.

(4) If the routing table does not contain any route to the destination address, discard the packets.

3. Definition of a Policy

A policy is a "**match-rule**..., **set-rule**..." statement, which indicates that "if certain conditions are matched, you can perform some processing actions"

Note

For details about the policies, see *Configuring Routing Policies*.

- Executing a policy

  A route map contains multiple policies. Each policy has a sequence number. A smaller sequence number means a higher priority. You can execute policies by sequence number. Once the matching condition of a policy

is met, the processing action of this policy is performed and the route map exits. If no matching condition of any policy is met, no processing is performed.

Policies support two working modes:

- permit: When the matching condition of a policy is met, perform the processing action of this policy and exit the route map.
- deny: When the matching condition of a policy is met, do not perform the processing action of this policy, but exit the route map.

- Matching conditions of policies

  The matching condition of a policy may contain zero, one, or more match rules.

  - If the matching condition contains 0 matching rule, no packet is matched.
  - If the matching condition contains one or more match rules, the matching condition is met only when all the rules are matched.

  In the route map mode, run the **match-rule** command to set matching rules. One **match-rule** command corresponds to one matching rule.

- Processing action for a policy

  The processing action of a policy contains 0, 1 or more set rules.

  - If the processing action contains 0 set rule, no processing action is performed and the route map directly exits.
  - If the processing action contains one or more set rules, all processing actions are performed and then the route map exits.

⚠ Caution

If set rules have different priorities, only the set rule with the highest priority takes effect.

4. Application of a Policy

Apply a route map including policies to interfaces or devices to implement PBR.

- If a route map is applied to an interface, packets received by the interface are routed based on the policy.

  The PBR is often used to control user packets received by a device. It is effective only for forwarded packets, but not for locally initiated packets.

- If a route map is applied to a device, locally initiated packets are routed based on the policy.

  The PBR is often used to control protocol packets exchanged between devices (such as ping packets sent locally). It is effective only for locally initiated packets, but not for forwarded packets.

5. Enhanced Reliability of PBR

- Redundant backup or load balancing

You can set multiple next hops in a policy. Either redundant backup or load balancing can be implemented among multiple next hops. Redundant backup is implemented by default.

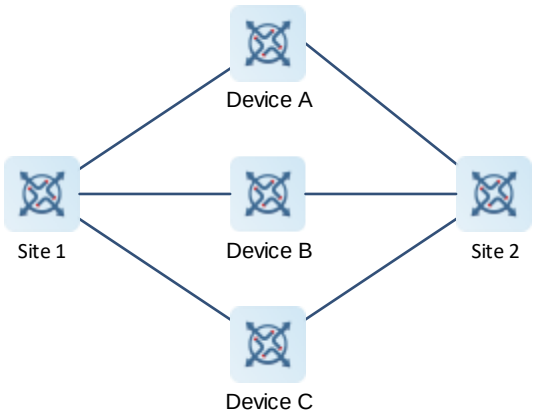
- Redundant backup

  As shown in Figure 1-1, the configuration sequence of the next hop from Site 1 to Site 2 is as follows: Device A > Device B > Device C, and the first accessible next hop takes effect. When the currently effective next hop (Device A) is faulty, the traffic automatically switches to the next accessible next hop (Device B). When Device A becomes accessible again, the traffic automatically switches back to Device A.

  A new next hop stays at the end of the sequence. If the original sequence is Device A > Device B > Device C, and if Device A is added after it is deleted, the current sequence becomes Device B > Device C > Device A.

  If no next hop is accessible, packets are discarded.

Figure 1-1 Redundant backup function



- Load balancing

  When multiple accessible next hops take effect concurrently, the Weighted Cost Multiple Path (WCMP) and Equal Cost Multiple Path (ECMP) are supported. After an accessible next hop fails, traffic will be balanced among the other accessible next hops.

- Correlation with BFD

  Correlation between PBR and BFD helps the PBR to better perceive network environment changes. When BFD finds the current next hop unaccessible, the BFD immediately notifies the PBR of switching the traffic to another accessible next hop (to implement redundant backup) or all the other accessible next hops (to implement load balancing).

- Correlation with Track

  Correlation between PBR and Track helps the PBR to better perceive network environment changes. When Track finds the current next hop unaccessible, the Track immediately notifies the PBR of switching the traffic to another accessible next hop (to implement redundant backup) or all the other accessible next hops (to implement load balancing).

6. PBR and VRF

If this feature is selected for VRF based on PBR, an interface to which PBR is applied can filter received IP packets by using the matching rules. If the packets are matched, the interface specifies a VRF instance for route selection in the set rules. The matching rules include the packet length and ACL (IP access list). Since the matching rules are flexible, you can allocate different traffic to different VRF instances based on actual requirements.

Generally, packets received on a VRF interface are forwarded from this VRF interface, and packets received on a global interface are forwarded based on a global routing table. PBR can break this limit and enable packets to be forwarded between a VRF instance and a global route map as follows:

- From a global routing table to a VRF instance: Packets received from a global interface jump to a specified VRF instance for forwarding.
- From a VRF instance to another VRF instance: Packets received from a VRF interface jump to another VRF interface for forwarding.
- From a VRF instance to a global routing table: Packets received from a VRF interface jump to the global routing table for forwarding.

7. Source-address-based PBR

If the source-address-based PBR is generated, you do not need to configure a route map.

- If only the source IPv4 address or source IPv6 address of packets forwarded by the interface needs to be matched, you can run the command of the source-address-based PBR. This command is more simply configured than that of interface-based PBR.
- If the source-address-based PBR is applied to a specified interface, packets received on this interface are routed according to the source addresses.

The PBR is often used to control user packets received by a device. It is effective only for forwarded packets, but not for locally initiated packets.

Note

The source-address-based PBR has a lower priority than the interface-based PBR. If the source-address-based PBR and the interface-based PBR are applied to the same interface, only the interface-based PBR takes effect.

8. PBR-based Traffic Diversion in Layer 3 Authentication

Divert traffic based on PBR for Layer 3 authentication.

1.2 Configuration Task Summary

The PBR configuration includes the following tasks:

(1) Configuring Basic Functions of PBR

- Applying PBR to an Interface

- Applying PBR to a Local Device

(2) (Optional) Configuring Redundant Backup or Load Balancing

(3) (Optional) Configuring Source-Address-Based PBR

1.3 Configuring Basic Functions of PBR

1.3.1 Overview

This section describes how to perform personalized routing management of user data streams by preparing flexible policies.

1.3.2 Restrictions and Guidelines

- A route map must be used when PBR is configured. Therefore, you must configure a route map on a device.
- If an ACL is used when the route map is configured, you must configure the ACL on the device.
- If the route map used in PBR does not exist, the PBR does not take effect.
- If PBR fails to be installed due to hardware resource insufficiency, run the **ip policy-sync all** or the **ipv6 policy-sync all** command to synchronize software entries to hardware when hardware resource is sufficient.
- Usage restrictions of PBR on matching rules in a route map:
  - IPv4 PBR uses IP standard or extended ACLs to define the source (and destination) IP address ranges of packets.
  - IPv6 PBR uses IPv6 extended IP ACLs to define the source (and destination) IPv6 address ranges of packets.
  - If PBR references an ACL but the ACL does not exist, this route submap is not used for packet matching, and packet matching continues from the next route submap. If a packet does not match any route submap, it is forwarded through the common route. If only an ACL is configured but no ACE is configured, PBR forwards packets just as it does when no ACL exists.
  - On switches, interface-based IPv4 PBR supports expert ACLs and MAC ACLs, while local PBR does not support expert ACLs and MAC ACLs.
  - On non-switches, both interface-based PBR and local PBR do not support expert ACLs and MAC ACLs.
  - On switches, if a route submap of PBR references multiple ACLs, only the first ACL is used for matching and subsequent ACLs are not used for matching.

Table 1-1 Matching Rules Supported by PBR

	Command	Description
IPv4 PBR	**match-rule ip address**	The source (and destination) IPv4 addresses are used as the matching condition. Multiple **match-rule ip address** commands are allowed in one policy.

IPv4 PBR	**match-rule length**	The match condition is the IPv4 packet length. One policy can be configured with only one **match-rule length** command.
IPv6 PBR	**match-rule ipv6 address**	The source (and destination) IPv6 addresses are used as the matching condition. Only one **match-rule ipv6 address** command is allowed in one policy.
IPv6 PBR	**match-rule ip policy**	The match conditions are the source IPv4 address (and destination IPv4 address) and the Layer 3 authentication traffic diversion domain type. One policy can be configured with only one **match-rule ip policy** command.

- Usage restrictions of PBR on set rules in a route map:
  - Priorities: **set-rule vrf > set-rule ip-address next-hop** > **set-rule ip-address next-hop recursive** > **set-rule interface** > common route > **set-rule ip-address default next-hop** > **set-rule default interface** > default route. The set commands can be configured at the same time but only the command with the highest priority takes effect.
  - Priorities: **set-rule vrf** > **set-rule ipv6-address next-hop** > **set-rule ipv6-address next-hop recursive** > common route > **set-rule ipv6-address default next-hop** > default route. The **set-rule** commands can be configured at the same time but only the command with the highest priority takes effect.
  - For switches, the **set-rule ipv6-address default next-hop** command does not take effect on IPv6 addresses with the mask length exceeding 64 bits.

Table 1-2 Set Rules Supported by PBR

	Command	Description
IPv4 PBR	**set-rule ip-address tos**	Changes the ToS field of IPv4 packets. This command cannot be configured together with the **set-rule ip-address dscp** command.
	set-rule ip-address precedence	Changes the precedence field of IPv4 packets. This command cannot be configured together with the **set-rule ip-address dscp** command.

	set-rule ip-address dscp	Changes the DSCP field of IPv4 packets. This command cannot be configured together with the **set-rule ip-address tos** and **set-rule ip-address precedence** commands.
	set-rule vrf	Sends IPv4 packets to a certain VRF for forwarding. The packets matching the Match rules are routed by the VRF specified in the **set-rule vrf** command, no matter whether the interface receiving the packets belongs to this VRF. This command cannot be used together with **set-rule interface** or **set-rule default interface** command**.**
	set-rule ip-address next-hop	Configures the next hop for IPv4 packet forwarding. The next hop must be directly connected. Otherwise, the command is invalid. Packets matching a matching rule are first forwarded to the next hop specified by the **set-rule ip-address next-hop** command regardless of whether the actual routing result of the packets in the routing table is the same as the next hop specified by PBR. On the switch, the outbound interface of the next hop supported by PBR can be an SVI, an L3 Ethernet interface, or an L3 aggregate interface.

	set-rule ip-address next-hop recursive	Configures the recursive next hop for IPv4 packet forwarding. The next hop can be directly connected or not. For routes with the next hop not directly connected, static or dynamic routes in the routing table are queried. This command supports recursion to multiple ECMP next hops of static or dynamic routes. A maximum of 32 next hops are supported. If a recursive route is a static recursive route, the next hop of only one static recursive route can be queried. The redundant backup or load balancing mode for recursively searching for multiple next hops is determined by the **ip policy-based** { **redundance** | **load-balance** } command. Packets matching a matching rule are first forwarded to the recursive next hop specified by the **set-rule ip-address next-hop recursive** command regardless of whether the actual routing result of the packets in the routing table is the same as the next hop specified by PBR. The recursive next hop of PBR takes effect only when a static or dynamic route with an egress and a next-hop IP address is queried.
	set-rule ip-address next-hop verify-accessibility	Verifies availability of the next-hop IPv4 address.
	set-rule interface	Specifies the outbound interface for IPv4 packet forwarding. The packets matching the Match rules are forwarded through the outbound interface specified in **set-rule interface** command, no matter whether the route of the packets specified in the routing table is the same as the next hop specified by PBR. This command cannot be used together with **set-rule vrf** command**.**
	set-rule ip-address default next-hop	Configures the default next hop for IPv4 packet forwarding. For IPv4 packets matching a matching rule, if the routing fails or a default route is selected in the routing table, the packets are forwarded to the default next hop specified by the command.

	set-rule ip-address default interface	Specifies the default outbound interface for IPv4 packets. If the routing of the packets matching the Match rules fails or a default route is selected, the packets are forwarded through the outbound interface specified in this command. This command cannot be used together with **set-rule vrf** command**.**
	set-rule ip-address policy l3-auth	Configures Layer 3 authentication for IPv4 packets. Layer 3 authentication will be performed for the packets matching the Match rules. This command is valid for the packets forwarded by interfaces, but invalid for the locally sent packets.
	set-rule ip-address policy load-balance	Configures the load balancing mode for IPv4 packets. If global load balancing mode is enabled, PBR selects an outbound interface for the packets matching the Match rules according to the configured load balancing mode This command is valid for the packets forwarded by interfaces, but invalid for the locally sent packets.
	set-rule ip-address policy no-ttl-decrease	Keeps the TTL of IPv4 packets not decreasing. The TTL in the headers of IPv4 packets that match the Match rules is not decreased by 1 during PBR-based forwarding. This command is applied to Layer 3 authentication traffic diversion. This command is valid for the packets forwarded by interfaces, but invalid for the locally sent packets.
IPv6 PBR	**set-rule ipv6-address precedence**	Changes the precedence field of IPv6 packets. IPv6 PBR does not support the **set-rule ipv6-address tos** and **set-rule ipv6-address dscp** commands.
	set-rule ipv6-address next-hop	Configures the next hop for IPv6 packet forwarding. IPv6 packets matching a matching rule are first forwarded to the next hop specified by the **set-rule ipv6-address next-hop** command regardless of whether the actual routing result of the IPv6 packets in the routing table is the same as the next hop specified by PBR. The next hop must be directly connected. Otherwise, the command is invalid.

	set-rule ipv6-address next-hop recursive	Configures the recursive next hop for IPv6 packet forwarding. The next hop can be directly connected or not. For routes with the next hop not directly connected, static or dynamic routes in the routing table are queried. This command supports recursion to multiple ECMP next hops of static or dynamic routes. A maximum of 32 next hops are supported. If a recursive route is a static recursive route, the next hop of only one static recursive route can be queried. The redundant backup or load balancing mode for recursively searching for multiple next hops is determined by the **ipv6 policy-based** { **redundance** | **load-balance** } command. Packets matching a matching rule are first forwarded to the recursive next hop specified by the **set-rule ipv6-address next-hop recursive** command regardless of whether the actual routing result of the packets in the routing table is the same as the next hop specified by PBR. The recursive next hop of PBR takes effect only when a static or dynamic route with an egress and a next-hop IPv6 address is queried.
	set-rule ipv6-address next-hop verify-availability	Verifies availability of the next-hop IPv6 address.
	set-rule ipv6-address reset next-hop	Configures the default next hop for IPv6 packet forwarding. For IPv6 packets matching a matching rule, if the routing fails or a default route is selected in the routing table, the packets are forwarded to the next hop specified by the command. The next hop must be directly connected. Otherwise, the command is invalid.

1.3.3 Configuration Tasks

The configuration of PBR basic functions includes the following tasks:

- Applying PBR to an Interface
- Applying PBR to a Local Device

1.3.4 Applying PBR to an Interface

1. Overview

This section describes how to apply PBR to an interface to implement personalized routing management on user data streams passing a device.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the loopback interface configuration mode.

  interface loopback *interface-number*

- Enter the VXLAN routed interface configuration mode.

  interface overlayrouter *interface-number*

(4) Apply PBR to IPv4 packets received by an interface.

ip policy-based route-map *route-map-name*

By default, no PBR is configured on an interface.

(5) Apply PBR to IPv6 packets received by an interface.

ipv6 policy-based route-map *route-map-name*

By default, no IPv6 PBR is configured on an interface.

1.3.5 Applying PBR to a Local Device

1. Overview

This section describes how to apply PBR to a local device to implement personalized management on packets initiated by the local device. A device has only one IPv4 PBR and one IPv6 PBR.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Apply PBR to IPv4 packets initiated by the local device.

ip local policy-based route-map *route-map-name*

By default, no IPv4 PBR is applied to the local device.

(4) Apply PBR to IPv6 packets initiated by the local device.

ipv6 local policy-based route-map *route-map-name*

By default, no IPv6 PBR is applied to the local device.

1.4 Configuring Redundant Backup or Load Balancing

1.4.1 Overview

Multiple next hops in the mutual backup mode can enhance the network reliability.

Load balancing mode among multiple next hops can expand the network bandwidth.

1.4.2 Restrictions and Guidelines

- If redundant backup is selected, the first next hop takes effect based on the configuration sequence.
- If load balancing is selected, all next hops take effect at the same time and share traffic by weight.
- This configuration applies to all PBRs configured on a device.
- The basic functions of PBR must be configured.
- Redundant backup and load balancing are effective only for the next hops set by the following **set-rule** commands, and only among multiple next hops set by the same **set-rule** command.

Table 1-3 set-rule commands for configuring the next hop

Command	Description
set-rule ip-address next-hop	Configures the next hop of IPv4 packets. This command carries the **weight** parameter to set the weight of the WCMP. The default value is **1**.
set-rule ip-address default next-hop	Configures the default next hop of IPv4 packets. This command carries the **weight** parameter to set the weight of the WCMP. The default value is 1.

Command	Description
set-rule ipv6-address next-hop	Configures the next hop of IPv6 packets. This command carries the **weight** parameter to set the weight of the WCMP. The default value is **1**.
set-rule ipv6-address default next-hop	Configures the default next hop of IPv6 packets. This command carries the **weight** parameter to set the weight of the WCMP. The default value is **1**.
set-rule ip-address next-hop recursive	Configures the recursive next hop of IPv4 packets. Only one command is configured for a route map, and packets can recur to multiple next hops (up to 32 next hops) of a static or dynamic ECMP route. The redundant backup or load balancing mode for recurring to multiple next hops is also determined by the **ip policy** { **redundance** | **load-balance** } command.
set-rule ipv6-address next-hop recursive	Configures the recursive next hop of IPv6 packets. Only one command is configured for a route map, and packets can recur to multiple next hops (up to 32 next hops) of a static or dynamic ECMP route. The redundant backup or load balancing mode for recurring to multiple next hops is also determined by the **ipv6 policy** { **redundance** | **load-balance** } command.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an IPv4 PBR to implement redundant backup or load balancing among multiple next hops.

ip policy { **redundance** | **load-balance** }

By default, multiple next hops of the PBR adopt the redundant backup mode.

(4) Configure an **IPv6** PBR to implement redundant backup or load balancing among multiple next hops.

ipv6 policy { **redundance** | **load-balance** }

By default, multiple next hops of the IPv6 PBR adopt the redundant backup mode.

1.5 Configuring Source-Address-Based PBR

1.5.1 Overview

This section describes how to perform personalized routing management for IPv4 or IPv6 addresses of user data streams by preparing flexible policies.

1.5.2 Restrictions and Guidelines

- Source-address-based PBR has a higher priority than interface-based PBR. When they are applied to an interface at the same time, interface-based PBR takes effect whereas source-address-based PBR does not take effect.
- The global configuration takes effect on the input interface of specified user data streams.
- You can specify multiple source-address-based PBRs for the same interface. The sequence numbers of source-address-based PBRs must be different. A smaller sequence number indicates a higher priority of the source-address-based PBR.

  If the sequence number is the same, the priorities of next hop types are as follows:

  vrf *vrf-name* **next-hop** *ipv4-address* > **interface** *out-interface-type* > **default next-hop** *ipv4-address*> **default interface** *out-interface-type*

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure applying source address policy routing for IPv4 packets received by the interface.

ip policy-source in-interface *interface-type interface-number sequence* { *source-address mask* | *source-address/mask* } { [**default**] **next-hop** { *ipv4-address* [*weight*] }&<1-4> | [**default**] **interface** { *out-interface-type out-interface-number* } &<1-4>| **vrf** *vrf-name* }

By default, no source-address-based PBR takes effect on an interface.

(4) Apply a source-address-based PBR for IPv6 packets received by an interface.

ipv6 policy-source in-interface *interface-type interface-number sequence source-address/prefix-length* { [**default**] **next-hop** [*ipv6-address* [*weight*]]&<1-4> [**default**] **interface** *out-interface-type*&<1-4>| **vrf** *vrf-name* }

By default, no source-address-based IPv6 PBR takes effect on an interface.

1.6 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

You can run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

You can run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-4 Monitoring of the PBR

Command	Purpose
display ip policy-based [*route-map-name*]	Displays the configurations of the IPv4 PBR.
display ipv6 policy-based	Displays the configurations of the IPv6 PBR.
display route-map [*name*]	Displays the configurations of a route map.
display access-control-list	Displays the configurations of an ACL.
display ip pbr bfd	Displays the correlation between the IPv4 PBR and BFD.
display ipv6 pbr bfd	Displays the correlation between the IPv6 PBR and BFD.
display ip pbr recursive	Displays routing information of IPv4 PBR after next-hop recursion.
display ipv6 pbr recursive	Displays routing information of IPv6 PBR after next-hop recursion.
display ip pbr route [**interface** *interface-type interface-number* | **local**]	Displays the routing information of the IPv4 PBR.
display ipv6 pbr route [**interface** *interface-type interface-number* | **local**]	Displays the routing information of the IPv6 PBR.
display ip pbr route-map *rmap-name*	Displays a route map used by the IPv4 PBR.

Command	Purpose
display ipv6 pbr route-map *rmap-name*	Displays a route map used by the IPv6 PBR.
display ip pbr source-routing [**interface** *interface-type interface-number*]	Displays the routing information of source-address-based IPv4 PBR.
display ipv6 pbr source-routing [**interface** *interface-type interface-number*]	Displays the routing information of source-address-based IPv6 PBR.
display ip pbr track	Displays the correlation between PBR and tracking.
debugging pbr error	Debugs PBR errors.
debugging pbr events	Debugs PBR events.
debugging pbr ms	Debugs multiple service cards supported by PBR.
debugging pbr msg	Debugs PBR message communication.
debugging pbr nsm	Debugs interaction between PBR and NSM.
debugging pbr packet	Debugs packet forwarding of PBR.
debugging pbr restart	Debugs PBR GR.

1.7 Configuration Examples

1.7.1 Configuring PBR to Implement Multi-Egress High Availability

1. Requirements

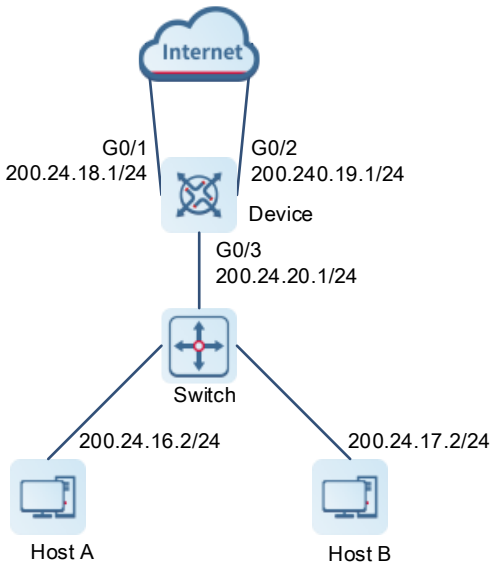
Configure PBR on the devices to achieve the following purposes:

- Packets from Host A (200.24.16.2/24) are sent from GigabitEthernet 0/1 first.
- Packets from Host B (200.24.17.2/24) are sent from GigabitEthernet 0/2 first.

The two output links work in mutual backup mode, and are correlated with Track. Thus, when one link is faulty, traffic is quickly switched over to the other link.

2. Topology

Figure 1-2 Configuring PBR to Implement Multi-Egress High Availability



3. Notes

- Configure two different ACLs to match packets in the Host A and Host B network segment respectively.
- Configure the route map RM_FOR_PBR: policy 10 is used to ensure that "packets in the Host A network segment are sent from GigabitEthernet 0/1first"; policy 20 is used to ensure that "packets in the Host B network segment are sent from GigabitEthernet 0/2 first".
- Configure a PBR to implement redundant backup among multiple next hops.
- Configure the track rules and correlation between PBR and Track.
- Configure PBR and apply the route map RM_FOR_PBR to packets received from GigabitEthernet 0/3.

4. Procedure

(1) Configure the IP addresses of interfaces to implement interworking between networks. (Omitted)

(2) Configure ACLs for subnets 200.24.16.0/24 and 200.24.17.0/24, respectively.

```
Device> privilege
Device# config-system terminal
Device(cfg)# access-control-list 1 permit 200.24.16.0 0.0.0.255
Device(cfg)# access-control-list 2 permit 200.24.17.0 0.0.0.255
```

(3) Configure the track rules.

```
Device(cfg)# ip rns 1
Device(cfg-ip-rns)# icmp-echo 200.24.18.1
```

```
Device(cfg-ip-rns)# exit
Device(cfg)# ip rns schedule 1 start-time now life forever
Device(cfg)# track 1 rns 1
Device(cfg)# ip rns 2
Device(cfg-ip-rns)# icmp-echo 200.24.19.1
Device(cfg-ip-rns)# exit
Device(cfg)# ip rns schedule 2 start-time now life forever
Device(cfg)# track 2 rns 2
```

(4) Configure the PBR **RM_FOR_PBR**, set the next hop of 200.24.16.0 to **200.24.18.1** and its backup next hop to **200.24.19.1**, and set the next hop of 200.24.17.0 to **200.24.19.1** and its backup next hop to **200.24.18.1**.

```
Device(cfg)# route-map RM_FOR_PBR 10
Device(cfg-routemap)# match-rule ip address 1
Device(cfg-routemap)# set-rule ip-address next-hop verify-accessibility 200.24.18.1 track 1
Device(cfg-routemap)# set-rule ip-address next-hop verify-accessibility 200.24.19.1 track 2
Device(cfg-routemap)# exit
Device(cfg)# route-map RM_FOR_PBR 20
Device(cfg-routemap)# match ip address 2
Device(cfg-routemap)# set-rule ip-address next-hop verify-accessibility 200.24.19.1 track 2
Device(cfg-routemap)# set-rule ip-address next-hop verify-accessibility 200.24.18.1 track 1
Device(cfg-routemap)# exit
```

(5) Apply the route map RM_FOR_PBR to GigabitEthernet 0/3, and enable load balancing.

```
Device(cfg)# interface GigabitEthernet 0/3
Device(cfg-if-GigabitEthernet 0/3)# ip policy-based route-map RM_FOR_PBR
Device(cfg-if-GigabitEthernet 0/3)# exit
Device(cfg)# ip policy redundance
```

5. Verification

Check the PBR configurations to verify that the route map is applied to the interface.

```
Device# display ip policy-based
Balance mode: redundance
Interface                    Route map                    state
local                        test                         ok
GigabitEthernet 0/3          test                         ok
```

Check the configurations of the route map to verify the policy rules.

```
Device# display route-map
route-map RM_FOR_PBR, permit, sequence 10
  Match clauses:
    ip address 1
  Set clauses:
    ip next-hop 200.24.18.1
```

```
    ip next-hop 200.24.19.1
route-map RM_FOR_PBR, permit, sequence 20
  Match clauses:
    ip address 2
  Set clauses:
    ip next-hop 200.24.19.1
    ip next-hop 200.24.18.1
```

Check the ACL configurations to verify the packet filtering rules.

```
Device# display access-lists
ip access-control-list standard 1
 10 permit 200.24.16.0 0.0.0.255
 10 permit 200.24.16.0 0.0.0.255
ip access-control-list standard 2
 10 permit 200.24.17.0 0.0.0.255
```

Check whether the track objects are up.

```
Device#display track
Track 1
  Reliable Network Service 1
  The state is Up
    1 change, current state last: 120 secs
  Delay up 30 secs, down 50 secs
Track 2
  Reliable Network Service 2
  The state is Up
    1 change, current state last: 130 secs
  Delay up 30 secs, down 50 secs
```

6. Configuration Files

Device configuration file

```
!
ip policy redundance
!
access-control-list 1 permit 200.24.16.0 0.0.0.255
access-control-list 2 permit 200.24.17.0 0.0.0.255
!
route-map RM_FOR_PBR 10
 match-rule ip address 1
 set-rule ip-address next-hop verify-accessibility 200.24.18.1 track 1
 set-rule ip-address next-hop verify-accessibility 200.24.19.1 track 2
!
route-map RM_FOR_PBR 20
```

```
 match-rule ip address 2
 set-rule ip-address next-hop verify-accessibility 200.24.19.1 track 2
 set-rule ip-address next-hop verify-accessibility 200.24.18.1 track 1
!
ip rns 1
 icmp-echo 200.24.18.1
!
ip rns 2
 icmp-echo 200.24.19.1
!
ip rns schedule 1 start-time now life forever
ip rns schedule 2 start-time now life forever
!
track 1 rns 1
!
track 2 rns 2
!
interface GigabitEthernet 0/3
 ip policy-based route-map RM_FOR_PBR
!
```

7. Common Errors

- After matching of ACLs and prefix-lists is configured, the corresponding ACLs and prefix lists are not defined.

1.7.2 Selecting an Output Link based on Source Addresses of Packets

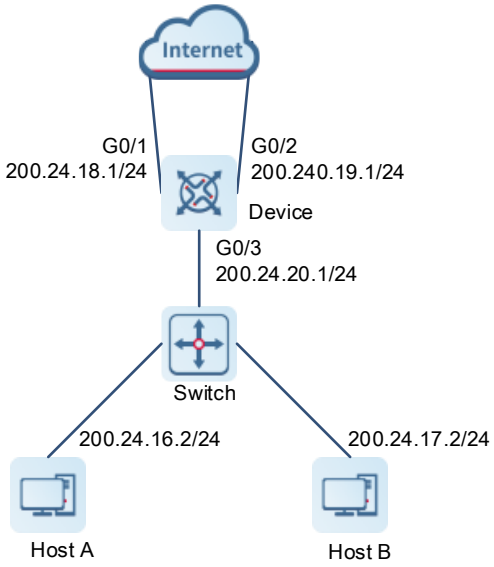
1. Requirements

This LAN has two output interfaces for connecting the Internet. The requirements are as follows:

- Route the Internet access data streams from Host A to GigabitEthernet 0/1.
- Route the Internet access data streams from Host B to GigabitEthernet 0/2.
- If the link on GigabitEthernet 0/1 is disconnected, switch over the data streams from GigabitEthernet 0/1 to GigabitEthernet 0/2. Vice versa.

2. Topology

Figure 1-3 Selecting an Output Link based on Source Addresses of Packets



3. Notes

- Configure a source-address-based PBR, and set the next hops for packets from the GigabitEthernet 0/3 of Host A to **GigabitEthernet 0/1** and **GigabitEthernet 0/2**. (Note the configuration sequence.)
- Configure a source-address-based PBR, and set the next hops for packets from the GigabitEthernet 0/3 of Host B to **GigabitEthernet 0/2** and **GigabitEthernet 0/1**. (Note the configuration sequence.)
- Configure a PBR to implement redundant backup among multiple next hops. (The default setting is redundant backup.)

4. Procedure

(1) Configure basic networks, such as the interface IP address.

(2) Select an output link based on source addresses of packets.

```
Device> privilege
Device# config-system terminal
Device (cfg)# ip policy-source in-interface gigabitEthernet 0/3 1 200.24.16.0/24 next-hop 200.24.18.1
200.24.19.1
Device (cfg)# ip policy-source in-interface gigabitEthernet 0/3 2 200.24.17.0/24 next-hop 200.24.19.1
200.24.18.1
Device (cfg)# ip policy redundance
```

5. Verification

Check the configurations of the source-address-based IPv4 PBR.

```
Device # display ip pbr source-routing
PBR IPv4 Source Route
Interface              : GigabitEthernet 0/3
   Sequence            : 1
   Source address : 200.24.16.0/24
   VRF ID              : 0
   Route Flags       :
     Route Type       : PBR
     Direct             : Permit
     Priority          : High
     Match_ipaddr    : Exist
   Mode                  : redundance
   Nexthop Count    : 2
     Nexthop[0]       : 200.24.18.1
     Weight[0]        : 1
     Ifindex[0]       : 1
     Nexthop[1]       : 200.24.19.1
     Weight[1]        : 1
     Ifindex[1]       : 2

Interface              : GigabitEthernet 0/3
   Sequence            : 2
   Source address : 200.24.17.0/24
   VRF ID              : 0
   Route Flags       :
     Route Type       : PBR
     Direct             : Permit
     Priority          : High
     Match_ipaddr    : Exist
   Mode                  : redundance
   Nexthop Count    : 2
     Nexthop[0]       : 200.24.19.1
     Weight[0]        : 1
     Ifindex[0]       : 2
     Nexthop[1]       : 200.24.18.1
     Weight[1]        : 1
     Ifindex[1]       : 1
```

6. Configuration Files

Device configuration file

```
!
ip policy-source in-interface gigabitEthernet 0/3 1 200.24.16.0/24 next-hop 200.24.18.1 200.24.19.1
ip policy-source in-interface gigabitEthernet 0/3 2 200.24.17.0/24 next-hop 200.24.19.1 200.24.18.1
ip policy redundance
!
```

Contents

1 Configuring Keys 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuring a Key Chain 1
1.2.1 Overview 1
1.2.2 Restrictions and Guidelines 1
1.2.3 Configuration Tasks 2
1.2.4 Creating a Key Chain 2
1.2.5 Configuring a Key ID 2
1.2.6 Configuring a Key String 3
1.2.7 Configuring the Send-Lifetime 4
1.2.8 Configuring the Accept-Lifetime 4
1.3 Monitoring 5
1.4 Configuration Examples 5
1.4.1 Configuring a Key Chain and Using the Key Chain in RIP Packet Authentication 5

1 Configuring Keys

1.1 Introduction

1.1.1 Overview

Keys are a kind of parameters that are used in algorithms for conversion from plaintext to ciphertext or from ciphertext to plaintext.

To allow packet authentication in a routing protocol to support plaintext authentication and ciphertext authentication, you must use keys.

Note

- Keys are used only for RIP, OSPF, and ISIS packet authentication.
- In this chapter, the term "router" refers to any network device that supports the routing function. These network devices can be Layer-3 switches, routers, or firewalls.

1.1.2 Principles

The key chain provides a tool for the authentication mechanism in a routing protocol.

1. Key Chain

A key chain contains the TCP authentication configuration, authentication algorithm ID, and multiple different keys. Each key has the following attributes:

- Key ID: Identifies a key. In the current key chain, each key corresponds to one ID.
- Authentication string: Specifies a set of key characters used for verifying the consistency of authentication strings in a routing protocol.
- Lifetime: Specifies the lifetime of the current key for sending or receiving packets. Different authentication keys can be used in different periods.
- Authentication algorithm: It is used by keys to authenticate routing protocols.

1.2 Configuring a Key Chain

1.2.1 Overview

This section defines a key chain to be used by a routing protocol.

1.2.2 Restrictions and Guidelines

A key chain can take effect only after it is associated with a routing protocol.

1.2.3 Configuration Tasks

The key configuration includes the following tasks:

(1) Creating a Key Chain

(2) Configuring a Key ID

(3) Configuring a Key String

(4) (Optional) Configuring the Send-Lifetime

(5) (Optional) Configuring the Accept-Lifetime

1.2.4 Creating a Key Chain

1. Overview

In the global configuration mode, you can run the **key chain** *key-chain-name* command to define a key chain and enter the key chain configuration mode.

2. Restrictions and Guidelines

- This configuration is mandatory if a key chain needs to be used.
- If there is no special requirement, you should perform this configuration on all routers that need to authenticate routing protocols.
- To make a key chain take effect, you must configure at least one key.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a key chain, and enter the key chain configuration mode.

key chain *key-chain-name*

No key chain is configured by default.

1.2.5 Configuring a Key ID

1. Overview

In the key chain configuration mode, you can run the **key** *key-id* command to define a key and enter the key chain key configuration mode.

2. Restrictions and Guidelines

- This configuration is mandatory if a key chain needs to be used.

- If there is no special requirement, you should perform this configuration on all devices that need to authenticate routing protocols.
- After the key is configured, it must meet two conditions before it takes effect: (1) the key-string is configured; (2) the key is in the lifetime (send-lifetime and accept-lifetime). If the lifetime is not configured, the key is considered effective permanently once the key-string is configured. Two effective states are defined: effective on the sending end and effective on the receiving end. The two states correspond to send-lifetime and accept-lifetime respectively.
- If there is no special demand, you can configure a key by incrementing the key-id to avoid the authentication state oscillation caused by the possible changes of effective keys. If multiple effective keys exist, each routing protocol uses the key with the smallest key-id.
- In the TCP enhanced authentication scenario, the key-id ranges from 0 to 63.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the key chain configuration mode.

key chain *key-chain-name*

(4) Create a key ID, and enter the key configuration mode of the key chain.

key *key-id*

No key is configured by default.

1.2.6 Configuring a Key String

1. Overview

In the key configuration mode of the key chain, you can specify a key string in plaintext or cyphertext. To ensure security, you are advised to use cyphertext. If you run the **service password-encryption-storage** command to enable the encryption service, you can forcibly convert a plaintext key string into a ciphertext string.

2. Restrictions and Guidelines

- This configuration is mandatory if a key chain needs to be used.
- If there is no special requirement, you should perform this configuration on all routers that need to authenticate routing protocols.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the key chain configuration mode.

key chain *key-chain-name*

(4) Enter the key configuration mode of the key chain.

key *key-id*

(5) Configure a key string.

key-string [**0 | 7**] *string-text*

No key string is configured by default.

1.2.7 Configuring the Send-Lifetime

1. Overview

This section defines the lifetime of a key in the sending direction.

2. Restrictions and Guidelines

If the send-lifetime is not configured, the key chain is always effective.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the key chain configuration mode.

key chain *key-chain-name*

(4) Enter the key configuration mode of the key chain.

key *key-id*

(5) Configure the send-lifetime.

accept-lifetime *start-time* { *end-time* | **duration** *duration-time* | **infinite** }

By default, the accept-lifetime of a key chain is disabled.

1.2.8 Configuring the Accept-Lifetime

1. Overview

This section defines the lifetime of a key in the receiving direction.

2. Restrictions and Guidelines

If the accept-lifetime is not configured, the key chain is always effective.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the key chain configuration mode.

key chain *key-chain-name*

(4) Enter the key configuration mode of the key chain.

key *key-id*

(5) Configure the accept-lifetime.

send-lifetime *start-time* { *end-time* | **duration** *duration-time* | **infinite** }

By default, the send-lifetime of a key chain is disabled.

1.3 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-1 Key Monitoring

Command	Purpose
display key chain config [*key-chain-name*]	Displays the configurations of a key chain.

1.4 Configuration Examples

1.4.1 Configuring Key-based Authentication

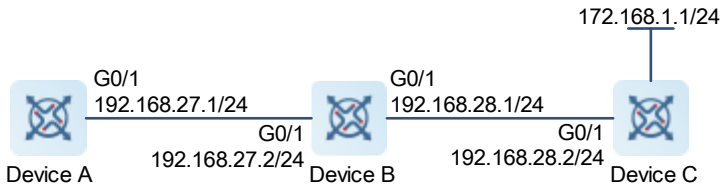
1. Requirements

Device A and Device B are interconnected using the RIP routing protocol. To improve security, you can configure the MD5 encryption algorithm between Device A and Device B to encrypt RIP packets.

Device B accesses 172.168.1.1/24 via the static route.

2. Topology

Figure 1-1 Configuring Key-based Authentication



3. Notes

- Configure keys on all routers.
- Configure RIP on all routers.
- Enable RIP authentication on all routers.

4. Procedure

(1) Configure the IP addresses of interfaces.

Configure Device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitEthernet 0/1
Device A(cfg-if)# ip address 192.168.27.1 255.255.255.0
Device A(cfg-if)# exit
```

Configure Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitEthernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.27.2 255.255.255.0
Device B(cfg-if-GigabitEthernet 0/1)# exit
```

(2) Configure keys and lifetime of each key.

Configure Device A.

```
Device A# config-system terminal
Device A(cfg)# key chain ripchain
Device A(cfg-keychain)# key 1
Device A(cfg-keychain-key)# key-string Hello
Device A(cfg-keychain-key)# accept-lifetime 16:30:00 Oct 1 2013 duration 43200
Device A(cfg-keychain-key)# send-lifetime 16:30:00 Oct 1 2013 duration 43200
Device A(cfg-keychain-key)# exit
Device A(cfg-keychain)# key 2
```

```
Device A(cfg-keychain-key)# key-string World
Device A(cfg-keychain-key)# accept-lifetime 04:00:00 Oct 2 2013 infinite
Device A(cfg-keychain-key)# send-lifetime 04:00:00 Oct 2 2013 infinite
Device A(cfg-keychain-key)# exit
```

Configure Device B.

```
Device B# config-system terminal
Device B(cfg)# key chain ripchain
Device B(cfg-keychain)# key 1
Device B(cfg-keychain-key)# key-string Hello
Device B(cfg-keychain-key)# accept-lifetime 16:30:00 Oct 1 2013 duration 43200
Device B(cfg-keychain-key)# send-lifetime 16:30:00 Oct 1 2013 duration 43200
Device B(cfg-keychain-key)# exit
Device B(cfg-keychain)# key 2
Device B(cfg-keychain-key)# key-string World
Device B(cfg-keychain-key)# accept-lifetime 04:00:00 Oct 2 2013 infinite
Device B(cfg-keychain-key)# send-lifetime 04:00:00 Oct 2 2013 infinite
Device B(cfg-keychain-key)# exit
```

(3) Configure the basic RIP.

Configure Device A.

```
Device A(cfg)# router-protocol rip
Device A(cfg-router)# version 2
Device A(cfg-router)# network-advertise 192.168.27.0
Device A(cfg-router)# exit
```

Configure Device B.

```
Device B(cfg)# router-protocol rip
Device B(cfg-router)# version 2
Device B(cfg-router)# network-advertise 192.168.27.0
Device B(cfg-router)# redistribute-route static
Device B(cfg-router)# exit
```

(4) Set the interface authentication mode to MD5, and specify the authentication key chain.

Configure Device A.

```
Device A(cfg)# interface gigabitEthernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip rip enable-auth keychain ripchain
Device A(cfg-if)# ip rip enable-auth mode md5
```

Configure Device B.

```
Device B(cfg)# interface gigabitEthernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.27.2 255.255.255.0 Device B(cfg-if-GigabitEthernet
0/1)# ip rip enable-auth keychain ripchain
Device B(cfg-if-GigabitEthernet 0/1)# ip rip enable-auth mode md5
```

5. Verification

Check whether Device A can receive routes from Device B.

```
Device A# display ip route-config rip
R       172.168.0.0/16 [120/1] via 192.168.27.2, 00:05:16, GigabitEthernet 0/1
```

6. Configuration Files

- Device A configuration file

```
!
key chain ripchain
 key 1
  key-string Hello
  accept-lifetime 16:30:00 Oct 1 2013 duration 43200
  send-lifetime 16:30:00 Oct 1 2013 duration 43200
 key 2
  key-string World
  accept-lifetime 04:00:00 Oct 2 2013 infinite
  send-lifetime 04:00:00 Oct 2 2013 infinite
!
interface gigabitEthernet 0/1
 ip address 192.168.27.1 255.255.255.0
 ip rip enable-auth keychain ripchain
 ip rip enable-auth mode md5
!
router-protocol rip
 version 2
 network-advertise 192.168.27.0
!
```

- Device B configuration file

```
!
key chain ripchain
 key 1
  key-string Hello
  accept-lifetime 16:30:00 Oct 1 2013 duration 43200
  send-lifetime 16:30:00 Oct 1 2013 duration 43200
key 2
  key-string World
  accept-lifetime 04:00:00 Oct 2 2013 infinite
  send-lifetime 04:00:00 Oct 2 2013 infinite
!
interface gigabitEthernet 0/1
 ip address 192.168.27.2 255.255.255.0
```

```
 ip rip enable-auth keychain ripchain
 ip rip enable-auth mode md5
!
router-protocol rip
 version 2
 network-advertise 192.168.27.0
 redistribute-route static
!
```

7. Common Errors

- A key is not correctly associated with a routing protocol, resulting in authentication failures.
- The keys configured on multiple routers are inconsistent, resulting in authentication failures.

Multicast Configuration

1. Configuring Multicast Basics
2. Configuring IPv4 Multicast Route Management
3. Configuring IGMP
4. Configuring PIM-SM
5. Configuring PIM-DM
6. Configuring IGMP Snooping
7. Configuring MSDP
8. Configuring IPv6 Multicast Route Management
9. Configuring MLD
10. Configuring PIM-SMv6
11. Configuring MLD Snooping

Contents

1 Configuring Multicast Basics 1
1.1 Basic Concepts of Multicast 1
1.2 Multicast Service Models 2
1.2.1 ASM Model 2
1.2.2 SSM Model 2
1.3 Multicast Address 2
1.3.1 IPv4 Multicast Address 2
1.3.2 IPv4 Multicast MAC Address 3
1.3.3 IPv6 Multicast Address 4
1.3.4 IPv6 Multicast MAC Address 5
1.4 IPv4 Multicast Protocols 6
1.5 IPv6 Multicast Protocols 6
1.6 Basic Concepts of PIM 7

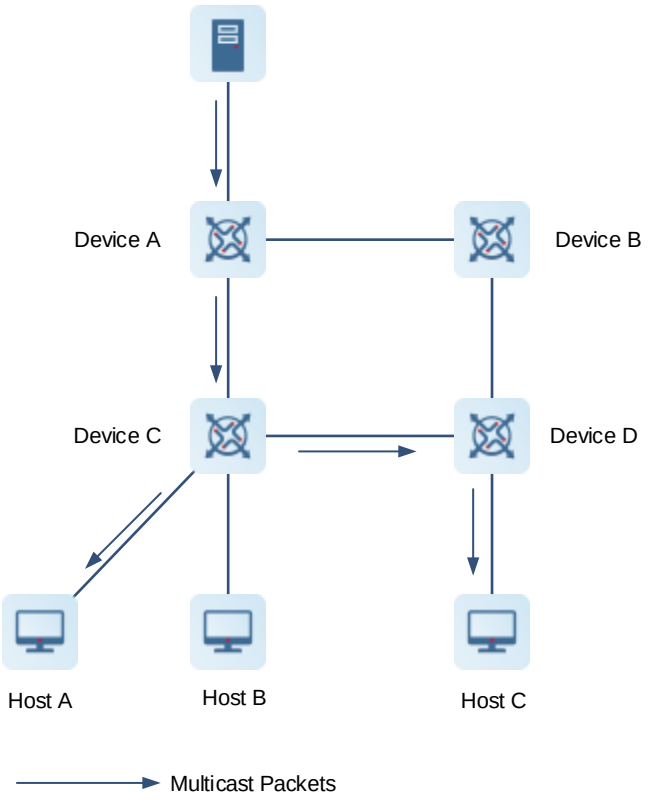
1 Configuring Multicast Basics

IP transmission methods are categorized into unicast, multicast, and broadcast. In IP multicast, an IP packet is sent from a source and forwarded to a specific group of receivers. Compared with unicast and broadcast, IP multicast saves bandwidth and reduces network loads. Therefore, IP multicast is applied to different network services that have high requirements for real timeliness, for example, Internet TV, distance education, live broadcast and multimedia conference.

1.1 Basic Concepts of Multicast

As shown in Figure 1-1, there are packet sending sources and receivers on a network. Users who are interested in information raise information requirements. Then, the information sending source sends information only to these users.

Figure 1-1 Topology of Multicast Data Transmission



- Multicast group: It is marked based on a multicast IP address. After a host is added to a multicast group, the host becomes a member of the multicast group and can receive multicast packets destined to the multicast group.
- Multicast source: Indicates a packet sender, for example, the source in Figure 1-1. Different multicast sources can send packets to a multicast group, and a multicast source can send multicast packets to different multicast groups.

- Member of multicast group: Indicates a host that is added to a multicast group, for example, Host A and Host C in Figure 1-1. Members in a multicast group can be dynamically added to or removed from the multicast group.
- Multicast device: the device that supports the multicast function. For example, the devices in Figure 1-1 can provide multicast routing and member management functions.

1.2 Multicast Service Models

Based on members in a multicast group, the service models of multicast can be categorized into Any-Source Multicast (ASM) and Source-Specific Multicast (SSM). The two service models use different address ranges of multicast groups.

1.2.1 ASM Model

In the ASM model, multicast data is assigned to multicast groups. The address of a multicast group is used as a network service ID. Any multicast source can send packets to this multicast group, and hosts join the group to receive multicast packets from this multicast group. The address of the multicast group in the ASM model is unique on the multicast network.

1.2.2 SSM Model

In the SSM model, multicast data is assigned to specific multicast sources and multicast groups. When you add a host to a multicast group, you can configure it to receive or not receive data from the specified multicast sources After the host is added to the multicast group, the host can receive only multicast data that is sent to this multicast group by the specified multicast sources. In the SSM model, different multicast sources can use the same address of a multicast group.

1.3 Multicast Address

The multicast technology provides network layer multicast for communication between a multicast source and members in a multicast group based on a multicast IP address. The multicast technology ensures transmission of multicast data in a local physical network based on a multicast MAC address. The IP address must be mapped to the MAC address to ensure correct transmission of the multicast data, because the members in the multicast group are unknown.

1.3.1 IPv4 Multicast Address

The four higher-order bits of a class D IP address are 1110, and the IP address range is from 224.0.0.0 to 239.255.255.255. These four bits are assigned by the Internet Assigned Numbers Authority (IANA) to IPv4 multicast. Table 1-1 lists how IPv4 multicast addresses are allocated.

Table 1-1 Assignment of IPv4 Multicast Addresses

Address Range	Description
224.0.0.0~224.0.0.255	Address of a permanent group
224.0.1.0~231.255.255.255 233.0.0.0~238.255.255.255	ASM multicast address

Address Range	Description
232.0.0.0~232.255.255.255	Default SSM multicast address
239.0.0.0~239.255.255.255	Address of a local management group

Table 1-2 Addresses of Common Permanent Groups

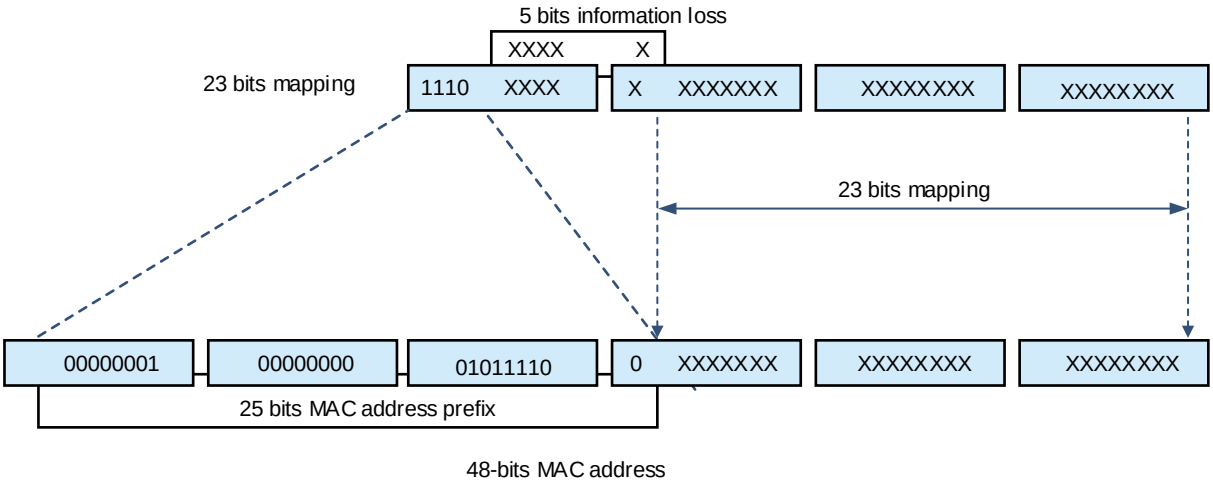
Address of a Permanent Group	Description
224.0.0.0	Unallocated
224.0.0.1	All hosts and routers in a network segment
224.0.0.2	All multicast devices
224.0.0.3	Unallocated
224.0.0.4	Distance Vector Multicast Routing Protocol (DVMRP) router
224.0.0.5	Open Shortest Path First (OSPF) router
224.0.0.6	OSPF specified router
224.0.0.7	Shared tree (ST) router
224.0.0.8	ST host
224.0.0.9	Routing Information Protocol version 2 (RIP-2) router
224.0.0.11	Mobile proxy
224.0.0.12	Dynamic Host Configuration Protocol (DHCP) server/relay proxy
224.0.0.13	All Protocol Independent Multicast (PIM) devices
224.0.0.14	Resource Reservation Protocol (RRP) encapsulation
224.0.0.15	All Core-Based Tree (CBT) routers
224.0.0.16	Specified Subnetwork Bandwidth Management (SBM)
224.0.0.17	All SBMs
224.0.0.18	Virtual Router Redundancy Protocol (VRRP)

1.3.2 IPv4 Multicast MAC Address

Based on IANA regulation, the higher-order 24 bits of an IPv4 multicast MAC address is 0x01005e, the 25th bit is 0, and the lower-order 23 bits are those of the IPv4 multicast address. Figure 1-2 shows the mapping. The four higher-order bits of an IPv4 multicast address is fixed at 1110, corresponding to the higher-order 25 bits of the multicast MAC address. The lower-order 23 bits in the lower-order 28 bits of the IPv4 multicast address map

to the lower-order 23 bits of the multicast MAC address. The lost information indicating by five bits causes the 32 IPv4 addresses to map to the same MAC address.

Figure 1-2 Mapping of IPv4 Multicast Address to IPv4 Multicast MAC Address



1.3.3 IPv6 Multicast Address

An IPv6 address consists of 128 bits. An IPv6 multicast address always starts with FF.

Table 1-3 Assignment of IPv6 Multicast Addresses

Address Range	Description
FF0x::/32	Addresses of a reserved group
FF1x::/32 (x cannot be 1 or 2) FF2x::/32 (x cannot be 1 or 2)	ASM multicast address
FF3x::/32 (x cannot be 1 or 2)	Default SSM multicast address

Table 1-4 Common IPv6 Multicast Addresses and Description

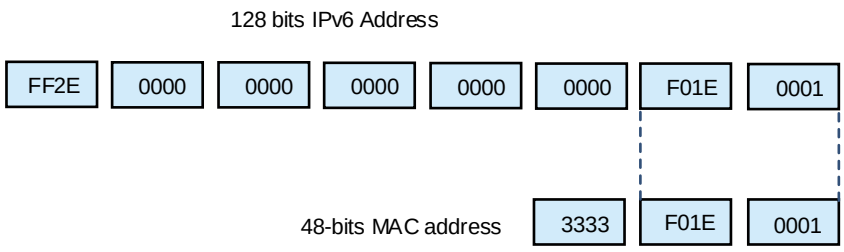
Range	IPv6 Multicast Address	Description
Local range of a node (or an interface)	FF01::1	Addresses of all nodes (interfaces)
	FF01::2	Addresses of all routers
Local range of a link	FF02::1	Addresses of all nodes
	FF02::2	Addresses of all routers
	FF02::3	Unallocated
	FF02::4	DVMRP router

Range	IPv6 Multicast Address	Description
	FF02::5	OSPF IGP router
	FF02::6	OSPF IGP DR
	FF02::7	ST router
	FF02::8	ST host
	FF02::9	RIP router
	FF02::A	EIGRP router
	FF02::B	Mobile proxy
	FF02::D	All PIM routers
	FF02::E	RSVP encapsulation
	FF02::1:1	Link name
	FF02::1:2	All DHCP proxies
	FF02::1:FFXX:XXXX	Solicited-Node address. XX:XXXX indicates the lower-order 24 bits of an IPv6 address of a node.
Local range of station	FF05::2	Addresses of all routers
	FF05::1:3	All DHCP servers
	FF05::1:4	All DHCP relays
	FF05::1:1000~FF05::1:13FF	Service location

1.3.4 IPv6 Multicast MAC Address

The higher-order 16 bits of an IPv6 multicast MAC address is 0x3333, and the lower-order 32 bits are those of the IPv6 multicast address. Figure 1-3 shows the mapping of the IPv6 multicast address FF2E::F01E:1 to the MAC address. There are more IPv6 multicast addresses mapped to the same MAC address.

Figure 1-3 Mapping of IPv6 Address to Multicast MAC Address



1.4 IPv4 Multicast Protocols

A complete set of protocols is developed during the development of multicast to complete multicast data transmission. IPv4 multicast covers the following protocols:

- Internet Group Management Protocol (IGMP): Manages IPv4 multicast members and runs in the network segment that hosts are connected to. IGMP supports data communication with the upstream multicast routing protocols, and maintains and manages relationship of members by adding or removing the members. Three versions of IGMP are available:
  - IGMPv1: Supports the ASM model, and supports the SSM model only when SSM mapping is configured.
  - IGMPv2: Supports the ASM model, and supports the SSM model only when SSM mapping is configured.
  - IGMPv3: Supports the ASM and SSM models.
- PIM: It is a multicast routing protocol in an IPv4 network, and used for route-based query and forwarding of multicast data. PIM includes the following two types:
  - Protocol Independent Multicast Dense Mode (PIM-DM): Applies to a small-size network in which members of a multicast group are densely distributed. The protocol is not distinguished in the ASM and SSM models.
  - Protocol Independent Multicast Sparse Mode (PIM-SM): Applies to a network in any form, and especially applies to a large-size network in which members of a multicast group are sparsely distributed. The protocol is distinguished in the ASM and SSM models based on the data and the multicast addresses in protocol packets.
- Internet Group Management Protocol Snooping (IGMP Snooping): Creates a multicast L2 forwarding table by listening to the IGMP packets between upstream L3 devices and member hosts, and manages and controls multicast packet forwarding on L2 network to suppress multicast data flooding.
- Multicast Source Discovery Protocol (MSDP): Connects multiple PIM-SM domains, shares multicast sources among the PIM-SM domains, and forwards multicast packets across the domains.

1.5 IPv6 Multicast Protocols

IPv6 multicast covers the following protocols:

- Multicast Listener Discovery (MLD): Manages IPv6 multicast members and runs in the network segment that hosts are connected to. MLD supports data communication with the upstream multicast routing protocols, and maintains and manages relationship of members by adding or removing the members. Two versions of MLD are available:
  - MLDv1: Supports the ASM model, and supports the SSM model only when SSM mapping is configured.
  - MLDv2: Supports the ASM and SSM models.
- PIM (IPv6): It is a multicast routing protocol in an IPv6 network, and used for route-based query and forwarding of multicast data. This protocol applies to a network in any form, and especially applies to a large-size network in which members of a multicast group are sparsely distributed. The protocol is distinguished in the ASM and SSM models based on the data and the multicast addresses in protocol packets.

- Multicast Listener Discovery Snooping (MLD Snooping): Creates a multicast L2 forwarding table by listening to the MLD packets between upstream L3 devices and member hosts, and manages and controls multicast packet forwarding on L2 network to suppress multicast data flooding.

1.6 Basic Concepts of PIM

- PIM device

  A device on which PIM is enabled. Multicast packets are forwarded on the PIM device.

- PIM interface

  An interface on which PIM is enabled. The PIM interface that receives multicast packets is an upstream interface, and the PIM interface that sends multicast packets is a downstream interface. The network segment where the upstream interface resides is an upstream network segment, and the network segment where the downstream interface resides is a downstream network segment.

- PIM network

  Connections of PIM devices through the PIM interfaces constitute a PIM network.

- PIM domain

  On some PIM interfaces, borders can be set to divide a large PIM network into multiple PIM domains. The borders can reject specified multicast packets or control transmission of PIM messages.

- Rendezvous point (RP)

  A PIM device functioning as an RP collects multicast source and group member information on a network.

- Candidate-RP (C-RP)

  A device selects RPs from C-RPs based on the election mechanism, and enables the RPs to serve different multicast groups.

- Designated router (DR)

  A PIM device can function as a DR. A DR that is connected to a multicast source and reports multicast source information to the RP is referred to as a source DR. A DR that is connected to members of a multicast group and reports the group member information to the RP is referred to as a member DR.

- Candidate-BSR (C-BSR)

  PIM devices can be configured as C-BSRs, from which a BSR is elected. When the BSR becomes faulty, other C-BSRs initiate a new election request to elect a new BSR. This avoids service interruption.

- Bootstrap router (BSR)

  A PIM device functioning as a BSR collects and releases RP information in a PIM-SM domain. A BSR is elected from candidate-BSRs. Only one BSR is available in a PIM-SM domain. However, there can be multiple candidate-BSRs in the PIM-SM domain.

- Multicast distribution tree

  Multicast packets are transmitted from one point to multiple points. The forwarding path is displayed as a tree. This forwarding path is referred to as a multicast distribution tree (MDT). There are two types of MDTs:

  - RP tree (RPT): The RP is regarded as the root, and the DRs that connect group members are regarded as leaves.
  - Shortest path tree (SPT): The DR that connects multicast sources is regarded as the root, and the RPs

or DRs that connect group members are regarded as leaves.

- (*, G) and (S, G)
  - (*, G) refers to a routing entry corresponding to packets that are sent from any multicast source to Group G. The forwarding path corresponding to the packets that are sent from any multicast source to Group G is referred to as an RPT.
  - (S, G) refers to a routing entry corresponding to packets that are sent from source S to Group G. The forwarding path corresponding to the packets that are sent from source S to Group G is an SPT.

Contents

1 Configuring IPv4 Multicast Route Management 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Topology of IPv4 Multicast 1
1.1.3 Basic Concepts of IPv4 Multicast 2
1.1.4 Principle of IPv4 Multicast Forwarding 3
1.1.5 RPF Checking 3
1.1.6 RPF Checking and Multicast Packet Forwarding 5
1.2 Restrictions and Guidelines 5
1.3 Configuration Task Summary 5
1.4 Configuring Basic Functions of IPv4 Multicast 6
1.4.1 Overview 6
1.4.2 Restrictions and Guidelines 6
1.4.3 Procedure 6
1.5 Configuring RPF Route Selection Based on the Longest Match Rule 7
1.5.1 Overview 7
1.5.2 Restrictions and Guidelines 7
1.5.3 Procedure 7
1.6 Configuring Static Routes of IPv4 Multicast 7
1.6.1 Overview 7
1.6.2 Procedure 8
1.7 Configuring IPv4 Multicast Borders 8

1.7.1 Overview 8
1.7.2 Restrictions and Guidelines 8
1.7.3 Procedure 8
1.8 Configuring L2 Direction Control for Multicast Streams 9
1.8.1 Overview 9
1.8.2 Restrictions and Guidelines 9
1.8.3 Procedure 9
1.9 Configuring TTL Threshold 9
1.9.1 Overview 9
1.9.2 Restrictions and Guidelines 10
1.9.3 Procedure 10
1.10 Configuring Forced Forwarding of Multicast Packets by Software 10
1.10.1 Overview 10
1.10.2 Procedure 10
1.11 Configuring Entry Capacity of IPv4 Multicast Routing Table 11
1.11.1 Overview 11
1.11.2 Procedure 11
1.12 Configuring the Overwriting Mechanism upon Overflow of Multicast Hardware Forwarding Entries 11
1.12.1 Overview 11
1.12.2 Procedure 11
1.13 Configuring Immediate Delivery of Entries to the Forwarding Plane 12
1.13.1 Overview 12
1.13.2 Procedure 12
1.14 Configuring Multicast Non-Stop Forwarding Parameters 12

1.14.1 Overview 12
1.14.2 Procedure 12
1.15 Configuring the Multicast Lossless Work Mode 12
1.15.1 Overview 12
1.15.2 Restrictions and Guidelines 13
1.15.3 Procedure 13
1.16 Monitoring 13
1.17 Configuration Examples 15
1.17.1 Configuring IP Multicast Static Route to Change RPF Routes 15
1.17.2 Configuring Multicast Static Routes to Connect RPF Routes 21

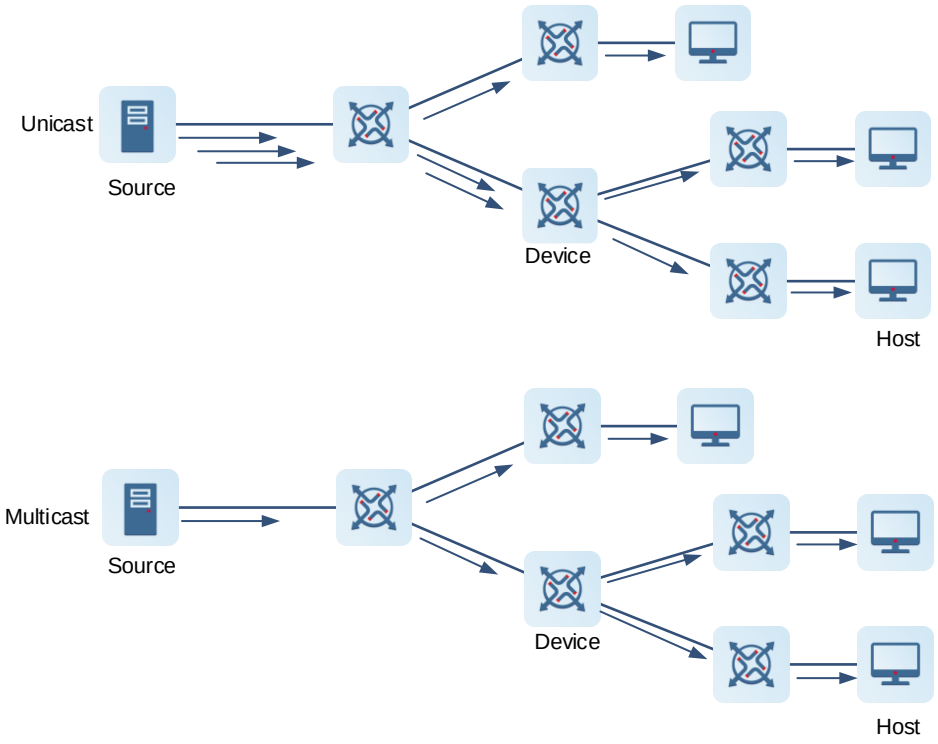
1 Configuring IPv4 Multicast Route Management

1.1 Introduction

1.1.1 Overview

IPv4 multicast transmits a packet from a source to multiple devices. The multicast packet is sent from the source, forwarded by intermediate devices, and then received by receivers in a specified group.

Figure 1-1 Multicast Topology vs. Unicast Topology



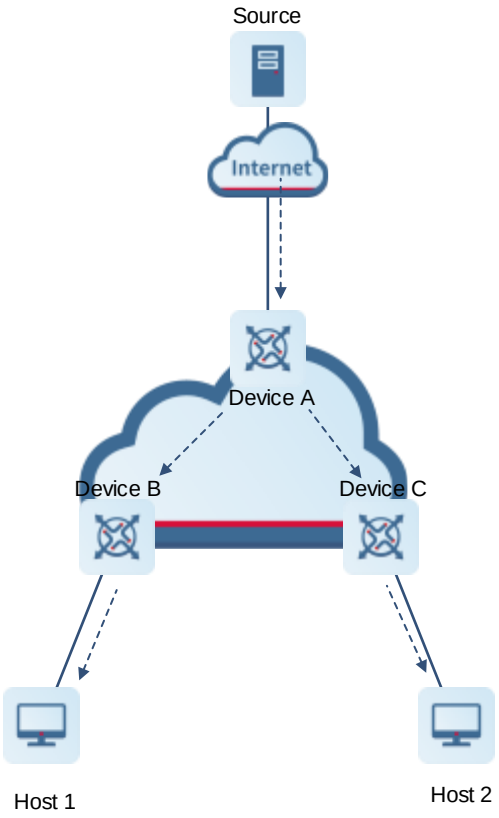
Compared with unicast, multicast reduces redundant traffic, relieves transmission loads of servers and intermediate devices, and completes one-to-multiple content transmission with high efficiency. Multicast is applied to different network services that have high requirements for real timeliness, for example, Internet TV, distance education, live broadcast and multimedia conference.

Proper multicast routes must be configured on the intermediate devices so that IPv4 multicast packets can be correctly forwarded. This chapter describes how to create and maintain IPv4 multicast routes and control multicast packet forwarding based on the multicast routes.

1.1.2 Basic Topology of IPv4 Multicast

Figure 1-2 shows a simple multicast topology that includes a multicast source, multicast devices (Device A, Device B, and Device C), and members in a multicast group.

Figure 1-2 Basic Topology of IPv4 Multicast



As shown in Figure 1-2, the transmission process of a multicast packet is as follows.

(1) The multicast source functions as a packet sender and sends a multicast packet to one or more multicast groups.

(2) The multicast packet is forwarded by the intermediate devices to the packet receivers (members in the multicast group). The intermediate devices are routers or switches that support the multicast function and can correctly forward the multicast packet. They are referred to as multicast devices collectively.

(3) This multicast group contains multiple members and they reside in different network segments on the network. Therefore, the multicast packet sent from the multicast source may be replicated into multiple copies by the multicast devices and then sent to the members in the multicast group.

(4) The members in the multicast group receive the multicast packet from the source.

1.1.3 Basic Concepts of IPv4 Multicast

1. PIM

Protocol Independent Multicast (PIM) is a multicast routing protocol that provides multicast routes based on unicast static routes or unicast routing tables that are generated based on the unicast route protocols. PIM is independent from the unicast routing protocols used and can use the static routes or unicast routing tables that are generated based on any unicast routing protocol, for example, Routing Information Protocol (RIP), Open Shortest Path First (OSPF), Intermediate System-to-Intermediate System (IS-IS), and Border Gateway Protocol (BGP).

PIM generates multicast routing entries based on the reverse path forwarding (RPF) mechanism, and controls multicast packet forwarding based on the entries. For more information about RPF, see 1.1.5 RPF Checking.

2. Multicast Routing Entry and Forwarding Entry

A multicast routing entry is generated based on multicast routing protocols. Specific routing entries vary with the routing protocols, and they are a comprehension of the routing protocol entries.

A multicast forwarding entry is used to guide multicast packet forwarding, and it is generated based on routing entry information.

3. Multicast Network

Numerous multicast devices constitute a multicast network, and multicast packets are transmitted between multicast sources and members in multicast groups on the network. Multicast packet query and forwarding between multicast devices are completed by running PIM. These PIM running devices are referred to as PIM routers.

PIM takes effect to device interfaces. An interface with PIM enabled is referred to as a PIM interface. A PIM interface that receives a multicast packet is an upstream interface and the interface resides in an upstream network segment. A PIM interface that sends a multicast packet is a downstream interface and the interface resides in a downstream network segment.

PIM routers are connected through PIM interfaces, which constitutes a PIM network.

4. Multicast Border

Data in a multicast network is not transmitted boundlessly. You can configure an access control list (ACL) on a PIM interface of a device to filter multicast packets and allow packets of specific multicast groups to pass. That is, set a border to control multicast packet transmission.

By configuring borders on PIM interfaces, you can divide a large PIM network into multiple PIM domains.

1.1.4 Principle of IPv4 Multicast Forwarding

To ensure correct forwarding of multicast packets by multicast devices, enable multicast routing protocols on the devices.

Upon receiving a multicast packet, a multicast device checks whether the multicast packet is valid and selects proper paths based on the routing table to forward the packet to one or more downstream devices. The basic process of IPv4 multicast forwarding is as follows:

(1) A multicast device receives a multicast packet from an upstream interface.

(2) The multicast device queries the multicast routing table to obtain proper forwarding paths. If no proper forwarding path is found, the multicast device performs RPF checking to determine forwarding paths and updates the multicast routing table.

(3) The multicast device forwards this packet to the outbound interfaces of all the determined forwarding paths.

1.1.5 RPF Checking

1. Basic Principles

In unicast transmission, the destination address in a packet corresponds to a specific address in a network. Provided that the destination address is determined, a network device can determine the outbound interface of the packet. In multicast transmission, a multicast address corresponds to members in a multicast group and a

multicast device does not know the IP address of each member. Therefore, the receiver addresses cannot be determined based on the destination address.

RPF checking is used to determine outbound interfaces of a multicast packet. The core idea behind RPF checking is that: An RPF route is generated to check whether the theoretical destination interface of a multicast packet is consistent with the actual interface. If they are not consistent, the packet is discarded. If they are consistent, the packet is forwarded to all the determined outbound interfaces.

2. RPF Checking Procedure

The RPF checking procedure consists of the following steps:

(1) Select optimal routes.

Take the address (the source IP address in a multicast packet) of the multicast source as the destination address and select optimal routes from the local device to the destination address.

The optimal routes are selected from the unicast routing table, Multiprotocol Extensions for BGP-4 (MBGP) routing table, and multicast static routing table, respectively.

(2) Determine an RPF route.

Select a route from the determined optimal routes of the routing tables as an RPF route based on rules. Table 1-1 lists the specific rules.

Table 1-1 RPF Routing Rules

Select RPF Route Based on the Longest Match	Rule Priority (a Smaller Number Indicates a Higher Priority)	Description
No	1	The route with the highest priority is selected.
	2	An RPF route is selected in the sequence of multicast static route, MBGP route, and unicast route.
Yes	1	The route with the longest match mask is selected.
	2	The route with the highest priority is selected.
	3	An RPF route is selected in the sequence of multicast static route, MBGP route, and unicast route.

An RPF route includes RPF interface information and the information is used for checking the RPF interface.

- In a unicast route and an MBGP route, the outbound interface in a routing entry is used as an RPF interface, with the next hop being an RPF neighbor.
- In a multicast static routing entry, the RPF interface and RPF neighbor are specified.

(3) Check the RPF interface.

Check whether the inbound interface of the multicast packet is consistent with the RPF interface.

- If yes, the RPF interface passes the checking.
- Otherwise, the RPF interface fails the checking.

1.1.6 RPF Checking and Multicast Packet Forwarding

When a multicast packet is forwarded, query a path from the routing table first. If no proper path is found, RPF checking is performed. Path information that is generated after RPF checking passes is updated to the routing table and can be queried for packet forwarding next time.

The rules of forwarding a multicast packet are as follows:

- No forwarding entry exists in a multicast routing table.

  RPF checking is performed for this packet. The RPF interface is used as an inbound interface and updated to the routing table.

  - If the RPF interface passes the checking, the packet is forwarded to all outbound interfaces.
  - If the RPF interface fails the checking, this packet is discarded.

- There are forwarding entries in a multicast routing table.
  - If the inbound interface of the packet matches the inbound interface in an entry, the packet is forwarded to all outbound interfaces.
  - If the inbound interface of the packet does not match the inbound interface in an entry, RPF checking is performed. If the RPF interface is consistent with the inbound interface in an entry, the packet is discarded. If the RPF interface is not consistent with the inbound interface in an entry, the inbound interface in the entry is updated as an RPF interface and the inbound interface in the packet is compared with the RPF interface to determine whether they are consistent.

1.2 Restrictions and Guidelines

- All the configuration items must be configured after the basic functions of IPv4 multicast are configured.
- The multicast routing function is implemented based on unicast routes. Before applying the multicast function, ensure that the unicast communication between network nodes is normal.

1.3 Configuration Task Summary

IPv4 multicast route management configuration includes the following tasks:

(1) Configuring Basic Functions of IPv4 Multicast

(2) (Optional) Configuring a packet forwarding rule. The following configuration tasks are optional. Select tasks for configuration according to actual condition.

- Configuring RPF Route Selection Based on the Longest Match Rule
- Configuring Static Routes of IPv4 Multicast
- Configuring IPv4 Multicast Borders
- Configuring L2 Direction Control for Multicast Streams
- Configuring TTL Threshold
- Configuring Forced Forwarding of Multicast Packets by Software

(3) (Optional) Configuring multicast entries. The following configuration tasks are optional. Select tasks for configuration according to actual condition.

- Configuring Entry Capacity of IPv4 Multicast Routing Table

- Configuring the Overwriting Mechanism upon Overflow of Multicast Hardware Forwarding Entries

(4) (Optional) Configuring Immediate Delivery of Entries to the Forwarding Plane

(5) (Optional) Configuring Multicast Non-Stop Forwarding Parameters

(6) (Optional) Configuring the Multicast Lossless Work Mode

1.4 Configuring Basic Functions of IPv4 Multicast

1.4.1 Overview

To configure basic functions of IPv4 multicast, enable the IPv4 multicast routing function and the IPv4 multicast protocols on a device so that the device has the multicast packet forwarding function and can provide IPv4 multicast services for multicast sources and hosts in multicast groups.

1.4.2 Restrictions and Guidelines

- The IPv4 multicast routing function should be enabled on each router unless otherwise specified.
- IPv4 unicast routes must be configured so that PIM can use the unicast routes to generate multicast routes.
- After L3 multicast is enabled for a private virtual local area network (PVLAN) and a super VLAN, if there is a multicast source in a sub VLAN, the entry count increases by 1. This is because the inbound validity will be checked during multicast forwarding, and therefore an extra entry will be generated, and the inbound of this entry is the sub VLAN that the multicast stream enters.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IPv4 multicast routing.

ip mrouting-enable

The IPv4 multicast routing function is disabled by default.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(5) Enable an IPv4 multicast protocol on an interface. Select at least one of them to configure.

- Enable Protocol Independent Multicast Dense Mode (PIM-DM) on the interface.

  ip pim dense-mode

PIM-DM is disabled for an interface by default.

PIM-DM is a multicast routing protocol that is used in dense mode and applies to small-sized networks where members in a multicast group are densely distributed. For more information about PIM-DM, see *Configuring PIM-DM*.

- Enable Protocol Independent Multicast Sparse Mode (PIM-SM) on the interface.

  ip pim sparse-mode

  PIM-SM is disabled for an interface by default.

  PIM-SM is a multicast routing protocol that is used in sparse mode and applies to large-sized networks where members in a multicast group are sparsely distributed. For more information about PIM-SM, see *Configuring PIM-SM*.

1.5 Configuring RPF Route Selection Based on the Longest Match Rule

1.5.1 Overview

In multicast routing, there may be different routing tables, for example, multicast static routing table, MBGP routing table, and unicast routing table. An optimal route should be selected from all the routing tables and used as an RPF route.

If the longest match rule is used, the route with the longest match mask is selected from the optimal routes in different routing tables as the RPF route. If the masks are equal in length, the route with the highest priority is selected from the optimal routes. If the priorities are consistent, a route is selected in the sequence of multicast static route, MBGP route, and unicast route.

1.5.2 Restrictions and Guidelines

The longest match rule can be configured on each device unless otherwise specified.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure RPF route selection based on the longest match rule.

ip multicast rpf-routing longest-match

By default, the route with the highest priority is selected as the RPF route. If they have the same priority, the RPF routes are selected in the sequence of multicast static route, MBGP route, and unicast route.

1.6 Configuring Static Routes of IPv4 Multicast

1.6.1 Overview

You can configure static routes of IPv4 multicast to specify RPF interfaces or RPF neighbors for multicast packets from specific multicast sources. This function can be used to change and connect RPF routes.

- Change an RPF route: If you want a device to receive multicast packets from a multicast source through a

specified interface, but this interface is not the required interface based on the RPF checking result, you need to configure a multicast static route to specify this required interface as an RPF interface.

- Connect an RPF route: If a unicast route is interrupted, a multicast route failed to be generated by RPF. You can configure a multicast static route to specify an RPF interface to complete RPF checking and realize multicast packet forwarding.

1.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a multicast static route.

ip rpf-mroute *ipv4-source-address mask* { [**bgp** | **isis** | **ospf** | **rip** | **static-route**] { *ipv4-rpf-address* | *interface-type interface-number* } } [*distance*]

No multicast static route is configured by default. To specify an outbound interface for a multicast static route, set the outbound interface to point-to-point type.

1.7 Configuring IPv4 Multicast Borders

1.7.1 Overview

After IPv4 multicast borders are configured, a multicast network can be divided into different domains so that specified multicast packets are transmitted in a domain.

The multicast border function filters out specified multicast packets based on an ACL to forbid packets of this type from entering or leaving an interface.

1.7.2 Restrictions and Guidelines

IP multicast borders can be configured on each PIM router interface unless otherwise specified.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.

interface loopback *interface-number*

(4) Configure an IP multicast border.

ip multicast scope-boundary { *acl-name* | *acl-number* } [**in** | **out**]

No multicast border is configured by default.

This command is associated with a standard access control list (ACL). If an extended ACL is used, an error occurs in filtering.

1.8 Configuring L2 Direction Control for Multicast Streams

1.8.1 Overview

By configuring L2 direction control for multicast streams, you can control multicast stream forwarding on an interface.

You can configure this command multiple times for a multicast stream. That is, configure multiple interfaces to forward the stream. After direction control is configured for a multicast stream, this multicast stream can be forwarded only by these configured interfaces. Other interfaces are not allowed to forward the stream.

1.8.2 Restrictions and Guidelines

This command controls only the forwarding of multicast streams on interfaces, but does not directly affect the processing of protocol packets by the multicast protocols. However, some features of the multicast protocols (PIM-DM or PIM-SM) depend on the multicast streams. Behaviors of the multicast routing protocols may be affected.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure L2 direction control for multicast streams.

ip multicast static *ipv4-source-address ipv4-group-address interface-type interface-number*

No L2 direction control is configured for multicast streams by default.

1.9 Configuring TTL Threshold

1.9.1 Overview

The time to live (TTL) value in an IP packet is used to define the maximum number of hops that this packet can transmit through.

To configure a TTL threshold means to configure the minimum TTL value of multicast packets allowed on an interface. Before forwarding a multicast packet on an interface, the device checks whether the TTL value in the packet is greater than the TTL threshold. If yes, the device forwards the packet. Otherwise, the device discards this packet.

1.9.2 Restrictions and Guidelines

- A TTL threshold can be configured on each PIM router interface unless otherwise specified.
- The TTL threshold must be configured on router interfaces and takes effect only to multicast packets.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure a TTL threshold for an interface.

ip multicast packet-ttl-threshold *ttl-threshold-value*

The default TTL threshold for an interface is **0**.

1.10 Configuring Forced Forwarding of Multicast Packets by Software

1.10.1 Overview

After the mechanism of forced forwarding of multicast packets by software is configured, multicast packets destined to the local device can be forwarded to other interfaces by software.

1.10.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable forced forwarding of multicast packets by software.

msf force-forwarding

The mechanism of forced forwarding of multicast packets by software is disabled by default.

1.11 Configuring Entry Capacity of IPv4 Multicast Routing Table

1.11.1 Overview

A device creates and maintains a multicast route forwarding entry for each received multicast packet. Therefore, the entry quantity increases continuously. If the number of entries maintained by a device surpasses the limited hardware resources of the device, hardware forwarding may be changed to software forwarding, which consumes device memory and reduces device performance.

The entry capacity of an IPv4 multicast routing table refers to the maximum number of entries that can be added to the IPv4 multicast routing table. You can configure a proper entry capacity based on actual network situation and service performance without affecting running of the device.

1.11.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure entry capacity of an IPv4 multicast routing table.

ip multicast max-route *max-route* [*max-threshold*]

The default entry capacity of an IPv4 multicast routing table is **1000** entries.

1.12 Configuring the Overwriting Mechanism upon Overflow of Multicast Hardware Forwarding Entries

1.12.1 Overview

The hardware forwarding entries are limited. If they are used up, the device uses software forwarding entries. Compared with hardware forwarding entries, software forwarding entries feature lower performance, occupy device memory, and deteriorate device performance.

After the overwriting mechanism upon overflow of multicast hardware forwarding entries is enabled, new entries can overwrite the earliest entries when hardware forwarding entries are used up.

1.12.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the overwriting mechanism upon overflow of multicast hardware forwarding entries.

msf ipmc-overflow override

The overwriting mechanism upon overflow of multicast hardware forwarding entries is disabled by default.

1.13 Configuring Immediate Delivery of Entries to the Forwarding Plane

1.13.1 Overview

A routing table generated based on a routing table contains entries guiding multicast packet forwarding. After the function of immediate delivery of entries to the forwarding plane is configured, the forwarding rules specified by the entries take effect immediately.

1.13.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the function of immediate delivery of IPv4 multicast L2 entries to the forwarding plane.

msf immediately-install

1.14 Configuring Multicast Non-Stop Forwarding Parameters

1.14.1 Overview

In certain active/standby switchover scenarios, multicast protocols such as PIM-SM and Internet Group Management Protocol (IGMP) Snooping need re-convergence during the switchover. You can configure non-stop forwarding parameters to ensure continuous forwarding of multicast streams during re-convergence of the multicast protocols.

1.14.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure multicast non-stop forwarding parameters.

msf nsf { **convergence-time** *convergence-time* | **leak** *leak-time* }

The maximum period for multicast protocol convergence is **20** seconds and the packet leakage time is **30** seconds by default.

After the configured protocol convergence period times out, all multicast forwarding entries that are not updated during the convergence period are deleted.

1.15 Configuring the Multicast Lossless Work Mode

1.15.1 Overview

Devices run independently in a distributed environment, and the hardware settings of multicast prefixes and multicast egresses on the devices cannot guarantee the time sequence. Therefore, there is a large probability that packet loss occurs on the multicast services when the multicast hardware is updated. The multicast

lossless work mode is a technology aiming to ensure that no packet loss occurs on the multicast services when multicast hardware is updated in distributed environments.

1.15.2 Restrictions and Guidelines

- Configuring the multicast lossless work mode can ensure that the multicast services do not encounter packet loss in distributed scenarios.
- After the multicast lossless work mode is enabled, the installation performance of multicast entries is reduced by about 80%.

1.15.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the multicast lossless work mode.

multicast xgv-lossless

1.16 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information of the **debugging** command occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-2 Monitoring

Command	Purpose
cleanup ip mrouting-table { *ipv4-group-address* [*ipv4-source-address*] | * }	Clears the IPv4 multicast routing table.
cleanup ip mrouting-table statistics { *ipv4-group-address* [*ipv4-source-address*] | * }	Resets the statistics in the IPv4 multicast routing table.
display ip mrouting-table [*ipv4-group-address* | *ipv4-source-address*] * [**dense** | **sparse**] [**summary**]	Displays the IPv4 multicast routing table.

Command	Purpose
display ip mrouting-table [*ipv4-group-address* | *ipv4-source-address*] * [**dense** | **sparse**] **count**	Displays the statistics in the IPv4 multicast routing table.
display ip mrouting-table static-route	Displays the information about IPv4 multicast static route.
display ip rpf-routing *ipv4-source-address*	Displays the RPF information of a specified IPv4 source address.
display ip mvif [*interface-type interface-number*]	Displays the information about IPv4 multicast interfaces.
display ip mrf mfc	Displays the IPv4 L3 multicast routing table.
display msf global	Displays details of the MSF process.
display msf nsf	Displays the configuration of IPv4 multicast non-stop forwarding.
display msf msc	Displays the IPv4 multi-layer multicast routing table.
debugging nsm mcast all	Debugs the running procedure of the multicast core.
debugging nsm mcast fib-msg	Debugs the communication procedure between the IPv4 multicast core and the protocol module.
debugging nsm mcast vif	Debugs the interface running procedure about the IPv4 multicast core.
debugging nsm mcast stats	Debugs the interface and entry statistics processing procedure about the IPv4 multicast core.
debugging ip mrf forwarding	Debugs the IPv4 L3 multicast packet forwarding procedure.
debugging ip mrf mfc	Debugs the operation procedure of L3 multicast forwarding entries on an IPv4 network.
debugging ip mrf event	Debugs the processing procedure of L3 multicast forwarding events on an IPv4 network.
debugging msf forwarding	Debugs the IPv4 multi-layer multicast packet forwarding procedure.
debugging msf msc	Debugs the operation procedure of multi-layer multicast forwarding entries on an IPv4 network.

Command	Purpose
debugging msf ssp	Debugs the bottom-layer hardware processing procedure of IPv4 multi-layer multicast forwarding.
debugging msf api	Debugs the API calling procedure of IPv4 multi-layer multicast forwarding.
debugging msf event	Debugs the processing procedure of multi-layer multicast forwarding events on an IPv4 network.

1.17 Configuration Examples

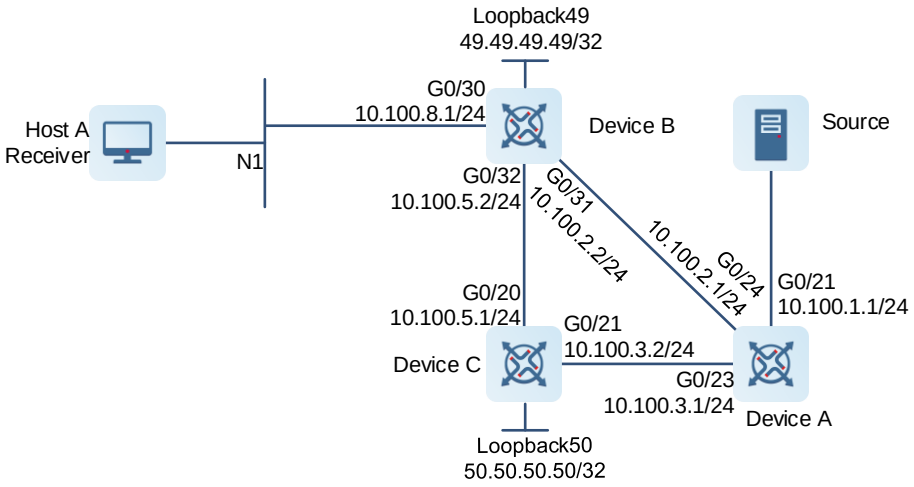
1.17.1 Configuring IP Multicast Static Route to Change RPF Routes

1. Requirements

As shown in Figure 1-3, Device A, Device B and Device C run the OSPF routing protocol, and the PIM-SM function is enabled on each interface of these devices. Multicast data from the multicast source Source is sent to Host A through the Device A -> Device B link. To reduce the burden of the Device A -> Device B link and separate multicast and unicast data streams, multicast data from the multicast source Source is to be transmitted through the Device A -> Device C -> Device B path.

2. Topology

Figure 1-3 Topology of Configuring IP Multicast Static Route to Change RPF Routes



3. Notes

As shown in Figure 1-3, configuring a multicast static route on Device B to change the RPF interface for receiving multicast data can separate multicast and unicast data streams shown in the figure. Unicast data is transmitted through the Device A -> Device B path, while multicast data through the Device A -> Device C -> Device B path. The specific configuration method is as follows:

- Configure an IP address and unicast routing protocol on the multicast device. Unicast route reachability is the basis for the multicast protocol to work properly.
- Enable the IP multicast routing function on all multicast devices; enable the PIM-SM function on the interface connecting the multicast device, and the interface connecting the host and the multicast source; and configure C-BSR and C-RP on Device C.
- Configure RPF multicast static routing function on Device B, and specify the RPF neighbor as Device C to change the forwarding path of multicast packets.

4. Procedure

(1) Configure an IP address and unicast routing protocol on the multicast device.

Configure an IP address and unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/21)# ip address 10.100.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/23)# ip address 10.100.3.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# interface GigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/24)# ip address 10.100.2.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.1.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.2.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.3.0 0.0.0.255 area 0
DeviceA(cfg-router)# exit
```

Configure an IP address and unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/30)# ip address 10.100.8.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/31)# ip address 10.100.2.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/32)# ip address 10.100.5.2 255.255.255.0
```

```
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ip address 49.49.49.49 255.255.255.255
DeviceB(cfg-if-Loopback 49)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.2.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.8.0 0.0.0.255 area 0
DeviceB(cfg-router)# exit
```

Configure an IP address and unicast routing protocol on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/20)# ip address 10.100.5.1 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ip address 10.100.3.2 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ip address 50.50.50.50 255.255.255.255
DeviceC(cfg-if-Loopback 50)# exit
DeviceC(cfg)# router-protocol ospf 49
DeviceC(cfg-router)# network-advertise 10.100.3.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 50.50.50.50 0.0.0.0 area 0
DeviceC(cfg-router)# exit
```

(2) Enable the IP multicast routing function on all multicast devices; enable the PIM-SM function on the interface connecting the multicast device, and the interface connecting the host and the multicast source; and configure C-BSR and C-RP on Device C.

Enable the multicast routing function and PIM-SMv6 related functions on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# interface GigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
```

Enable the multicast routing function and PIM-SMv6 related functions on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip mrouting-enable
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ip pim sparse-mode
DeviceB(cfg-if-Loopback 49)# exit
```

Enable the multicast routing function and PIM-SMv6 related functions on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ip mrouting-enable
DeviceC(cfg)# ipv6 mrouting-enable
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ip pim sparse-mode
DeviceC(cfg-if-Loopback 50)# exit
```

Configure C-BSR and C-RP on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip pim c-bsr GigabitEthernet 0/21
DeviceA(cfg)# ip pim c-rp GigabitEthernet 0/21
```

(2) Configure RPF multicast static routing function on Device B, and specify the RPF neighbor as Device C to change the forwarding path of multicast packets.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip rpf-mroute 10.100.1.0 255.255.255.0 10.100.5.1
```

5. Verification

(1) Run the **display ip mld groups** command on Device B to check whether Host A has successfully joined the multicast group 225.0.0.11, and whether the multicast source 10.100.1.2 has sent packets to the multicast group 225.0.0.11.

```
DeviceB# display ip igmp groups
IGMP Connected Group Membership
```

```
Group Address      Interface                    Uptime     Expires    Last Reporter
225.0.0.11         GigabitEthernet 0/30         00:01:03   00:03:37   10.100.8.2
DeviceB#
```

(2) Run the **display ip mrouting-table** command on Device A to check the multicast routing table and verify that the next hop is Device C and the outbound interface is GigabitEthernet 0/23.

```
DeviceA# display ip mrouting-table

IP Multicast Routing Table
Flags: I - Immediate Stat, T - Timed Stat, F - Forwarder installed,
    R - RPT, S - SPT, s - SSM Group
Timers: Uptime/Stat Expiry
Interface State: Interface (TTL)

(10.100.1.2, 225.0.0.11), uptime 00:01:58, stat expires 00:02:38
Owner PIMSM, Flags: TFS
 Incoming interface: GigabitEthernet 0/21
 Outgoing interface list:
  GigabitEthernet 0/23 (1)

DeviceA#
```

(3) Run the **display ip rpf-routing** command on Device B to verify that the next hop of the multicast source is Device C and the RPF interface is GigabitEthernet 0/32.

```
DeviceB# display ip rpf-routing 10.100.1.0
RPF information for 10.100.1.0
RPF interface: GigabitEthernet 0/32
  RPF neighbor: 10.100.5.1
  RPF route: 10.100.1.0/24
  RPF type: static-route
  RPF recursion count: 0
  Doing distance-preferred lookups across tables
Distance: 0
  Metric: 0
DeviceB#
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.1.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/23
```

```
 unset switch-port
 ip address 10.100.3.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/24
 unset switch-port
 ip address 10.100.2.1 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.1.0 0.0.0.255 area 0
 network-advertise 10.100.2.0 0.0.0.255 area 0
 network-advertise 10.100.3.0 0.0.0.255 area 0
!
!
ip pim c-bsr GigabitEthernet 0/21
ip pim c-rp GigabitEthernet 0/21
!
```

- Device B configuration file

```
hostname DeviceB
!
ip mrouting-enable
!
ip rpf-mroute 10.100.1.0 255.255.255.0 10.100.5.1
!
interface GigabitEthernet 0/30
 unset switch-port
 ip address 10.100.8.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/31
 unset switch-port
 ip address 10.100.2.2 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/32
 unset switch-port
 ip address 10.100.5.2 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 49
 ip address 49.49.49.49 255.255.255.255
 ip pim sparse-mode
!
router-protocol ospf 49
```

```
 graceful-restart
 network-advertise 10.100.2.0 0.0.0.255 area 0
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 10.100.8.0 0.0.0.255 area 0
 network-advertise 49.49.49.49 0.0.0.0 area 0
!
```

- Device C configuration file

```
hostname DeviceC
!
ip mrouting-enable
!
interface GigabitEthernet 0/20
 unset switch-port
 ip address 10.100.5.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.3.2 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 50
 ip address 50.50.50.50 255.255.255.255
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.3.0 0.0.0.255 area 0
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 50.50.50.50 0.0.0.0 area 0
!
```

7. Common Errors

- An IPv6 unicast route is incorrectly configured.
- IPv6 multicast routing is not enabled on a multicast device.
- No IPv6 multicast routing protocol is enabled on an interface.

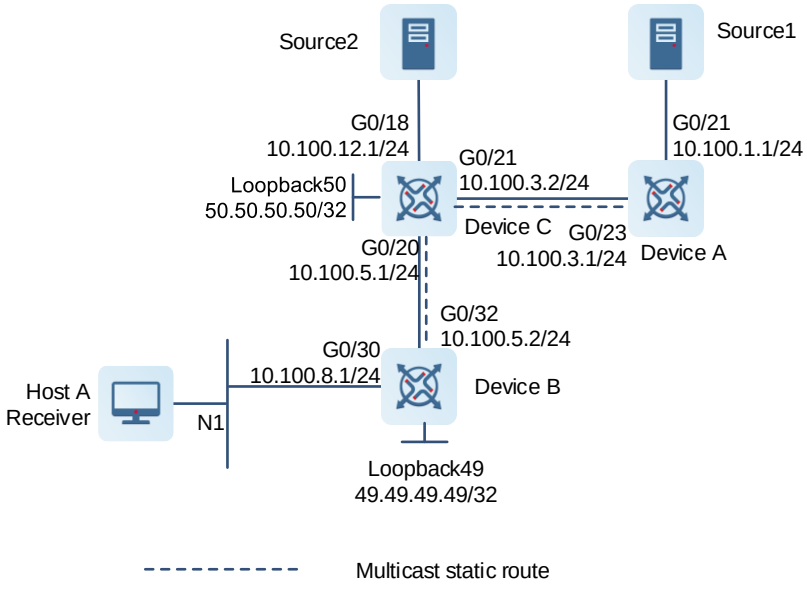
1.17.2 Configuring Multicast Static Routes to Connect RPF Routes

1. Requirements

As shown in Figure 1-4, the OSPF protocol is run on Device B and Device C, not Device A which runs a unicast routing protocol. The PIM-SM function is enabled on all interfaces on Device A, Device B and Device C. Host A can receive packets from Source 2 normally. It is required that Host A should also be able to receive multicast packets from Source 1.

2. Topology

Figure 1-4 Topology of Configuring IPv4 Multicast Static Routes to Connect RPF Routes



3. Notes

If a multicast static route is configured, an RPF route can be created to reach the multicast source Source 1 through the Device B ->Device C ->Device A path. The specific configuration method is as follows:

- Configure an IP address for each interface on these three devices, and configure the OSPF routing protocol on Device B and Device C, but not on Device A which runs a unicast routing protocol.
- Enable the IP multicast routing function, configure the PIM-SM protocol and C-BSR and C-RP on each Layer 3 interface of all devices, and configure the MLD function on the interface connected to the host.
- Configure a multicast static route on Device B and Device C to reach the multicast source Source 1.

4. Procedure

(1) Configure an IP address for each interface on each device, and configure the OSPF routing protocol on Device B and Device C, but not on Device A which runs a unicast routing protocol.

Configure an IP address for each interface on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/21)# ip address 10.100.1.1 24
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/23)# ip address 10.100.3.1 24
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# ip route-config 50.50.50.50 255.255.255.255 10.100.3.2
```

Configure an IP address for each interface and OSPF routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/30)# ip address 10.100.8.1 24
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/32)# ip address 10.100.5.2 24
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ip address 49.49.49.49 24
DeviceB(cfg-if-Loopback 49)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.8.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 49.49.49.49 0.0.0.0 area 0
DeviceB(cfg-router)# exit
```

Configure an IP address for each interface and OSPF routing protocol on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/18
DeviceC(cfg-if-GigabitEthernet 0/18)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/18)# ip address 10.100.12.1 24
DeviceC(cfg-if-GigabitEthernet 0/18)# exit
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/20)# ip address 10.100.5.1 24
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ip address 10.100.3.2 24
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ip address 50.50.50.50 32
DeviceC(cfg-if-Loopback 50)# exit
DeviceC(cfg)# router-protocol ospf 49
DeviceC(cfg-router)# network-advertise 10.100.3.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 50.50.50.50 0.0.0.0 area 0
DeviceC(cfg-router)# exit
```

(2) Enable the IP multicast routing function, configure the PIM-SM protocol and C-BSR and C-RP on each Layer 3 interface of all devices, and configure the MLD function on the interface connected to the host.

Enable the IP multicast routing function and the PIM-SM protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
```

Enable the IP multicast routing function and the PIM-SM protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip mrouting-enable
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ip pim sparse-mode
DeviceB(cfg-if-Loopback 49)# exit
```

Enable the IP multicast routing function and the PIM-SM protocol on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ip mrouting-enable
DeviceC(cfg)# interface GigabitEthernet 0/18
DeviceC(cfg-if-GigabitEthernet 0/18)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/18)# exit
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ip pim sparse-mode
DeviceC(cfg-if-Loopback 50)# exit
DeviceC(cfg)# ip pim c-bsr Loopback 50
DeviceC(cfg)# ip pim c-rp Loopback 50
```

Both Source 1 (10.100.1.2/24) and Source 2 (10.100.12.2/24) send multicast packets to the multicast group (225.0.0.10). Host A joins the multicast group 225.0.0.10 and can receive the multicast packets sent by Source 2, but not the multicast packets sent by Source 1.

Run the **display ip rpf-routing** 10.100.1.2 command on Device C. The route does not exist, showing that there is no RPF route from Device C to Source 1.

(3) Configure a multicast static route on Device B and Device C to reach the multicast source Source 1.

Configure a RPF multicast static route on Device B, and configure the RPF neighbor to Source 1 as Device C.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip rpf-mroute 10.100.1.0 255.255.255.0 10.100.5.1
```

Configure a RPF multicast static route on Device C, and configure the RPF neighbor to Source 1 as Device A.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ip rpf-mroute 10.100.1.0 255.255.255.0 10.100.3.1
```

5. Verification

Run the **display ip rpf-routing** 10.100.1.2 command on Device B to check the RPF information of Source 1. The RPF information is shown as follows.

```
DeviceB# display ip rpf-routing 10.100.1.2
RPF information for 10.100.1.2
RPF interface: GigabitEthernet 0/32
  RPF neighbor: 10.100.5.1
  RPF route: 10.100.1.0/24
  RPF type: static-route
  RPF recursion count: 0
  Doing distance-preferred lookups across tables
Distance: 0
  Metric: 0
```

Run the **display ip rpf-routing** 10.100.1.2 command on Device C to check the RPF information of Source 1. The RPF information is shown as follows.

```
DeviceC# display ip rpf-routing 10.100.1.2
RPF information for 10.100.1.2
RPF interface: GigabitEthernet 0/21
  RPF neighbor: 10.100.3.1
  RPF route: 10.100.1.0/24
  RPF type: static-route
  RPF recursion count: 0
  Doing distance-preferred lookups across tables
Distance: 0
  Metric: 0
```

Both multicast sources Source 1 and Source 2 send multicast packets to the multicast group 225.0.0.10, and Host A joins the multicast group 225.0.0.10.

Run the **display ip pim-info sparse-mode mroute** command on Device C to view the routing table information. There are entries for Source 1 in the multicast routing tables of Device C, indicating that Host A can receive multicast packets from Source 1.

```
DeviceC#
display ip pim-info sparse-mode mroute

IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 2
(S,G,rpt) Entries: 2
FCR Entries: 0
REG Entries: 0

(10.100.1.2, 225.0.0.10)
RPF nbr: 10.100.3.1
RPF idx: GigabitEthernet 0/21
SPT bit: 0
Upstream State: NOT JOINED
kat expires in 159 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
Outgoing
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(10.100.1.2, 225.0.0.10, rpt)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: RPT NOT JOINED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Pruned
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(10.100.12.2, 225.0.0.10)
RPF nbr: 10.100.11.2
```

```
RPF idx: GigabitEthernet 0/18
SPT bit: 0
Upstream State: NOT JOINED
kat expires in 202 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

(10.100.12.2, 225.0.0.10, rpt)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
```

```
Upstream State: RPT NOT JOINED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Pruned
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
DeviceB> privilege
DeviceB#
display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 1
(S,G) Entries: 0
(S,G,rpt) Entries: 0
FCR Entries: 2
REG Entries: 0

(*, 225.0.0.10)
RP: 50.50.50.50
RPF nbr: 10.100.5.1
```

```
RPF idx: GigabitEthernet 0/32
  Upstream State: JOINED
jt_timer expires in 51 seconds
    00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0   .   .   i   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
1   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
2   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
3   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
4   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
5   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
6   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
7   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
Joined
0   .   .   j   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
1   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
2   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
3   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
4   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
5   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
6   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
7   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
Asserted
0   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
1   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
2   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
3   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
4   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
5   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
6   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
7   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
FCR:
Source: 10.100.1.2
    00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Outgoing
0   .   .   0   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
1   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
2   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
3   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
4   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
5   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
6   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
7   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .
KAT timer running, 81 seconds remaining
  Packet count 1
Source: 10.100.12.2
```

```
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Outgoing
0  .  .  0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
KAT timer running, 81 seconds remaining
 Packet count 1
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.1.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/23
 unset switch-port
 ip address 10.100.3.1 255.255.255.0
 ip pim sparse-mode
!
ip route-config 50.50.50.50 255.255.255.255 10.100.3.2
!
```

- Device B configuration file

```
hostname DeviceB
!
ip mrouting-enable
!
ip rpf-mroute 10.100.1.0 255.255.255.0 10.100.5.1
!
interface GigabitEthernet 0/30
 unset switch-port
 ip address 10.100.8.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/32
 unset switch-port
```

```
 ip address 10.100.5.2 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 10.100.8.0 0.0.0.255 area 0
 network-advertise 49.49.49.49 0.0.0.0 area 0
!
```

- Device C configuration file

```
hostname DeviceC
!
ip mrouting-enable
!
ip rpf-mroute 10.100.1.0 255.255.255.0 10.100.3.1
!
interface GigabitEthernet 0/18
 unset switch-port
 ip address 10.100.12.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/20
 unset switch-port
 ip address 10.100.5.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.3.2 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 50
 ip address 50.50.50.50 255.255.255.255
 ip pim sparse-mode
!

router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 10.100.12.00.0.0.255 area 0
 network-advertise 50.50.50.50 0.0.0.0 area 0
!
ip pim c-bsr Loopback 50
ip pim c-rp Loopback 50
!
```

7. Common Errors

- An IP unicast route is incorrectly configured.
- An IPv4 unicast route is incorrectly configured.
- IPv4 multicast routing is not enabled on a router.
- IGMP function is not enabled on an interface.

Contents

1 Configuring IGMP....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Basic Principles of IGMPv1....1
1.1.3 Basic Principles of IGMPv2....3
1.1.4 Basic Principles of IGMPv3....4
1.1.5 IGMP SSM Mapping....5
1.1.6 IGMP Proxy....6
1.1.7 Protocols and Standards....6
1.2 Configuration Task Summary....6
1.3 Configuring Basic Functions of IGMP....7
1.3.1 Overview....7
1.3.2 Restrictions and Guidelines....7
1.3.3 Procedure....7
1.4 Configuring IGMP Parameters....8
1.4.1 Configuration Tasks....8
1.4.2 Configuring an IGMP Version....8
1.4.3 Configuring the Maximum Response Time for Query Packets....8
1.4.4 Configuring Common Group Query....9
1.4.5 Configuring Specific Group Query....10
1.4.6 Configuring Survival Period of Other Querier....10
1.5 Configuring IGMP Group Filtering....11

1.5.1 Configuration Tasks 11
1.5.2 Filtering a Multicast Group 11
1.5.3 Configuring the Maximum Number of Group Members 12
1.6 Configuring the IGMP Proxy Function 13
1.6.1 Overview 13
1.6.2 Restrictions and Guidelines 13
1.6.3 Procedure 14
1.7 Configuring the IGMP SSM Mapping Function 14
1.7.1 Overview 14
1.7.2 Restrictions and Guidelines 15
1.7.3 Procedure 15
1.8 Configuring the Router Alert Option of IGMP Packets 15
1.8.1 Overview 15
1.8.2 Restrictions and Guidelines 15
1.8.3 Procedure 15
1.9 Enabling Source Address Checking for IGMP Report Packets 16
1.9.1 Overview 16
1.9.2 Procedure 16
1.10 Enabling Fast Leave 16
1.10.1 Overview 16
1.10.2 Restrictions and Guidelines 16
1.10.3 Procedure 16
1.11 Configuring the Querier Robustness Variable 17
1.11.1 Overview 17

1.11.2 Procedure 17
1.12 Adding a Static Interface to a Group 17
1.12.1 Procedure 18
1.13 Simulating a Host to Join a Group 18
1.13.1 Overview 18
1.13.2 Procedure 18
1.14 Monitoring 19
1.15 Configuration Examples 20
1.15.1 Configuring Basic Functions of IGMP 20
1.15.2 Configuring the IGMP Proxy Function 39
1.15.3 Configuring SSM-MAP 43
1.16 Common Misconfigurations 50
1.16.1 On-demand Clients Failed to Request On-demand Multicast Services 50

1 Configuring IGMP

1.1 Introduction

1.1.1 Overview

Multicast devices in a multicast network must maintain members in a multicast group that are connected to the same network segment with the multicast devices. Hosts need to be added to the multicast group dynamically so that a multicast packet can be correctly transmitted to the hosts that need the packet. With the development of multicast applications and ever growing of the multicast network, a protocol is needed to manage the members in the multicast group.

The Internet Group Management Protocol (IGMP) is a TCP/IP protocol that manages members in an IPv4 multicast group and runs on the multicast devices and hosts residing on the stub of the multicast network, creating and maintaining membership of the multicast group between the hosts and connected multicast devices. At present, IGMP is available in three versions: IGMPv1, IGMPv2, and IGMPv3. The three versions support the Any-Source Multicast (ASM) and Source-Specific Multicast (SSM) models as follows:

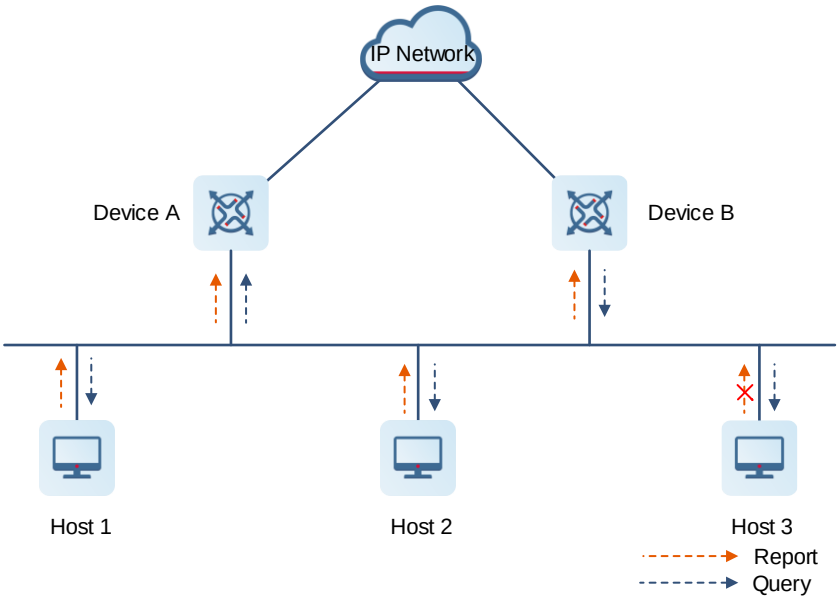
- The three versions of IGMP support the ASM model.
- IGMPv3 can be directly applied to the SSM model.
- IGMPv1 and IGMPv2 can be applied to the SSM model only when the IGMP SSM mapping technology is configured.

1.1.2 Basic Principles of IGMPv1

IGMPv1 runs on multicast devices and hosts, and manages members in a multicast group by communication between the multicast devices and hosts. The IGMP behavior of a multicast device is referred to as device behavior, and the IGMP behavior of a host is referred to as host behavior.

1. Topology of IGMPv1

Figure 1-1 Basic Topology of IGMP



The mechanism of IGMPv1 is introduced by taking the networking diagram shown in Figure 1-1. IGMP runs on hosts (host 1, host 2, and host 3) that need to be added to a multicast group and multicast devices (device A and device B) that are directly connected to the hosts.

A host sends an IGMP Report packet to a multicast device to report that the host joins the multicast group. The multicast device sends an IGMP Query packet to the host to discover the membership of the multicast group.

There are two important concepts involved in IGMPv1 running:

- Querier

  A querier is a multicast device that sends query packets and receives packets of a host. The querier is used to discover members of a multicast group on a connected network segment.

  If multiple multicast devices running IGMP reside on a network segment, one of the multicast devices is selected as a querier, and other devices function as non-queriers. In IGMPv1, a designated router (DR) elected based on the multicast routing protocols functions as a querier. For more information about DR, see *Configuring PIM-SM*.

- IGMPv1 packet

  IGMPv1 packets are classified into IGMP Query packets and IGMP Report packets.

  - IGMP Query packet: A packet sent by a multicast device to a host to query members of a multicast group on the connected network segment.
  - IGMP Report packet: A packet sent by a host to a multicast device to join a multicast group.

2. Work Process of IGMPv1

Management of multicast group members is divided into host joining a multicast group and host leaving from the multicast group.

- Group-specific queries

IGMPv1 only supports general queries, that is, query without specific multicast source and multicast group.

- Host joining a multicast group

  Communication after a host joins a multicast group falls into active joining and response query.

  - Active joining steps are as follows:

    When a host initiates a join request, the host sends an IGMP Report packet to the network segment it resides, to report that it joins a multicast group.

    The querier receives the IGMP Report packet, and creates a forwarding entry for the members in the multicast group.

  - Response query steps are as follows:

    The querier periodically sends an IGMP Query packet to all hosts in the local network segment in multicast mode.

    The host that joins the multicast group replies an IGMP Report packet to the querier in multicast mode. Other hosts that do not join the multicast group do not respond.

    The querier creates and maintains forwarding entries for the members in each multicast group.

- Host leaving a multicast group

  a A host leaves a multicast group without notifying the querier (types of packets for leaving a multicast group are not defined in IGMPv1).

  a The querier does not receive an IGMP Report packet from a multicast group in a specified period and removes the forwarding entry of this multicast group after the specified timer times out.

1.1.3 Basic Principles of IGMPv2

On the basis of the features of IGMPv1, IGMPv2 comes with a querier election mechanism and a group leaving mechanism.

1. Querier Election Mechanism

In IGMPv1, no querier election rule exists, a DR is selected as a querier using multicast routing protocols. Compared with IGMPv1, in IGMPv2, group-specific queries are supported, and a querier election mechanism is developed. The main communication process is as follows:

(1) A multicast device running IGMPv2 considers itself a querier and sends an IGMP Query (common group) packet (with destination address 224.0.0.1) to the connected network segment.

(2) Upon receiving the query packet, another multicast device compares the source IP address in the packet with the local IP address. If the source IP address is smaller than the local IP address, the querier changes to a non-querier. In this way, only one querier survives finally on the connected network segment.

(3) After the multicast device changes to a non-querier, a timer starts. The timeout period of the timer is referred to as "survival period of other querier". After timer times out, the multicast device considers itself a querier and re-initiates a querier election process. If the multicast device receives an IGMP Query packet from the querier before the timer times out, the timer is reset.

2. Group Leaving Mechanism

In IGMPv1, a host leaves a multicast group without notifying the multicast device. The multicast device detects the leaving of the host in hindsight only after the timer times out. In IGMPv2, an IGMP Leave packet is

developed. When a host leaves the multicast group, the host sends an IGMP Leave packet to notify the multicast device. The main communication process is as follows:

(1) When a host leaves a multicast group, the host sends an IGMP Leave packet to all multicast devices (destination address 224.0.0.2) on the connected network segment.

(2) Upon receiving the packet, the querier sends a query packet (the destination address is the multicast group address that the host resides) of the specified multicast group to the connected network segment.

(3) If this multicast group contains other hosts, the hosts reply an IGMP Report packet to the querier within the maximum response time upon receiving the query packet.

(4) If the querier receives replies from the hosts of the multicast group within the maximum response time, there are members in the multicast group. The querier continues maintaining the forwarding entry of this multicast group. Otherwise, the querier clears the entry of this multicast group.

1.1.4 Basic Principles of IGMPv3

On the basis of the features of IGMPv1 and IGMPv2, IGMPv3 comes with a multicast source filtering function and enhanced IGMP Query packet and IGMP Report packet.

1. Multicast Source Filtering Function

The multicast source filtering function requires the cooperation of the querier and a host.

- Host action

  A host supports filtering of a specific multicast source. When the host tries to join a multicast group, it adds information of one or multiple sources to an IGMP Report packet, expresses the correspondence between multicast group and multicast source list as (S1, S2, S3... Sn, G), and sets the mode of the multicast sources to INCLUDE or EXCLUDE. In INCLUDE mode, the host receives packets from this multicast source. In EXCLUDE mode, the host rejects packets from this multicast source. When the correspondence between the multicast group and the multicast source list changes, the IGMPv3 Membership Report message will store the relationship change in the group record field and send it to the IGMP querier.

  Group records are divided into the following types:

  - IS_IN: The filtering mode of a multicast source is INCLUDE, indicating that only packets from this specified multicast source to this multicast group are received.
  - IS_EX: The filtering mode of a multicast source is EXCLUDE, indicating that packets from this specified multicast source to this multicast group are rejected.
  - TO_IN: The filtering mode is changed from EXCLUDE to INCLUDE.
  - TO_EX: The filtering mode is changed from INCLUDE to EXCLUDE.
  - ALLOW: Multicast packets from specified multicast sources to this multicast group are received. If the filtering mode is INCLUDE, these multicast sources are added to the original source list. If the filtering mode is EXCLUDE, these multicast sources are removed from the original source list.
  - BLOCK: Multicast packets from specified multicast sources to this multicast group are rejected. If the filtering mode is INCLUDE, these multicast sources are removed from the original source list. If the filtering mode is EXCLUDE, these multicast sources are added to the original source list.

- Querier action

  When receiving multicast source information in INCLUDE mode, the multicast device acts as follows:

- If this multicast source is active, the multicast device forwards the packet from this multicast source.
- If this multicast source is inactive, the multicast device does not forward the packet from this multicast source.

When receiving multicast source information in EXCLUDE mode, the multicast device acts as follows:

- If other hosts need packets from this multicast source, the multicast device continues forwarding the packet from this multicast source.
- If this multicast source is inactive, the multicast device does not forward the packet from this multicast source.
- For other multicast sources not in EXCLUDE mode, the multicast device needs to forward the packets.

2. Enhanced IGMP Query Packet and IGMP Report Packet

- Enhanced IGMP Query Packet

  IGMPv1 supports common group query, and IGMPv2 supports specific group query. On the basis of the features of IGMPv1 and IGMPv2, IGMPv3 supports specific source and group query. Compared with IGMPv1 and IGMPv2, IGMPv3 supports specific source group queries on the basis of IGMPv1 and IGMPv2, that is, specific multicast source and multicast group queries. When receiving a report from a group member indicating a change in the correspondence between the multicast group and the multicast source list, the IGMP querier will send a specific source group query message. If a group member wants to receive multicast data traffic from any of the sources, it will send the feedback message. IGMP querier then updates the multicast source list corresponding to the group according to the feedback membership report.

- Enhanced IGMP Report Packet

  In IGMPv3, an IGMP Report packet carries records of one or multiple groups. The group records include the multicast source addresses and multicast group addresses. When a host joins a multicast group, it can explicitly express that it needs to receive or not receive multicast data traffic sent by a specific multicast source.

1.1.5 IGMP SSM Mapping

In actual network environment, most multicast devices support IGMPv3. However, there are a great number of old multicast devices that support only IGMPv1 or IGMPv2. Consequently, these old multicast devices cannot provide hosts with the multicast services in the SSM model using IGMPv3. After the IGMP SSM mapping technology is applied, the multicast devices can provide hosts with the multicast services in the SSM model without upgrading to IGMPv3.

The IGMP SSM mapping technology enables the multicast devices to support IGMPv3 by mapping the multicast group information in IGMPv1 Report packets and IGMPv2 Report packets to the multicast group information that carries specific multicast sources. After an SSM mapping rule is configured on the multicast devices, upon receiving IGMPv1 Report packets and IGMPv2 Report packets, the multicast devices process the packets based on the multicast group information in the packets.

- If a multicast group is not in the SSM group address range, the multicast services are provided in the ASM model.
- If the multicast group is in the SSM group address range and no SSM mapping rule corresponding to this multicast group is configured on the multicast device, the multicast device cannot provide multicast services in the SSM model and then discards the packets.

- If the multicast group is in the SSM group address range and an SSM mapping rule corresponding to this multicast group is configured on the multicast device, the multicast group information in the packets is converted to the multicast group information that carries specific multicast sources and the multicast device provides the multicast services in the SSM model.

1.1.6 IGMP Proxy

In a simple tree network topology, a border multicast device does not need to run complex multicast routing protocols such as Protocol Independent Multicast (PIM), you can configure the IGMP proxy function on this multicast device. This multicast device functions as an intermediate proxy between upstream IGMP querier and downstream hosts.

A device running the IGMP proxy functions as a multicast device and a host at the same time. For an upstream IGMP querier, this device plays the role of a host and executes host actions. For a downstream host, this device plays the role of an IGMP querier and executes querier actions.

1.1.7 Protocols and Standards

- RFC 1112: Host Extensions for IP Multicasting
- RFC 2236: Internet Group Management Protocol, Version 2
- RFC 3376: Internet Group Management Protocol, Version 3
- RFC 4605: Internet Group Management Protocol (IGMP)/Multicast Listener Discovery

1.2 Configuration Task Summary

IGMP configuration includes the following tasks:

- Configuring Basic Functions of IGMP
- (Optional) Configuring IGMP Parameters. All the configuration tasks below are optional. Select the configuration tasks as required.
  - Configuring an IGMP Version
  - Configuring the Maximum Response Time for Query Packets
  - Configuring Common Group Query
  - Configuring Specific Group Query
  - Configuring Survival Period of Other Querier
- (Optional) Configuring IGMP Group Filtering
- (Optional) Configuring the IGMP Proxy Function
- (Optional) Configuring the IGMP SSM Mapping Function
- (Optional) Configuring the Router Alert Option of IGMP Packets
- (Optional) Enabling Source Address Checking for IGMP Report Packets
- (Optional) Enabling Fast Leave
- (Optional) Configuring the Querier Robustness Variable
- (Optional) Adding a Static Interface to a Group
- (Optional) Simulating a Host to Join a Group

1.3 Configuring Basic Functions of IGMP

1.3.1 Overview

Basic functions of IGMP lay a foundation for other functions, enable multicast devices to run IGMP, and help collect and manage members in a multicast group.

You cannot directly enable the basic functions of IGMP. You need to enable the multicast routing function on a device to make the basic functions of IGMP take effect.

1.3.2 Restrictions and Guidelines

Before the basic functions of IGMP are enabled, you must enable the Protocol Independent Multicast Sparse Mode (PIM-SM) or Protocol Independent Multicast Dense Mode (PIM-DM) function on an interface to make the basic functions of IGMP take effect. To disable the basic functions of IGMP take effect, you need to remove the configuration of the PIM-SM or PIM-DM function.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IPv4 multicast routing.

ip mrouting-enable

The multicast routing is disabled by default.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(5) Enable the PIM-SM or PIM-DM function on an interface. Select at least one of them to configure.

- Enable PIM-DM on an interface.

  ip pim dense-mode

  The PIM-DM function is disabled on an interface by default.
- Enable PIM-SM on an interface.

  ip pim sparse-mode

  The PIM-SM function is disabled on an interface by default.

1.4 Configuring IGMP Parameters

1.4.1 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring an IGMP Version
- Configuring the Maximum Response Time for Query Packets
- Configuring Common Group Query
- Configuring Specific Group Query
- Configuring Survival Period of Other Querier

1.4.2 Configuring an IGMP Version

1. Overview

IGMP comes with three versions: IGMPv1, IGMPv2, and IGMPv3. A later version corresponds to stronger protocol functions. You can select an IGMP version based on network requirements and IGMP support of upstream and downstream devices.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure an IGMP version.

ip igmp specify-version { **1** | **2** | **3** }

IGMPv2 runs on an interface by default.

After this command is run, the IGMP function module automatically restarts.

1.4.3 Configuring the Maximum Response Time for Query Packets

1. Overview

After sending a query packet, the querier needs to receive a reply packet from a host in the specified time. Otherwise, the querier considers that no member exists in the corresponding multicast group and deletes the

group information. This specified time is referred to as the maximum response time for a query packet, and it can be set to a proper time value based on actual network topology.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the maximum response time for query packets on an interface.

ip igmp max-response-time *query-max-response*

The maximum response time for query packets on an interface is **10** seconds by default.

1.4.4 Configuring Common Group Query

1. Overview

Common group query means querying members in a multicast group. You can configure an interval for sending common group query packets.

2. Procedure

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the interval for querying a common group member.

ip igmp query-period *query-period*

The interval for querying a common group member is **125** seconds by default.

1.4.5 Configuring Specific Group Query

1. Overview

Specific group query means querying members in a specified multicast group. You can configure an interval and times for sending specific group query packets.

Upon receiving an IGMP Leave packet on an interface, the querier continuously sends specific group query packets and waits for responses from hosts. If no host replies in the specified time, the querier considers that this multicast group does not contain any member, and deletes this interface from the IGMP group member records. Then, this interface does not forward packets from this multicast group. The timeout period is a product of the sending interval of specific group query packets and the times for sending the specific group query packets.

2. Restrictions and Guidelines

The parameter of specific group query can be configured in IGMPv2 and IGMPv3. The parameter is not supported in IGMPv1.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the parameters of specific group query. Select at least one of them to configure.

- Configure the interval for sending specific group query packets.

  ip igmp last-member-query-period *last-member-query-interval*

  The interval for sending specific group query packets is **1** second by default.
- Configure the times for sending specific group query packets.

  ip igmp last-member-query-number *last-member-query-count-number*

  Specific group query packets are sent twice by default.

1.4.6 Configuring Survival Period of Other Querier

1. Overview

A querier election mechanism is introduced in IGMPv2. Based on the mechanism, non-queriers maintain a timer and the timeout period of the timer is referred to as "survival period of other querier". If the timer times out, a multicast device considers itself a querier and re-initiates a querier election request. During networking, you can determine proper survival period of other querier based on network topology configuration.

2. Restrictions and Guidelines

The querier election mechanism can be configured in IGMPv2 and IGMPv3. It is not supported in IGMPv1.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure survival period of other querier.

ip igmp other-querier-timeout *query-timeout*

The survival period of other querier is **255** seconds by default.

1.5 Configuring IGMP Group Filtering

1.5.1 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Filtering a Multicast Group
- Configuring the Maximum Number of Group Members

1.5.2 Filtering a Multicast Group

1. Overview

If you do not want hosts in the network segment where an interface resides to join a multicast group, you can configure access control list (ACL) rules on the interface as a filter. The interface filters IGMP Report packets of the hosts based on the ACL rules and allows the hosts to join specific groups.

2. Restrictions and Guidelines

- For more information about ACL, see *Configuring ACL*.
- Extended ACL rules can be configured in IGMPv3. When a received IGMP Report packet is (S1, S2, S3…Sn,

G), an ACL rule is used to match (0, G). Therefore, to normally filter (S1, S2, S3…Sn, G), a (0, G) extended ACL rule must be configured.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Filter out groups that you do not want a host to join.

ip igmp restricted-group { *acl-name* | *acl-number* }

Hosts can join any group by default.

1.5.3 Configuring the Maximum Number of Group Members

1. Overview

To configure the maximum number of group members means configuring the maximum number of multicast groups that a host can join.

The configuration can be completed in interface or global configuration mode. If the configuration is completed in global configuration mode, the number of group members for a multicast device is limited. If the configuration is completed in interface configuration mode, the number of group members on an interface is limited. If the number of group member records exceeds the interface limit or global limit, subsequent received IGMP Report packets are ignored.

2. Restrictions and Guidelines

- Group member records that are generated in EXCLUDE mode list are not counted in the group members.
- Interface and global limits can be configured separately. If the global limit is smaller than the interface limit, the global limit prevails.

3. Procedure (Global Configuration Mode)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of IGMP group members in global configuration mode.

ip igmp max-member-record *limit-number* [**except** *acl-name* | **except** *acl-number*]

The maximum number of IGMP group members is **64000** in global configuration mode by default.

4. Procedure (Interface Configuration Mode)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the maximum number of IGMP group members on an interface.

ip igmp max-member-record *limit-number* [**except** *acl-name* | **except** *acl-number*]

The maximum number of IGMP group members on an interface is **4000** by default.

1.6 Configuring the IGMP Proxy Function

1.6.1 Overview

In a simple tree network topology, you can configure the IGMP proxy function and use it as an intermediate proxy between upstream IGMP querier and downstream hosts without running complex multicast routing protocols such as PIM.

On the device that runs the IGMP proxy, interfaces are classified into proxy service interfaces and MRoute proxy interfaces.

The proxy service interfaces execute host behaviors, receive IGMP packets from upstream devices, and forward the packets to the MRoute proxy interfaces.

The MRoute proxy interfaces execute device behaviors, receive IGMP Report packets from downstream devices, and forward the packets to the proxy service interfaces.

1.6.2 Restrictions and Guidelines

- The basic functions of IGMP must be configured.
- The interfaces connected to upstream devices must be configured as proxy service interfaces and the interfaces connected to hosts must be configured as MRoute proxy interfaces. The two types of interfaces must be used at the same time. If only one type of interfaces are configured, the IGMP proxy function will be abnormal.

- If the proxy service interfaces are switched over from L3 interface to L2 interface, the routing proxy configured on the MRoute proxy interfaces is automatically disabled.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the IGMP proxy function. The configuration steps below are mutually exclusive. Please configure only one task.

- Configure an interface as a proxy service interface.

  ip igmp enable-proxy-service

  No interface is configured as a proxy service interface by default.

  The interface directly connected to an upstream device can be configured as a proxy service interface unless otherwise specified. A device supports 32 proxy service interfaces at most. When a proxy service interface receives an IGMP Report packet, the proxy service interface makes response based on the IGMP group member records.

- Configure an interface as an MRoute proxy interface.

  ip igmp enable-mrouting-proxy *interface-type interface-number*

  No interface is configured as an MRoute proxy interface by default.

  The interface directly connected to a downstream device can be configured as an MRoute proxy interface unless otherwise specified.

1.7 Configuring the IGMP SSM Mapping Function

1.7.1 Overview

At present, the SSM multicast model (supporting filtering multicast sources) is supported in IGMPv3 only. To enable the SSM multicast model to be supported in IGMPv1 and IGMPv2, you can configure the IGMP SSM mapping function on a device, and add multicast group information of specified multicast sources to IGMPv1 and IGMPv2 packets so that the hosts can use the SSM multicast services provided by the device on the basis of running IGMPv1 or IGMPv2.

1.7.2 Restrictions and Guidelines

- When a device receives an IGMPv1 or IGMPv2 Report packet, the device adds the static mapping source address.
- The IGMP SSM mapping function applies to only IGMPv1 and IGMPv2 packets. You are advised to enable IGMPv3 on an interface to ensure that IGMP packets of different versions can obtain the SSM multicast services.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IGMP SSM mapping function.

ip igmp ssm-policy enable

The IGMP SSM mapping function is disabled by default.

(4) Configure a static mapping entry.

ip igmp ssm-policy static { *acl-name* | *acl-number* } *source-address*

No SSM static mapping entry is configured by default.

1.8 Configuring the Router Alert Option of IGMP Packets

1.8.1 Overview

The Router Alert option in an IP packet header is used to warn a router device to check content in the IP packet. The router device can determine whether to add this option to a packet to be sent and whether to check this option in a received packet. The device does not check this option by default.

After the Router Alert option is configured, if an IGMP packet carries the Router Alert option, the multicast device checks the IGMP packet and updates the controlled content. If the IGMP packet does not carry the Router Alert option, the multicast device does not check the content in the packet and directly forwards the packet.

1.8.2 Restrictions and Guidelines

After the Router Alert option is configured, the multicast device must add this option to a packet to be sent and check this option in a received packet. Otherwise, an error occurs in IGMP packet receiving.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure to check the Router Alert option in an IGMP packet.

ip igmp enforce-router-alert

The Router Alert option in an IGMP packet is not checked by default.

After this option is configured, IGMP packets that do not carry the Router Alert option are discarded.

(4) Add the Router Alert option to an IGMP packet to be sent.

ip igmp send-router-alert

The Router Alert option is not added to an IGMP packet to be sent by default.

1.9 Enabling Source Address Checking for IGMP Report Packets

1.9.1 Overview

To prevent false packets, you can enable the source address checking function for IGMP packets so that a multicast device receives only IGMP Report packets whose source addresses are on the same network segment as the receiving interface.

1.9.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable source address checking for IGMP Report packets.

ip igmp enforce-source-subnet

The source address checking function for IGMP Report packets is disabled by default.

1.10 Enabling Fast Leave

1.10.1 Overview

After the fast leave function is enabled, if a device receives an IGMP Leave packet of a specified group, the device directly deletes this interface from the group member records to shorten the leave latency.

1.10.2 Restrictions and Guidelines

This function applies to an interface that runs IGMPv2 or IGMPv3 and that is connected to only one host.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Enable the fast leave function.

ip igmp leave-now group-acl { *acl-name* | *acl-number* }

The fast leave function on an interface is disabled by default.

1.11 Configuring the Querier Robustness Variable

1.11.1 Overview

The querier robustness variable is used to calculate the aging time of a forwarding entry after a device receives an IGMP Report packet. Aging time = Query interval × Robustness variable + 10

1.11.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Configure the querier robustness variable.

ip igmp robustness-variable *robustness-variable-number*

The default querier robustness variable is **2**.

1.12 Adding a Static Interface to a Group

In normal cases, an interface can join a multicast group only after the interface receives an IGMP Report packet from the multicast group.

If you manually add an interface to a group, the interface can be added to the group and exchanges multicast group information with the PIM router without IGMP packet exchange.

1.12.1 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Add a static interface to a group.

ip igmp static-group-entries *ipv4-group-address*

No static interface is added to an IPv4 multicast group by default.

1.13 Simulating a Host to Join a Group

1.13.1 Overview

By simulating a host to join a group, you can simulate host behaviors and send a Join packet to the upstream devices to join the group. This function is used to test the multicast function.

1.13.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*

- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Simulate a host to join a group.

ip igmp add-to-group *ipv4-group-address*

The behavior of simulating a host to join a multicast group is not performed by default.

1.14 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information of the **debugging** command occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup ip igmp-cache dynamic-group	Clears dynamic group member records in the IGMP cache.
cleanup ip igmp-cache interface *interface-type interface-number*	Clears all IGMP statistics and group member records on the interface.
display ip igmp-info groups [*interface-type interface-number*] [*ipv4-group-address*] [**detail**]	Displays groups directly connected to the device and group information learned from IGMP.
display ip igmp-info interface [*interface-type interface-number*]	Displays IGMP configurations of the interface.
display ip igmp-info ssm-policy [*ipv4-group-address*]	Displays IGMP SSM mapping information.
display debugging	Displays the status of the IGMP debugging switch.
debugging ip igmp-info all	Debugs all IGMP information.
debugging ip igmp-info decode	Debugs IGMP packet resolution.
debugging ip igmp-info encode	Debugs IGMP packet encoding.

Command	Purpose
debugging ip igmp-info events	Debugs IGMP event information.
debugging ip igmp-info fsm	Debugs IGMP Finite State Machine (FSM).
debugging ip igmp-info tib	Debugs IGMP state machine information.
debugging ip igmp-info warning	Debugs IGMP warning.

1.15 Configuration Examples

1.15.1 Configuring Basic Functions of IGMP

1. Requirements

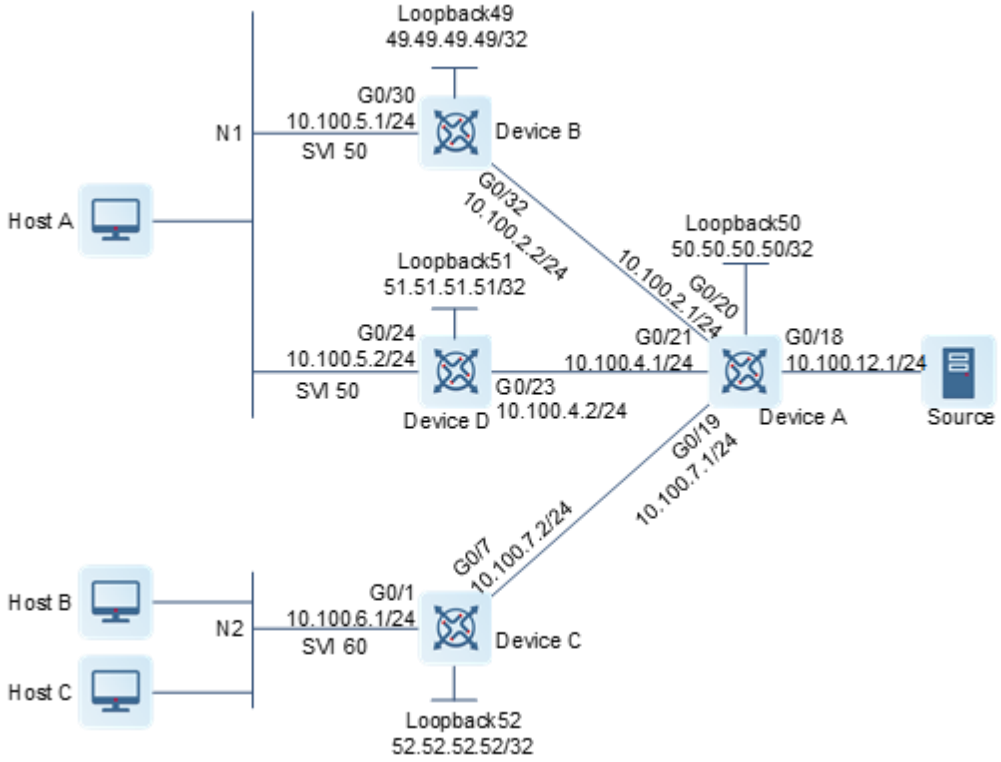
As shown in Figure 1-2, Host A, Host B and Host C on the shared networks N1 and N2 in the IP network want to receive video traffic from the source Source in multicast mode. This requires the static multicast group 225.0.0.10 to be configured on N1, while hosts on N2 can only join the multicast group 225.0.0.11.

- Run OSPF in the network to achieve reachability of any node through unicast routing.
- Device B and Device D are connected to N1 through SVI50. Because the IP address of Device B is lower than that of Device D, Device B acts as a querier, while Host A is the receiver of multicast traffic from N1.
- Device C is connected to N2 through SVI60, and Host B is the receiver of multicast traffic from N2.

Device B, Device C and Device D connect to other devices on the IP PIM network through Layer 3 interfaces.

2. Topology

Figure 1-2 Topology of IGMP with Basic Functions



3. Notes

- Configure an IP unicast routing protocol (OSPF) on multicast devices, and enable unicast routes to be reachable to the loopback interface.
- Enable the IP multicast routing function on all multicast devices.
- Enable the PIM-SM function on device interconnection interfaces and interfaces for connecting hosts and multicast sources.
- Configure the static multicast group function on device interfaces connecting N1.
- Configure the multicast group access control function on device interfaces connecting N2.

4. Procedure

(1) Configure IP addresses for interfaces on each device, and enable the unicast routing protocol.

Configure the IP address for each interface on Device A, and enable the OSPF protocol to realize inter-network communication.

```
DeviceA> privilege
DeviceA#config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/18
DeviceA(cfg-if-GigabitEthernet 0/18)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/18)# ip address 10.100.12.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/18)# exit
```

```
DeviceA(cfg)# interface gigabitethernet 0/19
DeviceA(cfg-if-GigabitEthernet 0/19)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/19)# ip address 10.100.7.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/19)# exit
DeviceA(cfg)# interface gigabitethernet 0/20
DeviceA(cfg-if-GigabitEthernet 0/20)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/20)# ip address 10.100.2.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/20)# exit
DeviceA(cfg)# interface gigabitethernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/21)# ip address 10.100.4.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface loopback 50
DeviceA(cfg-if-Loopback 53)# ip address 50.50.50.50 255.255.255.255
DeviceA(cfg-if-Loopback 53)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.2.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.4.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.7.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 50.50.50.50 0.0.0.0 area 0
DeviceA(cfg-router)# exit
```

Configure the IP address for each interface on Device B, and enable the OSPF protocol to realize inter-network communication.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan 50
DeviceB(cfg-vlan)# exit
DeviceB(cfg)# interface vlan 50
DeviceB(cfg-if-VLAN 50)# ip address 10.100.5.1 255.255.255.0
DeviceB(cfg-if-VLAN 50)# exit
DeviceB(cfg)# interface gigabitethernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# switch-port access vlan 50
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface gigabitethernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/31)# switch-port trunk native vlan 50
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface gigabitethernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/32)# ip address 10.100.2.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface loopback 49
DeviceB(cfg-if-Loopback 49)# ip address 49.49.49.49 255.255.255.255
DeviceB(cfg-if-Loopback 49)# exit
DeviceB(cfg)# router-protocol ospf 49
```

```
DeviceB(cfg-router)# network-advertise 10.100.2.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 49.49.49.49 0.0.0.0 area 0
DeviceB(cfg-router)# exit
```

Configure the IP address for each interface on Device C, and enable the OSPF protocol to realize inter-network communication.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# vlan 60
DeviceC(cfg-vlan)# exit
DeviceC(cfg)# interface vlan 60
DeviceC(cfg-if-VLAN 60)# ip address 10.100.6.1 255.255.255.0
DeviceC(cfg-if-VLAN 60)# exit
DeviceC(cfg)# interface gigabitethernet 0/1
DeviceC(cfg-if-GigabitEthernet 0/1)# switch-port access vlan 60
DeviceC(cfg-if-GigabitEthernet 0/1)# exit
DeviceC(cfg)# interface gigabitethernet 0/7
DeviceC(cfg-if-GigabitEthernet 0/7)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/7)# ip address 10.100.7.2 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/7)# exit
DeviceC(cfg)# interface loopback 52
DeviceC(cfg-if-Loopback 52)# ip address 52.52.52.52 255.255.255.255
DeviceC(cfg-if-Loopback 52)# exit
DeviceC(cfg)# router-protocol ospf 49
DeviceC(cfg-router)# network-advertise 10.100.6.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.7.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 52.52.52.52 0.0.0.0 area 0
DeviceC(cfg-router)# exit
```

Configure the IP address for each interface on Device D, and enable the OSPF protocol to realize inter-network communication.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# vlan 50
DeviceD(cfg-vlan)# exit
DeviceD(cfg)# interface GigabitEthernet 0/23
DeviceD(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/23)# ip address 10.100.4.2 255.255.255.0
DeviceD(cfg-if-GigabitEthernet 0/23)# exit
DeviceD(cfg)# interface GigabitEthernet 0/24
DeviceD(cfg-if-GigabitEthernet 0/24)# switch-port mode trunk
DeviceD(cfg-if-GigabitEthernet 0/24)# switch-port trunk native vlan 50
DeviceD(cfg-if-GigabitEthernet 0/24)# exit
DeviceD(cfg)# interface Loopback 51
DeviceD(cfg-if-Loopback 51)# ip address 51.51.51.51 255.255.255.255
DeviceD(cfg-if-Loopback 51)# exit
```

```
DeviceD(cfg)# router-protocol ospf 49
DeviceD(cfg-router)# network-advertise 10.100.4.0 0.0.0.255 area 0
DeviceD(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceD(cfg-router)# network-advertise 51.51.51.51 0.0.0.0 area 0
DeviceD(cfg-router)# exit
```

(2) Enable the IP multicast routing function on all devices, and enable the PIM-SM function on device interconnection interfaces and interfaces for connecting hosts and multicast sources.

Enable the IP multicast routing function and the PIM-SM function on Device A.

```
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface gigabitethernet 0/18
DeviceA(cfg-if-GigabitEthernet 0/18)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/18)# exit
DeviceA(cfg)# interface GigabitEthernet 0/19
DeviceA(cfg-if-GigabitEthernet 0/19)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/19)# exit
DeviceA(cfg)# interface GigabitEthernet 0/20
DeviceA(cfg-if-GigabitEthernet 0/20)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/20)# exit
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
```

Configure RP. Configure C-BSR and C-RP on Device A.

```
DeviceA(cfg)# ip pim c-bsr Loopback 50
DeviceA(cfg)# ip pim c-rp Loopback 50
```

Enable the IP multicast routing function and the PIM-SM function on Device B.

```
DeviceB(cfg)# ip mrouting-enable
DeviceB(cfg)# interface VLAN 50
DeviceB(cfg-if-VLAN 50)# ip pim sparse-mode
DeviceB(cfg-if-VLAN 50)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ip pim sparse-mode
DeviceB(cfg-if-Loopback 49)# exit
```

Enable the IP multicast routing function and the PIM-SM function on Device C.

```
DeviceC(cfg)# ip mrouting-enable
DeviceC(cfg)# interface VLAN 60
DeviceC(cfg-if-VLAN 60)# ip pim sparse-mode
DeviceC(cfg-if-VLAN 60)# exit
DeviceC(cfg)# interface GigabitEthernet 0/7
DeviceC(cfg-if-GigabitEthernet 0/7)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/7)# exit
DeviceC(cfg)# interface Loopback 52
```

```
DeviceC(cfg-if-Loopback 52)# ip pim sparse-mode
DeviceC(cfg-if-Loopback 52)# exit
```

Enable the IP multicast routing function and the PIM-SM function on Device D.

```
DeviceD(cfg)# ip mrouting-enable
DeviceD(cfg)# interface GigabitEthernet 0/23
DeviceD(cfg-if-GigabitEthernet 0/23)# ip pim sparse-mode
DeviceD(cfg-if-GigabitEthernet 0/23)# exit
DeviceD(cfg)# interface vlan 50
DeviceD(cfg-if-vlan 50)# ip pim sparse-mode
DeviceD(cfg-if-vlan 50)# exit
DeviceD(cfg)# interface Loopback 51
DeviceD(cfg-if-Loopback 51)# ip pim sparse-mode
DeviceD(cfg-if-Loopback 51)# exit
```

(3) Configure the static multicast group function on device interfaces connecting N1.

Configure the static multicast group function on the SVI50 interface of Device B, and configure the static multicast group address as 225.0.0.10.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface VLAN 50
DeviceB(cfg-if-VLAN 50)# ip igmp static-group-entries 225.0.0.10
DeviceB(cfg-if-VLAN 50)# exit
```

Configure the static multicast group function on the SVI50 interface of Device D, and configure the static multicast group address as 225.0.0.10.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# interface VLAN 50
DeviceD(cfg-if-VLAN 50)# ip igmp static-group-entries 225.0.0.10
DeviceD(cfg-if-VLAN 50)# exit
```

(4) Configure the multicast group access control function on device interfaces connecting N2.

Configure the multicast group access control function on the SVI60 interface of Device C. Member hosts on N2 can only join the multicast group 225.0.0.11.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ip access-control-list standard ip-group
DeviceC(cfg-stdacl)# permit 225.0.0.11 0.0.0.0
DeviceC(cfg-stdacl)# exit
DeviceC(cfg)# interface VLAN 60
DeviceC(cfg-if-VLAN 60)# ip igmp restricted-group ip-group
DeviceC(cfg-if-VLAN 60)# exit
```

5. Verification

(1) Run the **display ip igmp-info interface** command to check the operation of IGMP configuration on each device. It can be seen that Device B is the querier, because on N1, the IP address of SVI50 on Device B is low.

Run the **display ip igmp-info interface** command. It can be seen that Device B is the querier.

```
DeviceB# display ip igmp-info interface VLAN 50
Interface VLAN 50 (Index 4146)
IGMP Enabled, Active, Querier, Version 2 (default)
Internet address is 10.100.5.1
IGMP interface limit is 4000
IGMP interface has 1 group-record states
IGMP interface has 1 static-group-entries records
IGMP activity: 8 joins, 0 leaves
IGMP query interval is 125 seconds
IGMP querier timeout is 255 seconds
IGMP max query response time is 10 seconds
Last member query response interval is 10
Last member query count is 2
Group Membership interval is 260 seconds
Robustness Variable is 2
 Querier timer on, remain time:25
DeviceB#
```

Run the **display ip igmp-info interface** command. It can be seen that Device D is not selected as the querier.

```
DeviceD# display ip igmp-info interface VLAN 50
Interface VLAN 50 (Index 4146)
IGMP Active, Non-Querier, Version 2 (default)
Internet address is 10.100.5.2
IGMP interface limit is 4000
IGMP interface has 0 group-record states
 IGMP interface has 0 static-group records
 IGMP activity: 0 joins, 0 leaves
 IGMP querying router is 10.100.5.1
 IGMP query interval is 125 seconds
 IGMP querier timeout is 255 seconds
 IGMP max query response time is 10 seconds
 Last member query response interval is 10
 Last member query count is 2
 Group Membership interval is 260 seconds
 Robustness Variable is 2
 Querier timer OFF
DeviceB#
```

(2) Run the **display ip pim-info sparse-mode mroute** command to check multicast routing table entries on each device.

Multicast routing table entries on Device A are shown as follows. There are (*,225.0.0.10), (10.100.12.2, 225.0.0.10) and (*,225.0.0.11), (10.100.12.2, 225.0.0.11) entries.

```
DeviceA# display ip pim-info sparse-mode mroute
IP Multicast Routing Table
```

```
(*,*,RP) Entries: 0
(*,G) Entries: 2
(S,G) Entries: 2
(S,G,rpt) Entries: 2
FCR Entries: 0
REG Entries: 0

(*, 225.0.0.10)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: JOINED
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  j  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
FCR:

(10.100.12.2, 225.0.0.10)
RPF nbr: 10.100.11.2
RPF idx: GigabitEthernet 0/18
SPT bit: 1
```

```
Upstream State: JOINED
jt_timer expires in 6 seconds
kat expires in 177 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  .  .  o  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

(10.100.12.2, 225.0.0.10, rpt)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: PRUNED
```

```
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Pruned
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  .  .  o  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

(*, 225.0.0.11)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: JOINED
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  j  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
FCR:

(10.100.12.2, 225.0.0.11)
RPF nbr: 10.100.11.2
RPF idx: GigabitEthernet 0/18
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 18 seconds
kat expires in 144 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
```

```
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
*Jun 24 09:19:18: %CLI-5-EXEC_CMD: Executed from vty0(172.26.1.78) command: display ip pim-info
sparse-mode mroute
Outgoing
0 . . . . o . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

(10.100.12.2, 225.0.0.11, rpt)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: PRUNED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Pruned
0 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Outgoing
0 . . . . o . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
```

```
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

DeviceA#
```

Multicast routing table entries on Device B are shown as follows. There are (*, 225.0.0.10), and FCR entries with a source address of 10.100.12.2.

```
DeviceB# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 1
(S,G) Entries: 0
(S,G,rpt) Entries: 0
FCR Entries: 1
REG Entries: 0

(*, 225.0.0.10)
RP: 50.50.50.50
RPF nbr: 10.100.2.1
RPF idx: GigabitEthernet 0/32
Upstream State: NOT JOINED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  i  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
FCR:
Source: 10.100.12.2
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Outgoing
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
KAT timer running, 197 seconds remaining
  Packet count 1

DeviceB#
```

Multicast routing table entries on Device C are shown as follows. There are (*, 225.0.0.11), and FCR entries with a source address of 10.100.12.2.

```
DeviceC# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 1
(S,G) Entries: 0
(S,G,rpt) Entries: 0
FCR Entries: 1
REG Entries: 0

(*, 225.0.0.11)
RP: 50.50.50.50
RPF nbr: 10.100.7.1
RPF idx: GigabitEthernet 0/7
Upstream State: JOINED
jt_timer expires in 56 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  i  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
8 . . . . . .
9 . . . . . .
10 . . . . . .
11 . . . . . .
12 . . . . . .
13 . . . . . .
14 . . . . . .
15 . . . . . .
Joined
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
8 . . . . . .
9 . . . . . .
10 . . . . . .
11 . . . . . .
12 . . . . . .
13 . . . . . .
14 . . . . . .
15 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
8 . . . . . .
9 . . . . . .
10 . . . . . .
11 . . . . . .
12 . . . . . .
13 . . . . . .
14 . . . . . .

```
15 . . . . . .
FCR:
Source: 10.100.12.2
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Outgoing
0 . . 0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
8 . . . . . .
9 . . . . . .
10 . . . . . .
11 . . . . . .
12 . . . . . .
13 . . . . . .
14 . . . . . .
15 . . . . . .
KAT timer running, 8 seconds remaining
 Packet count 1

DeviceC#
```

Run the display ip igmp-info groups command to check that SVI50 interfaces on Device B and Device D join the static multicast group 225.0.0.10.

```
DeviceB# display ip igmp-info groups
IGMP Connected Group Membership
Group Address      Interface                Uptime      Expires     Last Reporter
225.0.0.10         VLAN 50                              01:48:10    stopped     0.0.0.0
DeviceB#
```

```
DeviceD# display ip igmp-info groups
IGMP Connected Group Membership
Group Address      Interface                Uptime      Expires     Last Reporter
225.0.0.10         VLAN 50                              01:01:55    stopped     0.0.0.0
DeviceD#
```

The host on SVI60 joins multicast groups 225.0.0.10 and 225.0.0.11. Run the **display ip igmp-info groups** command, and it can be seen that the host on SVI60 can only join the multicast group 225.0.0.11.

```
DeviceC# display ip igmp-info groups
IGMP Connected Group Membership
Group Address      Interface                Uptime      Expires     Last Reporter
225.0.0.11         VLAN 60                              01:16:11    00:04:03    10.100.6.2
DeviceC#
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
interface GigabitEthernet 0/18
 unset switch-port
 ip address 10.100.11.1 255.255.255.0
ip pim sparse-mode
!
interface GigabitEthernet 0/19
 unset switch-port
 ip address 10.100.7.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/20
 unset switch-port
 ip address 10.100.2.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.4.1 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 50
 ip address 50.50.50.50 255.255.255.255
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.2.0 0.0.0.255 area 0
 network-advertise 10.100.4.0 0.0.0.255 area 0
 network-advertise 10.100.7          .0 0.0.0.255 area 0
 network-advertise 10.100.12.0 0.0.0.255 area 0
 network-advertise 50.50.50.50 0.0.0.0 area 0
!
ip pim c-bsr Loopback 50
ip pim c-rp Loopback 50
!
end
```

- Device B configuration file

```
hostname DeviceB
!
```

```
ip mrouting-enable
!
vlan range 1,50
!
interface GigabitEthernet 0/30
 switch-port access vlan 50
!
interface GigabitEthernet 0/32
 unset switch-port
 ip address 10.100.2.2 255.255.255.0
 ip pim sparse-mode
!
interface TenGigabitEthernet 0/31
 switch-port mode trunk
 switch-port trunk native vlan 50
!
interface Loopback 49
 ip address 49.49.49.49 255.255.255.255
 ip pim sparse-mode
!
interface VLAN 50
 ip address 10.100.5.1 255.255.255.0
 ip igmp static-group-entries 225.0.0.10
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.2.0 0.0.0.255 area 0
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 49.49.49.49 0.0.0.0 area 0
!
End
```

- Device C configuration file

```
hostname DeviceC
!
ip access-control-list standard ip-group
 20 permit host 225.0.0.11
!
ip mrouting-enable
!
vlan range 1,60
!
interface GigabitEthernet 0/1
 switch-port access vlan 60
!
interface GigabitEthernet 0/7
```

```
 unset switch-port
 ip address 10.100.7.2 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 52
 ip address 52.52.52.52 255.255.255.255
 ip pim sparse-mode
!
interface VLAN 60
 ip address 10.100.6.1 255.255.255.0
 ip igmp restricted-group ip-group
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.6.0 0.0.0.255 area 0
 network-advertise 10.100.7.0 0.0.0.255 area 0
 network-advertise 52.52.52.52 0.0.0.0 area 0
!
end
```

- Device D configuration file

```
hostname DeviceD
!
ip mrouting-enable
!
interface GigabitEthernet 0/23
 unset switch-port
 ip address 10.100.4.2 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/24
 switch-port mode trunk
 switch-port trunk native vlan 50
!
interface Loopback 51
 ip address 51.51.51.51 255.255.255.255
 ip pim sparse-mode
!
interface VLAN 50
 ip address 10.100.5.2 255.255.255.0
 ip igmp static-group-entries 225.0.0.10
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.4.0 0.0.0.255 area 0
```

```
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 51.51.51.51 0.0.0.0 area 0
!
end
```

7. Common Errors

- Multicast routing is disabled on multicasting devices on the network.
- No multicast interface is available on the network.

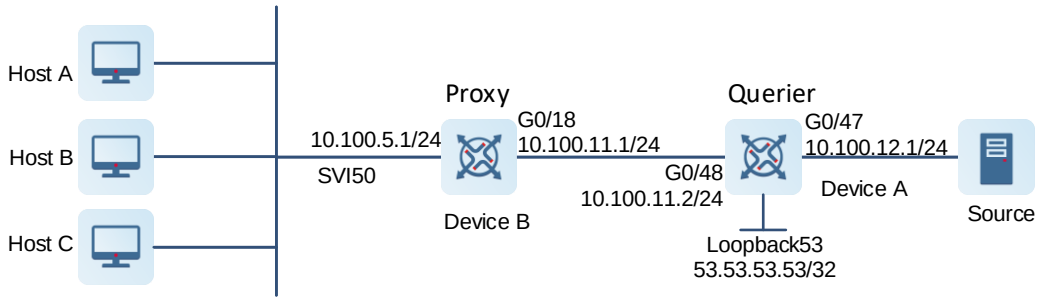
1.15.2 Configuring the IGMP Proxy Function

1. Requirements

As shown in Figure 1-3, member hosts Host A and Host B on the IP stub network want to receive video traffic from the source Source in multicast mode, and require Device B to use the IGMP proxy function to maintain multicast membership and normal data forwarding without running PIM-SM. PIM-SM is running on the network. The member hosts Host A and Host B on the stub network receive the multicast traffic sent by the source Source through the multicast group addresses 225.0.0.10 and 225.0.0.11. IGMP proxy function is configured on Device B to maintain group membership and maintain normal multicast data forwarding without running PIM-SM.

2. Topology

Figure 1-3 Topology of IGMP Proxy Configuration



3. Notes

- Configure an IP unicast routing protocol (OSPF) on multicast devices, and enable unicast routes to be reachable to the loopback interface.
- Enable the IP multicast routing function on all multicast devices, and the PIM-SM function on device interconnection interfaces and interfaces for connecting hosts and multicast sources.
- Configure a proxy service interface.
- Configure an MRoute proxy interface.

4. Procedure

(1) Configure an IP unicast routing protocol (OSPF) on multicast devices, and enable unicast routes to be reachable to the loopback interface.

Configure the IP address for each interface on Device A, and enable the OSPF protocol to realize inter-network communication.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/47)# ip address 10.100.12.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/48)# ip address 10.100.11.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# interface loopback 53
DeviceA(cfg-if-Loopback 53)# ip address 53.53.53.53 255.255.255.255
DeviceA(cfg-if-Loopback 53)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 53.53.53.53 0.0.0.0 area 0
DeviceA(cfg-router)# exit
```

Configure the IP address for each interface on Device B, and enable the OSPF protocol to realize inter-network communication.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan 50
DeviceB(cfg)# interface VLAN 50
DeviceB(cfg-if-VLAN 50)# ip address 10.100.5.1 255.255.255.0
DeviceB(cfg-if-VLAN 50)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 50
DeviceB(cfg-if-GigabitEthernet 0/19)#interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 50
DeviceB(cfg-if-GigabitEthernet 0/20)#interface GigabitEthernet 0/21
DeviceB(cfg-if-GigabitEthernet 0/21)# switch-port access vlan 50
DeviceB(cfg-if-GigabitEthernet 0/21)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/18)# ip address 10.100.11.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceB(cfg-router)# exit
```

(2) Enable the IP multicast routing function on all multicast devices, and the PIM-SM function on device interconnection interfaces and interfaces for connecting hosts and multicast sources.

Enable the IP multicast routing function and the PIM-SM function on Device A.

```
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# interface loopback 53
DeviceA(cfg-if-Loopback 0)# ip pim sparse-mode
DeviceA(cfg-if-Loopback 0)# exit
```

Configure RP. Configure C-BSR and C-RP on Device A.

```
DeviceA(cfg)# ip pim c-bsr Loopback 53
DeviceA(cfg)# ip pim c-rp Loopback 53
```

(3) Enable basic IGMP functions on Device B.

```
DeviceB(cfg)# ip mrouting-enable
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface vlan 50
DeviceB(cfg-if-VLAN 50)# ip pim sparse-mode
DeviceB(cfg-if-VLAN 50)# exit
```

(4) Configure a proxy service interface.

Configure the proxy service on interface GigabitEthernet 0/24 on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)#ip igmp enable-proxy-service
DeviceB(cfg-if-GigabitEthernet 0/18)#exit
```

(5) Configure an MRoute proxy interface.

Configure the MRoute proxy function on interface SVI50 on Device B.

```
DeviceB(cfg)# interface VLAN 50
DeviceB(cfg-if-VLAN 50)# ip igmp enable-mrouting GigabitEthernet 0/18
```

5. Verification

Run the **display ip igmp-info groups** command on Device B to check proxy entries.

```
DeviceB# display ip igmp-info groups
IGMP Connected Group Membership
Group Address      Interface                Uptime      Expires     Last Reporter
225.0.0.10         VLAN 50                              00:17:44    00:03:13    10.100.5.2
225.0.0.11         VLAN 50                              00:17:42    00:03:08    10.100.5.4
```

```
IGMP Proxy-server Connected Group Membership
Group Address      Interface                Uptime
225.0.0.10         GigabitEthernet 0/18     00:12:49
225.0.0.11         GigabitEthernet 0/18     00:12:49
DeviceB#
```

Run the **display ip igmp-info groups** command on Device A to verify that member addresses that join the multicast groups 225.0.0.10 and 225.0.0.11 are the IP addresses of Device B.

```
DeviceA# display ip igmp-info groups
IGMP Connected Group Membership
Group Address      Interface                Uptime      Expires     Last Reporter
225.0.0.10         GigabitEthernet0/48                  00:14:16    00:03:06    10.100.11.1
225.0.0.11         GigabitEthernet0/48                  00:14:16    00:03:06    10.100.11.1
DeviceA#
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
interface GigabitEthernet 0/47
 unset switch-port
 ip address 10.100.12.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/48
 unset switch-port
 ip address 10.100.11.2 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 53
 ip address 53.53.53.53 255.255.255.255
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 10.100.12.0 0.0.0.255 area 0
 network-advertise 53.53.53.53 0.0.0.0 area 0
!
ip pim c-bsr Loopback 53
ip pim c-rp Loopback 53
!
End
```

- Device B configuration file

```
hostname DeviceB
```

```
!
ip mrouting-enable
!
vlan range 1,50
!
interface GigabitEthernet 0/18
 unset switch-port
 ip address 10.100.11.1 255.255.255.0
ip igmp enable-proxy-service
 ip pim sparse-mode
!
interface GigabitEthernet 0/19
 switch-port access vlan 50
!
interface GigabitEthernet 0/20
 switch-port access vlan 50
!
interface GigabitEthernet 0/21
 switch-port access vlan 50
!
interface VLAN 50
 ip address 10.100.5.1 255.255.255.0
 ip igmp enable-mrouting-proxy GigabitEthernet 0/18
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 10.100.11.0 0.0.0.255 area 0
!
End
```

7. Common Errors

- The basic functions of IGMP are not configured on the proxy service interface.

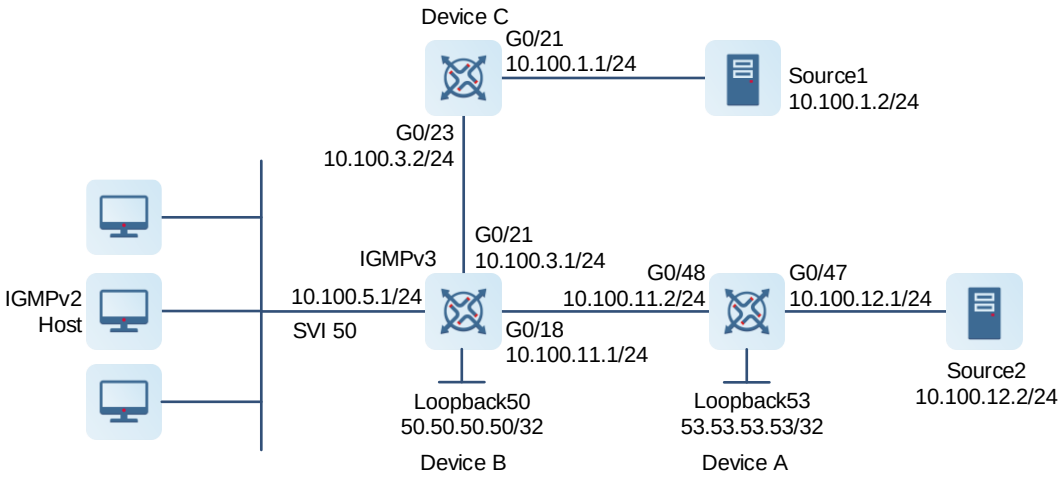
1.15.3 Configuring SSM-MAP

1. Requirements

As shown in Figure 1-4, on the IP network, the member host Host wants to receive video traffic from Source 2, but not video traffic from Source 1 without IGMPv3. The SSM mode can be used to provide multicast services. The SSM group address range on the network is 232.0.0.0/8 by default. The member host runs the IGMP2 protocol, and does not support IGMPv3. The host wants to receive only the multicast traffic sent by Source 2, not multicast traffic sent by Source 1.

2. Topology

Figure 1-4 Topology of SSM-MAP Configuration



3. Notes

- Configure an IP unicast routing protocol (OSPF) on multicast devices, and enable unicast routes to be reachable to the loopback interface.
- Enable the IP multicast routing function on all multicast devices, and the PIM-SM function on device interconnection interfaces and interfaces for connecting hosts and multicast sources.
- Configure the SSM group address range.
- Enable the SSM-MAP function.
- Configure the SSM-MAP static linked list.

4. Procedure

(1) Configure an IP unicast routing protocol (OSPF) on multicast devices, and enable unicast routes to be reachable to the loopback interface.

Configure the IP address for each interface on Device A, and enable the OSPF protocol to realize inter-network communication.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/47)# ip address 10.100.12.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/48)# ip address 10.100.11.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# interface loopback 53
DeviceA(cfg-if-Loopback 53)# ip address 53.53.53.53 255.255.255.255
```

```
DeviceA(cfg-if-Loopback 53)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 53.53.53.53 0.0.0.0 area 0
DeviceA(cfg-router)# exit
```

Configure the IP address for each interface on Device B, and enable the OSPF protocol to realize inter-network communication.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan 50
DeviceB(cfg)# interface VLAN 50
DeviceB(cfg-if-VLAN 50)# ip address 10.100.5.1 255.255.255.0
DeviceB(cfg-if-VLAN 50)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 50
DeviceB(cfg-if-GigabitEthernet 0/19)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 50
DeviceB(cfg-if-GigabitEthernet 0/20)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/21)# switch-port access vlan 50
DeviceB(cfg-if-GigabitEthernet 0/21)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/18)# ip address 10.100.11.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/21
DeviceB(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/21)# ip address 10.100.3.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface loopback 50
DeviceA(cfg-if-Loopback 50)# ip address 50.50.50.50 255.255.255.255
DeviceA(cfg-if-Loopback 50)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.3.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 50.50.50.50 0.0.0.0 area 0
DeviceB(cfg-router)# exit
```

Configure the IP address for each interface on Device C, and enable the OSPF protocol to realize inter-network communication.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ip address 10.100.1.1 255.255.255.0
```

```
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface GigabitEthernet 0/23
DeviceC(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/23)# ip address 10.100.3.2 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/23)# exit
DeviceC(cfg)# router-protocol ospf 49
DeviceC(cfg-router)# network-advertise 10.100.1.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.3.0 0.0.0.255 area 0
DeviceC(cfg-router)# exit
```

(2) Enable the IP multicast routing function on all multicast devices, and the PIM-SM function on device interconnection interfaces and interfaces for connecting hosts and multicast sources.

Enable the IP multicast routing function and the PIM-SM function on Device A.

```
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# interface loopback 53
DeviceA(cfg-if-Loopback 53)# ip pim sparse-mode
DeviceA(cfg-if-Loopback 53)# exit
```

Enable the IP multicast routing function and the PIM-SM function on Device B.

```
DeviceB(cfg)# ip mrouting-enable
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/21
DeviceB(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/21)# exit
DeviceB(cfg)# interface VLAN 50
DeviceB(cfg-if-VLAN 50)# ip pim sparse-mode
DeviceB(cfg-if-VLAN 50)# exit
DeviceB(cfg)# interface loopback 50
DeviceB(cfg-if-Loopback 50)# ip pim sparse-mode
DeviceB(cfg-if-Loopback 50)# exit
```

Enable the IP multicast routing function and the PIM-SM function on Device C.

```
DeviceC(cfg)# ip mrouting-enable
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface GigabitEthernet 0/23
DeviceC(cfg-if-GigabitEthernet 0/23)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/23)# exit
```

Configure the IGMP version as IGMPv3 on the interface SVI50 connecting the member host on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface VLAN 50
DeviceB(cfg-if-VLAN 50)# ip igmp specify-version 3
DeviceB(cfg-if-VLAN 50)# exit
```

(3) Configure the SSM group address range.

Configure the SSM multicast group address range on all devices. The following takes the configuration of SSM multicast group address range on Device A as an example. The configuration procedure is the same for Device B and Device C and is omitted.

```
DeviceA(cfg)# ip pim enable-ssm default
```

(4) Enable the SSM-MAP function.

Enable the SSM-MAP function on Device B.

```
DeviceB(cfg)# ip igmp ssm-policy enable
```

(5) Configure the SSM-MAP static linked list.

Configure multicast addresses in 232.0.0.0/24 on Device B to map to Source 1.

```
DeviceB(cfg)# ip access-control-list standard ipv4-map-acl
DeviceB(cfg-stdacl)# permit 232.0.0.0 0.0.0.255
DeviceB(cfg)# ip igmp ssm-policy static ipv4-map-acl 10.100.12.2
```

5. Verification

Verify the SSM-MAP information.

```
DeviceB# display ip igmp-info ssm-policy
SSM Mapping : Enabled
Database      : Static mappings configured
DeviceB#
```

Verify the multicast source information mapped to the multicast group 232.0.0.10.

```
DeviceB# display ip igmp-info ssm-policy 232.0.0.10
Group address: 232.0.0.10
Database      : Static
Source list   : 10.100.12.2
DeviceB#
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
interface GigabitEthernet 0/47
 unset switch-port
 ip address 10.100.12.1 255.255.255.0
 ip pim sparse-mode
```

```
!
interface GigabitEthernet 0/48
 unset switch-port
 ip address 10.100.11.2 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 53
 ip address 53.53.53.53 255.255.255.255
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 10.100.12.0 0.0.0.255 area 0
 network-advertise 53.53.53.53 0.0.0.0 area 0
!
ip pim enable-ssm default
!
end
```

- Device B configuration file

```
hostname DeviceB
!
ip access-control-list standard ipv4-map-acl
 10 permit 232.0.0.0 0.0.0.255
!
ip mrouting-enable
!
vlan range 1,50
!
interface GigabitEthernet 0/18
 unset switch-port
 ip address 10.100.11.1 255.255.255.0
 ip igmp enable-proxy-service
 ip pim sparse-mode
!
interface GigabitEthernet 0/19
 switch-port access vlan 50
!
interface GigabitEthernet 0/20
 switch-port access vlan 50
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.3.1 255.255.255.0
 ip pim sparse-mode
!
```

```
interface GigabitEthernet 0/22
 switch-port access vlan 50
!
interface Loopback 50
 ip address 50.50.50.50 255.255.255.255
 ip pim sparse-mode
!
interface VLAN 50
 ip address 10.100.5.1 255.255.255.0
ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.3.0 0.0.0.255 area 0
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 50.50.50.50 0.0.0.0 area 0
!
ip igmp ssm-policy enable
ip igmp ssm-policy static ipv4-map-acl 10.100.12.2
ip pim enable-ssm default
!
End
```

- Device C configuration file

```
hostname DeviceC
!
ip mrouting-enable
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.1.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/23
 unset switch-port
 ip address 10.100.3.2 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
graceful-restart
 network-advertise 10.100.1.0 0.0.0.255 area 0
 network-advertise 10.100.3.0 0.0.0.255 area 0
!
ip pim enable-ssm default
!
end
```

7. Common Errors

- The basic functions of IGMP are not enabled.
- SSM service is not enabled.
- SSM address range is not configured.
- SSM service address is not configured when using the SSM service.
- The SSM service address range configured on each multicast device is inconsistent.

1.16 Common Misconfigurations

1.16.1 On-demand Clients Failed to Request On-demand Multicast Services

- Phenomenon
  The multicast function is enabled on the network, but the client failed to request one-demand multicast services.
- Troubleshooting
  a Check that the connectivity between the on-demand client and the multicast server is normal.
  Run the **display ip route-config** command to check that there is a unicast route table entry to the multicast source on the device. If not, configure a unicast routing protocol.
  b Check that the configuration of each device is correct.
  Run the **display ip pim-info** { **dense-mode | sparse-mode** } **interface** command to check the PIM information on the interface. If no information is displayed on the interface, it means that PIM-SM function is not enabled. In this case, run the **ip pim sparse-mode** command on the interface to enable the PIM-SM function.
  c Check that the multicast receiver joins the multicast group.
  Run the **display ip igmp-info groups** command to check that the member host has successfully joined the corresponding multicast group on the multicast device on the same network segment as the member host.
  d Check that the multicast table entries of the multicast source and receiver are correct.
  Run the **display ip pim-info** { **dense-mode | sparse-mode** } command to check that the multicast table entries on each device are correct.

Contents

1 Configuring PIM-SM 1
1.1 Introduction 1
1.2 Principles 1
1.2.1 ASM Model 1
1.2.2 SSM Model 6
1.2.3 Protocols and Standards 7
1.3 Restrictions and Guidelines 8
1.4 Configuration Task Summary 8
1.5 Enabling the PIM-SM Function 8
1.5.1 Overview 8
1.5.2 Restrictions and Guidelines 8
1.5.3 Prerequisites 9
1.5.4 Procedure 9
1.6 Configuring PIM-SM in the ASM Model 9
1.6.1 Overview 9
1.6.2 Configuration Tasks 10
1.6.3 Prerequisites 10
1.6.4 Configuring a Static RP 10
1.6.5 Configuring C-BSRs and Related Parameters 11
1.6.6 Configuring C-RPs and Related Parameters 12
1.6.7 Configuring Registration Parameters 15
1.6.8 Configuring the Last Hop Device to Switch over from the RPT to SPT 18

1.7 Configuring PIM-SM in the SSM Model....18
1.7.1 Overview....18
1.7.2 Restrictions and Guidelines....18
1.7.3 Prerequisites....18
1.7.4 Procedure....19
1.8 Configuring PIM Neighbor Parameters....19
1.8.1 Overview....19
1.8.2 Configuration Tasks....19
1.8.3 Prerequisites....19
1.8.4 Configuring the Neighbor Message Time Parameters....19
1.8.5 Configuring Neighbor Tracking....20
1.8.6 Configuring Neighbor Filtering....21
1.9 Configuring DR Priority....21
1.9.1 Overview....21
1.9.2 Prerequisites....22
1.9.3 Procedure....22
1.10 Configuring Parameters of the Join/Prune Packets....22
1.10.1 Overview....22
1.10.2 Restrictions and Guidelines....22
1.10.3 Configuration Tasks....23
1.10.4 Prerequisites....23
1.10.5 Configuring Join/Prune Packet Sending Interval....23
1.10.6 Configuring the Prune Delay....23
1.11 Enabling PIM-SM on the Super VLAN Interface....24

1.12 Enabling PIM BFD Correlation 25

1.13 Monitoring 26

1.14 Configuration Examples 27

1.14.1 Configuring Basic Functions of PIM-SM in the ASM Model 27

1.14.2 Configuring Basic Functions of PIM-SM in the SSM Model 47

1 Configuring PIM-SM

1.1 Introduction

The Protocol Independent Multicast (PIM) is an intra-domain multicast routing protocol that is independent from unicast routing protocols. Any unicast routing protocol, for example, static route, Routing Information Protocol (RIP), and Open Shortest Path First (OSPF), can provide routing information for IP multicast. Multicast routing is independent from unicast routing, and simply uses unicast routing tables to generate multicast routing information. The Protocol Independent Multicast-Sparse Mode (PIM-SM) is used most widely, and applies to networks of any form, especially large networks where group members are scattered in a wide area. Compared with other intra-domain multicast routing protocols, PIM-SM has irreplaceable advantages.

1.2 Principles

PIM-SM considers that not all hosts need to receive multicast data, and sends multicast data to hosts that require the multicast data.

A PIM multicast device provides multicast service in any-source multicast (ASM) and source-specific multicast (SSM) models. For groups in the SSM address range, the SSM model is used. For other groups, the ASM model is used. In the ASM model, a host joins a specified multicast group G without the multicast source S specified. In the SSM model, a host can join a specified multicast group G to receive multicast data from a specified source S.

1.2.1 ASM Model

1. Basic Principles

PIM-SM in the ASM model is implemented as follows:

- A rendezvous point (RP) is used as a root of a rendezvous point tree (RPT) for transition. The designated router (DR) connected to a host sends a join message to the RP of a multicast group, and the multicast devices on the route create (*, G) entries hop by hop. Based on the entries, an RPT is created with the RP as the root.
- To send data to a multicast group, the source DR sends a register message to the RP to create an (S, G) entry on the RP. After the registration succeeds, the multicast source sends multicast data to the RP. After the multicast data arrives at the RP, it is copied and sent to the host via the RPT.

In the ASM model, the key mechanisms of PIM-SM include PIM neighbor creation, DR election, RP, bootstrap router (BSR), multicast source registration, shortest path tree (SPT) switchover, and assertion.

2. PIM Neighbor Creation

Neighbor relationship must be created between PIM multicast devices to form a PIM network, so that PIM control messages and multicast packets can be transmitted on the network. A Hello message is sent from a PIM interface. In the multicast packet with the Hello message encapsulated, the destination address is 224.0.0.13 (indicating all PIM multicast devices in the same network segment), the source address is the IP address of the PIM interface, and the time to live (TTL) value is 1. The Hello message is used to:

- Discover PIM neighbors.

  PIM multicast devices in the same network segment receive the multicast packet with the destination address as 224.0.0.13, obtain neighbor information, and create neighbor relationship. After the neighbor relationship is created, the PIM interfaces can receive and process other PIM packets and create multicast routing entries.

  When a PIM interface is enabled or detects a new neighbor, a Triggered-Hello-Delay message is used to generate a random time. Within the time, the interface sends Hello messages.

- Negotiate protocol parameters. The Hello message contains the following key parameters:
  - **DR Priority**: The priority of a multicast device for DR election. The device with a higher priority is selected.
  - **Holdtime**: The neighbor is considered reachable before timeout.
  - **Propagation Delay**: Duration between the time when a PIM device receives a prune packet and the time when the PIM device sends the prune packet to the uplink devices.
  - **Override-Interval**: The interval to override the prune.

- Maintain neighbor relationship.

  A Hello message is sent periodically between PIM multicast devices. If no new Hello message is received from a PIM neighbor within the hold time, this neighbor is considered unreachable and then removed from the neighbor list. Any change of PIM neighbors will cause change of the multicast topology in the network. If an uplink or a downlink neighbor in a multicast distribution tree (MDT) is unreachable, multicast routes converge again and the MDT is reshaped.

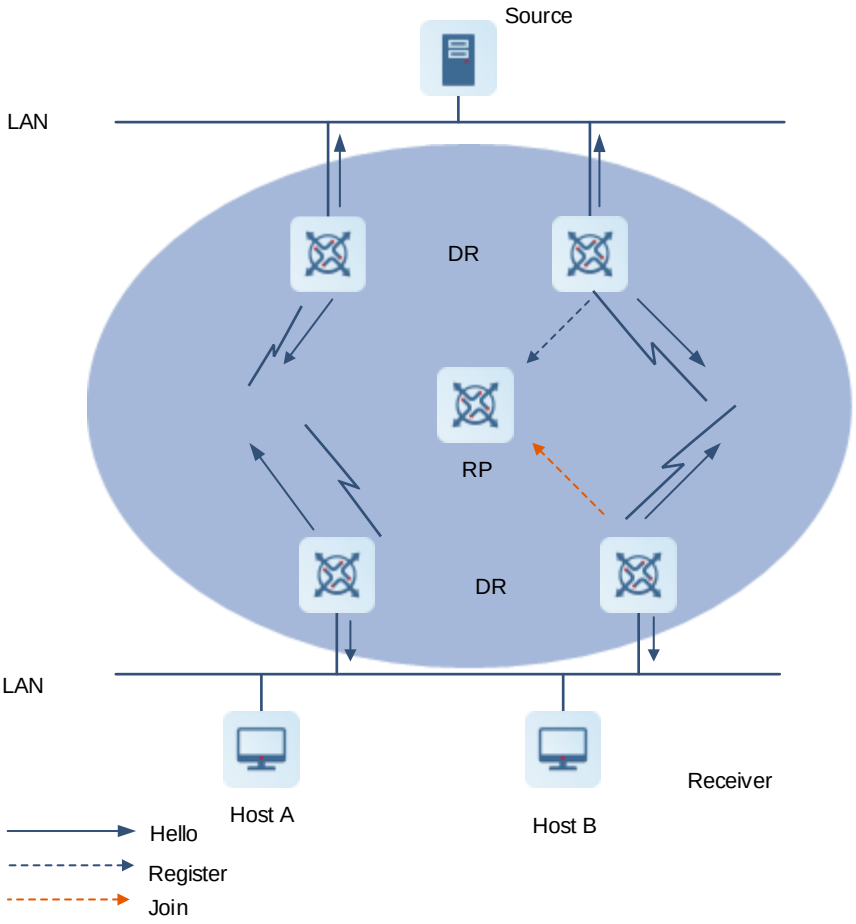
3. DR Election

As shown in Figure 1-1, the source DR sends a register message to the RP. Member DRs send join messages to the RP and forward multicast data to members in a multicast group. In the network segment to which members and multicast source are connected, multiple PIM devices obtain IP addresses and DR priorities of neighbors based on the Hello messages sent between the PIM devices during PIM neighbor creation. The PIM device that is elected becomes a member DR or source DR. The election rules are as follows:

(1) The DR priorities are compared. The device with a higher priority is selected.

(2) If the DR priorities are the same or there is any neighbor in the shared network segment that does not support DR priority, the IP addresses of the devices are compared. The device with a larger IP address is selected.

Figure 1-1 Schematic Diagram of DR Election



4. BSR Election

In the PIM network, the BSR periodically generates bootstrap messages (BSMs) that include a series of candidate-RPs (C-RPs) and corresponding group addresses. The BSMs are sent in the entire domain hop by hop. All PIM multicast devices in the network receive the BSMs and record the C-RPs and corresponding group addresses.

In the PIM-SM domain, one or multiple candidate-BSRs (C-BSRs) are configured and the BSR is elected based on rules.

5. RP Election

In the PIM network, the RP processes register messages of the source DR and join packets of group members. The process from the multicast source to members of the multicast group is divided into two sections: The SPT that uses the source DR as the root and the RP as the leaf, and the RPT that uses the RP as the root and member DRs as leaves. In the PIM-SM network with the ASM model, the RP must act as a transition to create or forward data.

In the PIM network, an RP can serve one or more groups, and a group can be served by only one RP. In the PIM network, the RP can be configured as static or dynamic.

- Static RP

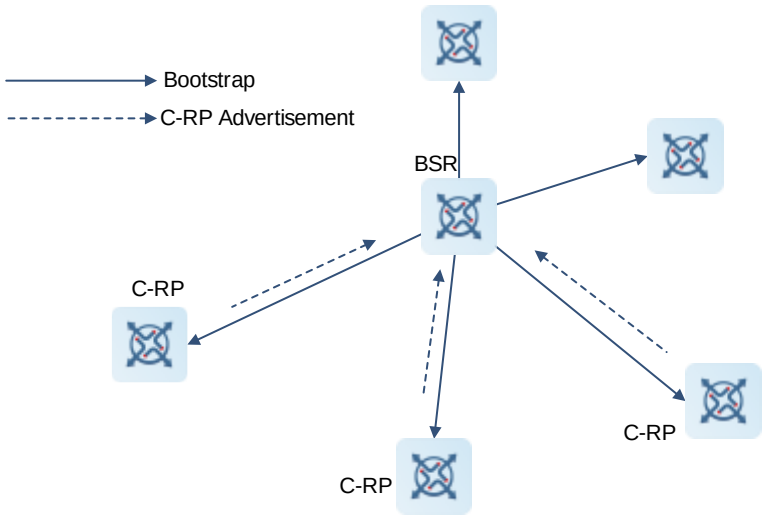
A static RP sets the RP address on each PIM multicast device to enable the entire PIM network to know the RP address.

- Dynamic RP

  As shown in Figure 1-2, multiple C-RPs can be configured in a PIM-SM domain to elect an RP dynamically. These C-RPs add their own addresses and served multicast groups to advertisement packets and sent the packets to the BSR periodically in unicast mode. The BSR is elected from the configured C-BSRs based on rules. The BSR periodically generates BSMs that include a series of C-RPs and corresponding group addresses. The BSMs are sent in the entire domain. Each PIM device in the domain receives and saves these BSMs, and applies the same hash algorithms to compute and elect an RP for each group. The BSR and RP election process is as follows:

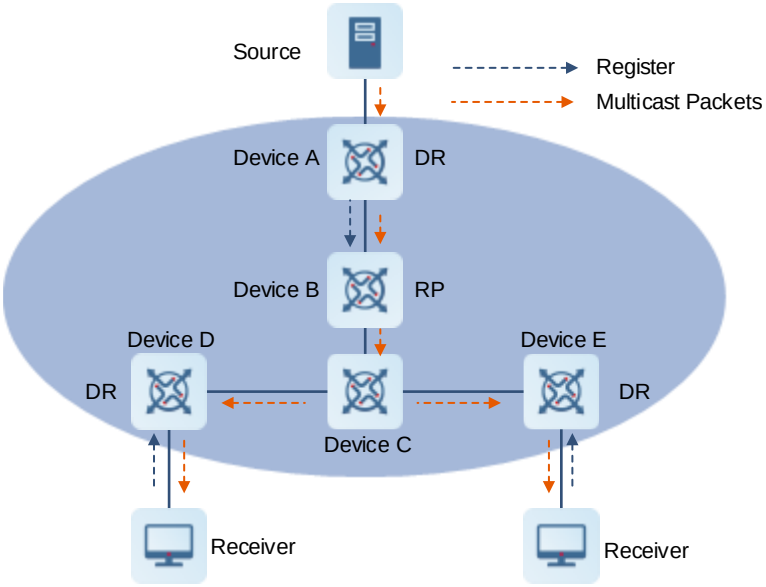
  - BSR election process: Initially, each C-BSR considers itself a BSR and sends a BSM throughout the network. The BSMs carry the C-BSR addresses and priorities. Upon receiving the BSMs, each PIM device compares the C-BSR addresses and priorities to elect the BSR.
  - RP election process: The C-RPs periodically send advertisement packets to the BSR in unicast mode to advertise the C-RP addresses, priorities, and addresses of the served groups. The BSR collects the information as RP-Set, encapsulates the information in a BSM, and sends the messages to each PIM device in the network. Upon receiving the RP-Set, the PIM devices in the network use the same algorithms for computation to obtain the RP that serves each group. Each PIM device obtains the RP position of each group.

Figure 1-2 Schematic Diagram of Dynamic RP Election



6. Multicast Source Registration

Figure 1-3 Schematic Diagram of Multicast Source Registration



As shown in Figure 1-3, when a multicast source sends data to a group, the source DR sends a register message to the RP so that the RP obtains the source information and a multicast packet and forwards the multicast packet to the members in the group. The forwarding process is as follows:

(1) The multicast source sends a multicast packet to the source DR.

(2) The source DR encapsulates the multicast packet in a register message and sends the message to the RP in unicast mode.

(3) Upon receiving the register message, the RP decapsulates the message and forwards the multicast packet to the members in a group via the RPT. Meanwhile, the RP sends an (S, G) join packet to the source. The PIM devices on the route create (S, G) entries, and finally create an SPT from the source to the RP.

(4) The source DR S forwards multicast packets based on this SPT. The multicast data is sent to the RP in the form of multicast packets.

(5) Upon receiving multicast packets from the source, the RP discards the register message and sends a register-stop message to the source.

(6) Upon receiving the register-stop message, the source DR stops encapsulating the register message, forwards multicast packets via the SPT, and periodically sends a register-probe message to the RP.

(7) Upon receiving the register-probe message, the RP sends a register-stop message to the source DR and updates the (S, G) entries.

(8) Upon receiving the register-stop message, the source DR continues forwarding multicast packets via the SPT without encapsulating the register message.

7. RPT Construction

When a group member joins the network, the DR connecting to the group member sends a join packet to the RP. If a multicast source has sent a multicast packet to the RP, the multicast packet will be sent to the group member via the RPT.

The RPT construction procedure is as follows:

- When a member DR receives an IGMP (*, G) INCLUDE Report packet, if the member DR is not an RP of the group G, the member DR sends a (*, G) join packet to the RP and the multicast devices on the route create the (*, G) entries hop by hop to generate an RPT that uses the RP as the root.
- During RPT construction, each PIM device checks the reverse path forwarding (RPF) when sending a join packet to the RP to find the unicast route to the RP. The outbound interface of the unicast route is an uplink interface, and the next hop of the unicast route is the RPF neighbor. The join packet is sent from the member DR to the RP.

8. SPT Switchover

In the PIM-SM network of the ASM model, one or multiple multicast groups correspond to one RP. Before SPT switchover, all packets of multicast groups sent to the RP service must be transitioned by the RP. When the sending rate of multicast packets grows, the RP load is increased. To solve this problem, the PIM-SM switches over the SPT to relieve the RP load. SPT switchover can be triggered by the RP or a member DR.

- SPT switchover triggered by the RP

  After receiving a register message from the source DR, the RP decapsulates the register message, forwards the multicast packets in the message to the group members via the RPT, and sends an (S, G) join packet to the source DR. The PIM multicast devices on the route create (S, G) entries hop by hop to create an SPT from the RP to the source.

- SPT switchover triggered by a member DR

  a When receiving the first multicast packet from the RP, the member DR obtains the source address.

  b The member DR sends an (S, G) join packet to the source. The PIM devices on the route create (S, G) entries and create an SPT from the member DR to the source DR.

  c After the SPT is created, the member DR sends a prune packet to the RP hop by hop and removes the downlink interface from the (*, G) entries. After the pruning, the RP does not forward multicast packets to the member DR via the RPT.

  d The SPT from the member DR to the source DR does not pass the RP. The RP sends a prune packet to the source DR to remove the downlink interface from the (S, G) entries. After the pruning, the source DR does not forward packets to the RP via the original "RP-source DR" SPT.

1.2.2 SSM Model

1. Basic Principles

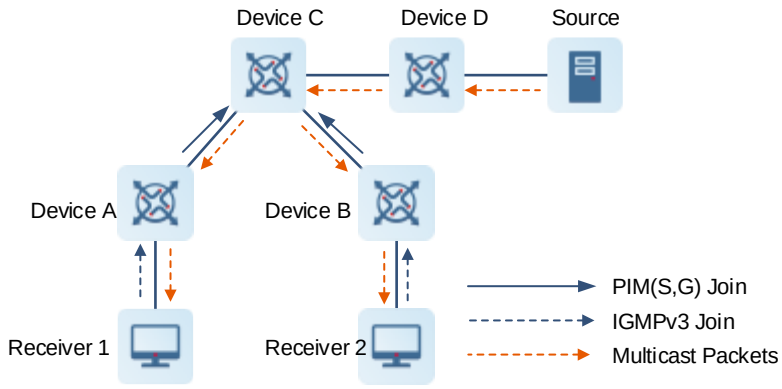
In the SSM model, members in a multicast group can specify to receive or reject data from a multicast source. The group members get to know the (S, G) entries of sources through access to the server or receiving advertisements. After understanding requirement of group members, the member DR sends a join packet to the source DR. This join packet is transmitted to the multicast source hop by hop to create an SPT from the source to the group members.

Note

- To implement the SSM model on IPv4, complete multicast interconnection through the PIM-SM and configure IGMPv3 to manage relationship between the hosts and devices.
- SSM model requirements: The hosts must know the addresses of multicast sources in advance using other network services so that the hosts can select multicast sources.

- On the devices that reside on the route from the hosts to the multicast source, the PIM-SM and SSM functions must be enabled.

Figure 1-4 SSM Model



Note

After the SSM function is enabled, the default group range of the SSM is **232/8**. You can modify the group range of the SSM using commands.

In the SSM model, the key mechanisms of PIM-SM include neighbor creation, DR election, and SPT creation.

2. PIM Neighbor Creation

For more information about the PIM neighbor creation mechanism in the SSM model, see the PIM neighbor creation mechanism in the ASM model in 1.2.1 2. PIM Neighbor Creation.

3. DR Election

For more information about the DR election mechanism in the SSM model, see the DR election mechanism in the ASM model in 1.2.1 3. DR Election.

4. SPT Construction

The SPT construction mechanism is as follows:

(1) The member DR learns through IGMPv3 that a group member wants to receive data of a multicast source and sends an (S, G) join packet to the source hop by hop.

(2) The PIM multicast devices on the route receive the (S, G) join packet and create the (S, G) entries, creating an SPT from the group member to the source.

(3) After the SPT is created, the multicast source distributes the multicast packets to the group member via the SPT.

1.2.3 Protocols and Standards

- RFC 4601: Protocol Independent Multicast-Sparse Mode (PIM-SM)
- RFC 5059: Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)
- RFC 3962: Protocol Independent Multicast-Dense Mode protocol
- RFC 4607: Source-Specific Multicast for IP

1.3 Restrictions and Guidelines

- Before PIM-SM is enabled, enable the multicast routing and forwarding function in global configuration mode. Otherwise, multicast packets cannot be sent even if PIM-SM is enabled.
- In the IPv4 multicast network, the PIM (IPv4) protocol must be run on all L3 devices. Otherwise, a forwarding route cannot be created.

1.4 Configuration Task Summary

The PIM-SM configuration includes the following tasks:

(1) Enabling the PIM-SM Function

(2) Configuring a model for PIM-SM. Select at least one of the configuration tasks to complete.

- Configuring PIM-SM in the ASM Model
- Configuring PIM-SM in the SSM Model

(3) Configuring PIM Neighbor Parameters. All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the Neighbor Message Time Parameters
- Configuring Neighbor Tracking
- Configuring Neighbor Filtering

(4) (Optional) Configuring DR Priority

(5) Configuring Parameters of the Join/Prune Packets. All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring Join/Prune Packet Sending Interval
- Configuring the Prune Delay

(6) (Optional) Enabling PIM-SM on the Super VLAN Interface

(7) (Optional) Enabling PIM BFD Correlation

1.5 Enabling the PIM-SM Function

1.5.1 Overview

Neighbor relationship must be created between PIM multicast devices to form a PIM-SM network, so that PIM control messages and multicast packets can be transmitted on the network. The PIM-SM function must be configured on all interfaces so that PIM neighbor creation, multicast entry creation, and multicast packet forwarding can be successful.

In a PIM network, if you want an interface, for example, a stub network device interface connecting to hosts, to receive multicast packets without involving the PIM network topology construction, you can set the interface to the PIM-SM passive mode. After the PIM-SM passive mode is configured, the interface does not process PIM packets. This improves device performance.

1.5.2 Restrictions and Guidelines

- The IPv4 multicast routing function must be enabled on each multicast device.

- The PIM-SM function should be enabled on the following interfaces: interfaces interconnecting multicast devices, interfaces connecting the static RP, C-RP and C-BSRs, and interfaces connecting multicast sources and hosts.
- PIM interfaces must be L3 interfaces, including: Layer 3 Ethernet interface, Layer 3 link aggregation, SVI, loopback interface, and VXLAN routed interface.

1.5.3 Prerequisites

All PIM interfaces can be reached via IPv4 unicast routes.

1.5.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IPv4 multicast routing.

ip mrouting-enable

The IPv4 multicast routing function is disabled by default.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(5) Enable the PIM-SM function.

ip pim sparse-mode

The PIM-SM function is disabled by default.

(6) (Optional) Enable the PIM-SM passive mode.

ip pim sparse-mode passive

The PIM-SM passive mode is disabled by default.

1.6 Configuring PIM-SM in the ASM Model

1.6.1 Overview

The PIM-SM network supports the multicast service in the ASM and SSM models. Either one model or two models can be configured.

PIM-SM in the ASM model can provide hosts with multicast services of any source. A host joining a multicast group can receive multicast packets destined to this multicast group from any source.

If you require a PIM network to support the multicast service in the ASM model, an RP must be configured. There are three methods to configure an RP:

- Configure a static RP.
- Configure a dynamic RP.
- Configure static and dynamic RPs. If static and dynamic RPs are available at the same time, dynamic RPs are preferred.

1.6.2 Configuration Tasks

The PIM-SM ASM model configuration includes the following tasks. Select the tasks to configure a static RP or a dynamic RP, or configure both.

- Configuring a Static RP
- Configuring a Dynamic RP
  - Configuring C-BSRs and Related Parameters
  - Configuring C-RPs and Related Parameters
  - (Optional) Configuring Registration Parameters
- (Optional) Configuring the Last Hop Device to Switch over from the RPT to SPT

1.6.3 Prerequisites

In the PIM-SM multicast network, all nodes are reachable via IPv4 unicast routes.

1.6.4 Configuring a Static RP

1. Overview

You can manually configure a static RP to prevent frequent packet interaction between the BSR and C-RPs or between the BSR and PIM multicast devices and avoid bandwidth occupation. If the multicast network is simple or small, you can configure a static RP.

2. Restrictions and Guidelines

- Static RPs must be configured on every multicast device.
- The multicast device selected as a static RP must boast high performance.
- You are advised to use the loopback interface as a static RP address.
- Static RPs configured on multicast devices should be consistent, including RP address and address range of the served groups.
- If multiple static RPs are configured to correspond to different addresses of groups, the RPs can implement load balancing.
- If multiple static RPs serve the same group, the static RP with a larger IP address is selected first.
- You can use the control list to configure an RP to serve multiple (referencing an ACL) or all (not referencing an ACL) multicast groups. However, an RP can adopt only one configuration. The last configuration overrides the preceding configurations.

3. Prerequisites

Before an address range of a group is configured for an RP, ensure that the corresponding ACL is created and specify a multicast group range for the static RP to serve. For more information about ACL configuration, see *Configuring ACL*.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static RP.

ip pim static-rp-addr *ipv4-rp-address* [{ *acl-name* | *acl-number* }]

No static RP is configured by default.

1.6.5 Configuring C-BSRs and Related Parameters

1. Overview

In the PIM-SM domain of the ASM model, a unique BSR must be available. The BSR collects and releases RP information.

- The BSR is elected from C-BSRs based on rules. The election rules are as follows:
  - a The priorities of the C-BSRs are compared. The device with a higher priority is selected.
  - b If the priorities are the same, the IP addresses are compared. The device with a larger IP address is selected.
  - c The winner is elected as the BSR in the domain and periodically sends BSMs in the domain. A BSM contains the BSR IP address, RP-Set and hash mask information. The hash mask is used for hash calculation during C-RP election. In the hold time of a BSM, if other C-BSRs do not receive the BSM from the BSR, the BSR is considered unreachable and a new round of BSR election is initiated.
- To avoid fraud by an illegitimate BSR, a BSR range can be specified on all PIM devices in the domain so that the PIM devices ignore a BSM that is sent from an illegitimate BSR.
- To limit BSM flooding, you can configure a BSR border on an interface. Upon receiving BSMs, this interface discards the BSMs without forwarding them. This limits the BSM flooding.

2. Restrictions and Guidelines

- C-BSRs must be configured on one or multiple multicast devices.
- The multicast devices selected as C-BSRs must boast high performance. The C-RPs and C-BSRs can reside on the same or different multicast devices.
- You are advised to use the loopback interface as a C-BSR address.
- If multiple C-BSRs are configured, they can implement redundancy backup.
- You are advised to configure a valid BSR address range on all PIM devices in the domain.
- A BSR border must be configured on the interface between domains.

3. Prerequisites

- Enabling the PIM-SM Function.
- Before a valid BSR address range is configured on PIM devices, ensure that the corresponding ACL is created and specify a valid BSR range. For more information about ACL configuration, see *Configuring ACL*.

4. Procedure

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Configure C-BSRs and related parameters.

ip pim c-bsr *interface-type interface-number* [*hash-mask-length* [*priority-value*]]

No C-BSR is configured by default.

(2) (Optional) Configure a valid BSR address range on all PIM devices.

ip pim accept-bsr { **list** *acl-name* | **list** *acl-number* }

All BSMs are received by default.

(3) (Optional) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) (Optional) Configure a BSR border on the interface between PIM-SM domains.

ip pim bsr-boundary

No BSR border is configured by default.

If multiple domains exist, the BSR border must be configured on the interface between two domains.

1.6.6 Configuring C-RPs and Related Parameters

1. Overview

- After a multicast device is configured as a C-RP, the C-RP periodically sends an advertisement packet to the BSR in unicast mode. The packet includes the RP priority, advertisement packet hold time (greater than the interval of sending the advertisement packet), RP address, address of the served group, and prefix of the multicast address. Upon receiving the advertisement packet, the BSR collects the information in the packet as RP-Set in the advertisement packet hold time, encapsulates the information in a BSM, and sends the message to all PIM devices. Upon receiving the BSM from the BSR, the PIM devices in the domain apply the same algorithms for hash calculation to calculate the C-RP serving each multicast group. After the advertisement packet hold timer times out, if the BSR does not receive a next advertisement packet from the

C-RP, this C-RP is considered unreachable.

- To prevent fraud by an illegitimate C-RP, a C-RP range can be specified on the BSR. If an advertisement packet comes from an illegitimate C-RP, the packet is ignored.
- The BSR can determine whether to receive an advertisement packet with the prefix of the multicast address being 0 from a C-RP. When the prefix of the multicast address is 0, this C-RP serves all multicast groups.
- When receiving the BSM from the BSR, each PIM multicast device in the domain applies rules to select the RP. The RP election rules are as follows:
  a The address of the multicast group that a host joins is compared with the mask length of the group address range that the C-RP serves. The device with the longest mask is selected.
  b If the addresses are consistent with the mask length, the priorities of the C-RPs are compared. The device with the smallest priority value is selected.
  c If the priorities are consistent, hash values are calculated and compared. The device with the largest hash value is selected.
  d If the hash values are the same, the IP addresses of the C-RPs are compared. The device with the largest IP address is selected.

2. Restrictions and Guidelines

- C-RPs must be configured on one or multiple multicast devices.
- The multicast devices selected as C-RPs must boast high performance. The C-RPs and C-BSRs can reside on the same or different multicast devices.
- You are advised to use the loopback interface as a C-RP address.
- If multiple C-RPs are configured to serve the same group, they can implement redundancy backup.
- If multiple C-RPs are configured to serve different groups, they can implement load balancing.
- A valid C-RP range can be configured on the BSR.
- To specify a group range for C-RPs to serve, you can add the ACL option to the **ip pim c-rp** command. The group range is calculated based on the permit access control entry (ACE), rather than the deny ACE.

3. Prerequisites

- Enabling the PIM-SM Function.
- If the **group-list** parameter is used to limit the range of groups served by the RP, ensure that the corresponding ACL is created and specify a multicast group range for the RP to serve. For more information about ACL configuration, see *Configuring ACL*.

4. Procedure

- Configure C-RPs and related parameters.
  a Enter the privileged EXEC mode.
  privilege
  b Enter the global configuration mode.
  config-system terminal
  c Configure C-RPs.

ip pim c-rp *interface-type interface-number* [**interval** *interval* | **priority** *priority-value* | { **group-acl** *acl-name* | **group-acl** *acl-number* }] *

No C-RP is configured by default.

- (Optional) Configure the BSR to receive C-RP advertisement packets with the prefix of the multicast address being 0. This C-RP serves all the multicast groups.

  a Enter the privileged EXEC mode.

  privilege

  b Enter the global configuration mode.

  config-system terminal

  c Configure the BSR to receive advertisement packets from the C-RP that serves all multicast groups.

  ip pim accept-crp-with-null-group

  The function for the BSR to receive advertisement packets from the C-RP that serves all multicast groups is disabled by default.

  You are advised to enable this function on all C-BSRs.

- (Optional) Configure a valid C-RP range on the BSR.

  a Enter the privileged EXEC mode.

  privilege

  b Enter the global configuration mode.

  config-system terminal

  c Configure a valid C-RP range.

  ip pim accept-crp { **list** *acl-name* | **list** *acl-number* }

  The C-BSRs receive all external C-RP advertisement packets by default.

  You are advised to enable this function on all C-BSRs.

- (Optional) Configure to ignore RP priority on all PIM multicast devices and reselect an RP.

  a Enter the privileged EXEC mode.

  privilege

  b Enter the global configuration mode.

  config-system terminal

  c Configure to ignore the RP priority for RP election.

  ip pim ignore-rp-set-priority

  If the address of the multicast group that a host joins is consistent with the mask length of the group address range that the C-RPs serve, the RP with the highest priority is selected by default.

  You can configure to ignore C-RP priority on each multicast device.

1.6.7 Configuring Registration Parameters

1. Overview

When a multicast source sends multicast data, the data arrives at the source DR first. Upon receiving the multicast data, the source DR encapsulates the multicast data as a register message and sends the message to the RP. Related parameters of the register message are configured on the source DR and RP.

- You can configure the following parameters on the source DR:
  - RP reachability checking: Before sending a register message, the source DR determines whether the RP is reachable. If the RP is reachable, the message is sent. Otherwise, the message is not sent.
  - Limit on the rate of sending the register message: The rate of sending each register message in (S, G) status is limited. If multiple register messages in (S, G) status are sent from the same source, the rate of the register messages is limited to the product of *N* × Limit value.
  - Source address in the register message: The source address in the register message is the address of the DR interface connected to the multicast source by default. If a security policy is configured on the RP to filter out this address, the RP does not receive register messages from this source DR. In this case, you can specify a reachable source address as the source address of the register message.
  - Register suppression time: When the RP receives multicast packets sent by the multicast source via the SPT, the RP sends a register-stop message to the source DR. Within the register- suppression time, the source DR stops encapsulating the register message and sends multicast packets to the RP via the SPT. After the register suppression timer times out, if the source DR does not receive another register-stop message from the RP, the source DR continues to encapsulate the register message and send the message to the RP.
  - Null message detection time: Before the register suppression timer times out, the RP receives a null message from the source DR, refreshes the local (S, G) entries, and sends a register-stop message to the source DR so that the source DR refreshes the register suppression timer. Otherwise, the source DR continues to encapsulate the register message and send multicast packets to the RP in the form of the register message.
- You can configure the following parameters on the RP:
  - (S, G) address range of source: Provide multicast services for hosts in the permitted multicast group in the ASM model. After this parameter is configured, the RP replies a register-stop message upon receiving data from an unauthorized source.
  - Length of register message checksum: Adjust the calculation method of the register message checksum and calculate the checksum of the entire PIM protocol packets (including encapsulated multicast packets) or the checksum of PIM headers of separate register messages.
  - Register message decapsulation and delivery to downlink interface: Configure this parameter on the RP to decapsulate the received register message and forward the multicast packets in the register message.
  - Register message (S, G) entry timeout period: Configure this parameter to configure the timeout period of (S, G) entries. The (S, G) entries age if no packet is received within the register message timeout period.

2. Restrictions and Guidelines

- The source address in the specified register message must be reachable in unicast mode. You are advised to use the loopback interface of the source DR as the source address of the register message.
- If the **ip pim rp-register-kat** command is not configured, modifying the register suppression time on the RP changes the (S, G) entry timeout period.
- Configuring a limit to the register message rate reduces the load of the source DR and RP. Only register messages under the limit are sent.
- The register-probe time cannot be greater than a half of the register suppression time. Otherwise, the configuration failed and an alarm is generated.
- The sum of the three times of register suppression time and the register-probe time does not exceed 65,535. Otherwise, the configuration failed and an alarm is generated.

3. Prerequisites

- Enabling the PIM-SM Function.
- Before an (S, G) address range is configured for the source, if the **list** parameter is specified, ensure that the corresponding ACL is created and specify an (S, G) address range for the source. For more information about ACL configuration, see *Configuring ACL*.
- Before an (S, G) address range is configured for the source, if the **route-map** parameter is specified, ensure that the corresponding route map is created and specify an (S, G) address range for the source. For more information about route map configuration, see *Configuring Route Policies*.
- Before the register message checksum calculation is applied to entire packets, if the **group-list** parameter is specified, ensure that the corresponding ACL is created and specify the register message checksum calculation to apply to the entire packets in a multicast group address range. For more information about ACL configuration, see *Configuring ACL*.

4. Procedure

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configure registration parameters on the source DR.
  - a Enter the privileged EXEC mode.

    privilege
  - b Enter the global configuration mode.

    config-system terminal
  - c Enable RP reachability checking on the source DR.

    ip pim register-rp-reachability

    The RP reachability checking function is disabled by default.
  - d Limit the rate of sending register messages on the source DR.

    ip pim limit-register-rate *rate*

    The sending rate of register messages is not limited by default.
  - e Configure a source address in the register messages on the source DR.

    ip pim register-source-addr { *ipv4-local-address* | *interface-type interface-number* }

The source IP address in the register messages is the address of the DR interface connected to the source by default.

f Configure the register suppression time on the source DR.

ip pim register-suppression *suppression-time*

The default register suppression time is **60** seconds.

g Configure the null message detection time on the source DR.

ip pim probe-interval *probe-interval*

The default register-probe time is **5** seconds.

- Configure registration parameters on the RP.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Configure the (S, G) address range for the source on the RP.

ip pim restricted-register { { **list** *acl-name* | **list** *acl-number* } | **route-map** *route-map-name* } *

The (S, G) address range of register messages is not limited by default. The RP receives register messages of all (S, G) addresses.

d Enable register message checksum calculation to be applied to the entire packets on the RP.

ip pim register-checksum-wholepkt [{ **group-acl** *acl-name* | **group-acl** *acl-number* }]

By default, the register message checksum of headers of only PIM packets and register messages, rather than the entire packets, is calculated.

If the **group-acl** parameter is not specified, this configuration applies to all group addresses.

e Enable the RP to decapsulate the received register messages and forward the multicast packets in the received register messages on the RP.

ip pim register-decapsulate-forward

The function for the RP to decapsulate the register messages and forward the multicast packets is disabled by default.

The register message decapsulation and multicast packet forwarding function is implemented by the corresponding software. If there are too many register messages to be decapsulated and forwarded, the CPU is overloaded. Therefore, you are not advised to enable this function.

f Configure the timeout period for the (S, G) entries created upon receiving the register message on the RP.

ip pim rp-register-kat *interval*

By default, the timeout period for the (S, G) entries is the sum of three times of the register suppression time and the register-probe time.

1.6.8 Configuring the Last Hop Device to Switch over from the RPT to SPT

1. Overview

If the multicast traffic in a network is heavy, the RP easily gets overloaded and then faulty. To relieve the RP load, you can initiate SPT switchover through the member DR. You can specify multicast groups to perform SPT switchover.

2. Prerequisites

- Enabling the PIM-SM Function.
- Before an address range of a group is configured for SPT switchover, ensure that the corresponding ACL is created and specify a multicast group range for SPT switchover. For more information about ACL configuration, see *Configuring ACL*.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the SPT switchover function.

ip pim spt-switching [{ **group-acl** *acl-name* | **group-acl** *acl-number* }]

The SPT switchover function is disabled by default.

If the **group-acl** parameter is not specified, all groups support SPT switchover.

1.7 Configuring PIM-SM in the SSM Model

1.7.1 Overview

The PIM-SM function in the SSM model can provide specified source services for hosts so that the hosts can determine to receive or reject multicast data of specified sources based on needs.

1.7.2 Restrictions and Guidelines

- If you require a PIM network to support the multicast service in the SSM model, the SSM function must be enabled.
- The SSM function must be enabled on each multicast device.
- The address range of SSM groups configured on all multicast devices must be consistent.
- IGMPv3 or SSM mapping must be enabled.

1.7.3 Prerequisites

- In the multicast network, all nodes are reachable via IPv4 unicast routes.
- Enabling the PIM-SM Function.
- Before an SSM group range is configured using an ACL, ensure that the corresponding ACL is created and specify an SSM multicast group range. For more information about ACL configuration, see *Configuring ACL*.

1.7.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the SSM function.

ip pim enable-ssm { **default** | **range** { *acl-name* | *acl-number* } }

The SSM function is disabled by default.

1.8 Configuring PIM Neighbor Parameters

1.8.1 Overview

PIM multicast devices transmit hello messages to discover neighbors, coordinate protocol parameters and maintain neighbor relationships.

1.8.2 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the Neighbor Message Time Parameters
- Configuring Neighbor Tracking
- Configuring Neighbor Filtering

1.8.3 Prerequisites

- In the multicast network, all nodes are reachable via IPv4 unicast routes.
- Enabling the PIM-SM Function.

1.8.4 Configuring the Neighbor Message Time Parameters

1. Overview

The neighbor message time parameters include hello message sending interval and hello message sending delay.

- The hello message sending interval can be configured, and the hello message hold time is 3.5 times the hello message sending interval. If no new hello message is received within the hello message hold time, the neighbor is considered unreachable.
- When the PIM interface is enabled or detects a new neighbor, random time is generated and it is referred to as hello message sending delay. Within the delay time, the interface sends hello messages out.

2. Restrictions and Guidelines

If the hello message hold time is greater than 65,535 seconds, the hello message sending interval is automatically set to 18,725 seconds.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the neighbor message time parameters. Select at least one of them to configure.

- Configure the hello message sending interval.

  ip pim query-period *hello-interval*

  The hello message is sent at an interval of **30** seconds by default.
- Configure the hello message sending delay.

  ip pim triggered-hello-delay *delay*

  The hello message sending delay is **5** seconds by default.

1.8.5 Configuring Neighbor Tracking

1. Overview

After the suppression capability of an interface is enabled, when a multicast device plans to send a join packet to an uplink multicast device but it receives a join packet sent from the neighbor to the uplink multicast device, this local multicast device suppresses its own join packet. If the suppression capability of the interface is disabled, the join packet can be sent. When the suppression capability of downlink hosts is disabled, an uplink device can determine the number of the downlink hosts based on the quantity of received join packets. This is neighbor tracking.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.

interface vlan *interface-number*

- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Enable the neighbor tracking function.

ip pim neighbor-tracking

The neighbor tracking function is disabled by default.

1.8.6 Configuring Neighbor Filtering

1. Overview

The neighbor filtering function can protect neighborhood. Controlling neighbors can strengthen PIM network security and limit the valid address range of neighbors. If a neighbor is filtered out based on an access filtering list, PIM-SM does not create peer relationship with the neighbor or stops the peer relationship with this neighbor.

2. Prerequisites

Before an access filtering list is configured, ensure that the corresponding ACL is created and specify a valid neighbor address range. For more information about ACL configuration, see *Configuring ACL*.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Enable the neighbor filtering function.

ip pim peering-filter { *acl-name* | *acl-number* }

The neighbor filtering function is disabled by default.

1.9 Configuring DR Priority

1.9.1 Overview

In a network segment shared by multicast sources and group members, multiple PIM multicast devices exist. The PIM multicast devices elect a DR based on neighbor messages. A device with the highest priority is

selected. If the priorities are the same or some PIM devices do not support the DR priority function, the IP addresses are compared and the PIM multicast device with the largest IP address is selected. The selected PIM device functions as a DR, transmits protocol packets between the sources and members, and forwards multicast packets.

1.9.2 Prerequisites

- In the multicast network, all nodes are reachable via IPv4 unicast routes.
- Enabling the PIM-SM Function.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure a DR priority in hello messages.

ip pim dr-elected-priority *priority-value*

The default DR priority in hello messages is **1**.

1.10 Configuring Parameters of the Join/Prune Packets

1.10.1 Overview

A device sends a join packet to request an uplink device to forward multicast packets and sends a prune packet to request the uplink device to stop forwarding multicast packets. The request route is referred to as an RPT or SPT.

1.10.2 Restrictions and Guidelines

- The network administrator must ensure that the prune override interval is smaller than the join-prune packet hold time. Otherwise, a short interrupt may occur.
- When the PIM multicast device receives a prune message from an uplink interface, there are other downlink devices existent in the shared network segment. If this PIM multicast device still needs to receive multicast data, the PIM multicast device must send a prune override message to the uplink device within the override interval.

1.10.3 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring Join/Prune Packet Sending Interval
- Configuring the Prune Delay

1.10.4 Prerequisites

Enabling the PIM-SM Function.

1.10.5 Configuring Join/Prune Packet Sending Interval

1. Overview

A PIM device periodically sends the join/prune packet to an uplink device to update the forwarding status. The join packet is used to request data forwarding and the prune packet is used to stop data forwarding. Upon receiving the join/prune packet, the uplink device starts the join/prune packet hold time timer.

- If a join packet hold time timer times out, the PIM device is suppressed to forward multicast packets of the corresponding multicast group through the downlink interface.
- If a prune packet hold time timer times, the PIM device is restored to forward the multicast packets of the corresponding multicast group through the downlink interface.

2. Restrictions and Guidelines

The join/prune packet sending interval must be smaller than the join/prune packet hold time. Generally, the join/prune packet hold time is 3.5 times the join/prune packet sending interval.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the join/prune packet sending interval.

ip pim jp-timer *jp-timer-interval*

The join/prune packet is sent at an interval of **60** seconds by default.

1.10.6 Configuring the Prune Delay

1. Overview

Upon receiving a prune packet from a downlink device, the multicast device does not perform prune immediately and it waits for a period of time. This period of time is divided into two parts: The part from the time when the multicast device receives a prune message to the time when the multicast device sends the prune message to the uplink device is referred to as the prune propagation delay; the part from the time when the PIM device receives a prune packet to the time when the device continues to forward the multicast packets to a downlink device is referred to as the override interval. The period from the time when the multicast device receives a prune message to the time when the prune action is performed is the sum of the propagation delay and the override interval.

2. Restrictions and Guidelines

In the period from the time when the multicast device receives a prune message to the time when the prune action is performed, if the PIM device receives a prune override message from the downlink interface, the PIM device cancels pruning.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the prune message propagation delay.

ip pim propagation-delay *propagation-delay*

The default propagation delay of an interface is **500** ms.

(5) Configure the prune override interval.

ip pim override-interval *override-interval*

The default override interval is **2500** ms.

1.11 Enabling PIM-SM on the Super VLAN Interface

1. Overview

In most scenarios of a PIM network, the PIM-SM protocol does not need to be enabled on a super VLAN interface. A super VLAN interface is connected to many sub VLANs. If the PIM-SM protocol is enabled on the super VLAN interface, the device replicates and sends a multicast protocol packet to all the sub VLANs, which easily surpasses the processing capability of the device and causes packet loss even protocol flapping. In some scenarios, the PIM-SM protocol must be enabled on the super VLAN interface. You can configure the command to enable the protocol packets to be sent to the specified sub VLANs or all sub VLANs. If the packets are sent to all VLANs, protocol flapping may occur.

2. Restrictions and Guidelines

- This function must be enabled on the super VLAN interface.
- If you enable protocol packets to be sent to all sub VLANs, the packets may surpass the processing capability of the device, causing protocol flapping.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI configuration mode.

interface vlan *interface-number*

(4) Enable PIM-SM protocol packets to be sent to specified sub VLANs.

ip pim sparse-mode subvlan [*subvlan-id* | **all**]

The PIM-SM function is disabled on a super VLAN interface by default.

1.12 Enabling PIM BFD Correlation

1. Overview

In some shared multicast networks, PIM neighbor status change triggers re-initiation of some mechanisms, for example, DR election mechanism. Generally, a PIM neighbor triggers a new round of election after the neighborhood times out, causing multicast data interruption. PIM bidirectional forwarding detection (BFD) correlation can quickly detect link status and neighbor faults in a shared network segment and trigger a new round of DR election. This reduces multicast data transmission interrupt time and ensures reliability of the multicast network.

2. Restrictions and Guidelines

- If DR election or assertion election is required in a multicast network, you are advised to enable PIM-SM BFD correlation.
- The PIM BFD function must be enabled on the interconnection interface of the multicast devices.
- PIM interfaces must be L3 interfaces, including: Layer 3 Ethernet interface, Layer 3 link aggregation, SVI, loopback interface, VXLAN routed interface.

3. Prerequisites

- In the network, all nodes are reachable via IPv4 unicast routes.
- Enabling the PIM-SM Function.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Enable PIM BFD correlation.

ip pim bfd

PIM BFD correlation is disabled by default.

1.13 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup ip mroute { *ipv4-group-address* [*ipv4-source-address*] | ***** }	Clears multicast routing entries.
cleanup ip mroute statistics { *ipv4-group-address* [*ipv4-source-address*] | ***** }	Clears statistical information about multicast routing.
cleanup ip pim-info sparse-mode bsr rp-set *	Clears dynamic RP information.
cleanup ip pim-info sparse-mode track	Clears statistical information about PIM-SM protocol packets.
display ip pim-info sparse-mode bsr-router	Displays details about the BSR.
display ip pim-info sparse-mode interface [*interface-type interface-number*] [**detail**]	Displays PIM-SM information of an interface.
display ip pim-info sparse-mode local-members [*interface-type interface-number*]	Displays local IGMP information about a PIM-SM interface.
display ip pim-info sparse-mode mroute [*group-or-source-address* [*group-or-source-address*]]	Displays information about the PIM-SM multicast routing table.
display ip pim-info sparse-mode mroute [*ipv4-group-or-source-address* [*ipv4-group-or-source-address*]]	Displays PIM-SM neighbor information.

Command	Purpose
display ip pim-info sparse-mode nexthop	Displays the PIM-SM next hop information in an NSM.
display ip pim-info sparse-mode rp-group-hash *ipv4-group-address*	Displays RP information corresponding to a group address.
display ip pim-info sparse-mode rp mapping	Displays all RPs and the groups served by the RPs.
display ip pim-info sparse-mode track	Displays the number of PIM packets sent and received since the statistic start time.

1.14 Configuration Examples

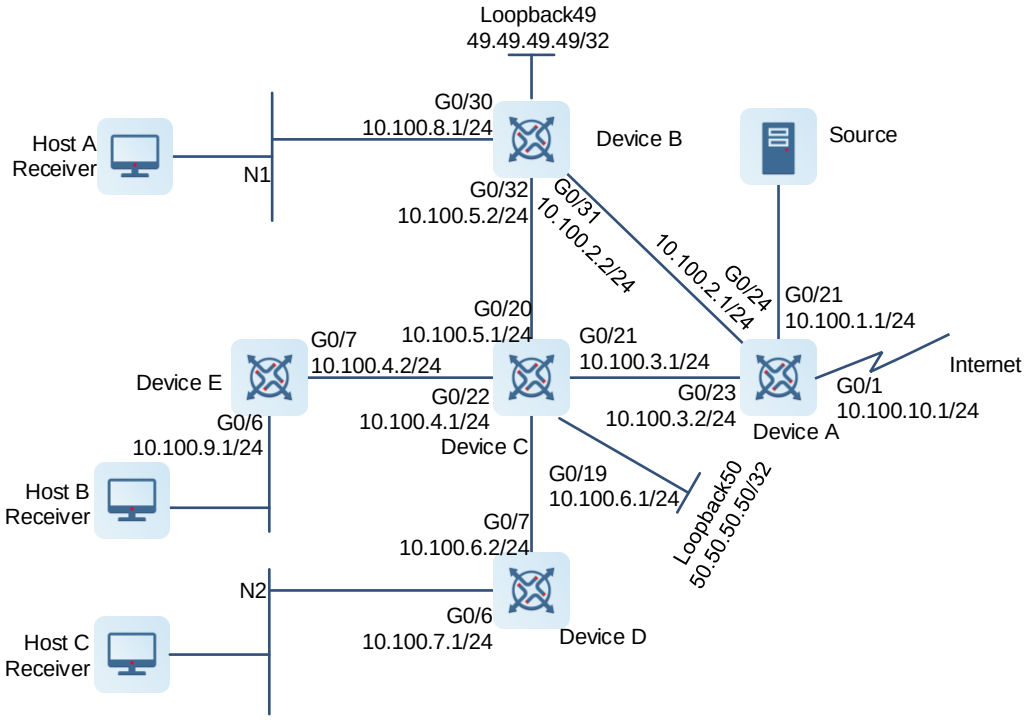
1.14.1 Configuring Basic Functions of PIM-SM in the ASM Model

1. Requirements

As shown in Figure 1-5, the internal network is connected to the Internet. The PIM-SM protocol is configured on each multicast device to provide ASM multicast services for hosts in networks N1 and N2 so that hosts joining a multicast group can receive multicast data destined to this multicast group from any multicast source, including video-on-demand information.

2. Topology

Figure 1-5 PIM-SM Multicast Network Topology in the ASM Model



3. Notes

- Configure IP addresses and unicast routing protocols for multicast devices. The PIM routing protocol in a multicast domain depends on unicast routing protocols. Normal communication of unicast routing lays a foundation of multicast routing. Ensure that each PIM node is reachable via unicast routes.
- Enable the multicast function on all multicast devices. Enabling the multicast function is a prerequisite of PIM-SM configuration.
- Enable the basic functions of PIM-SM on all interfaces of the multicast devices. The other functions of PIM-SM can be enabled only after the basic functions of PIM-SM are enabled.
- Enable the IGMP function on the interface connected to hosts so that the hosts can join or leave a multicast group by sending IGMP messages. The multicast devices acting as leaf nodes maintain group membership based on the IGMP protocol.
- Configure RPs. In the PIM-SM network in the ASM model, the RP is a transitional station for forwarding multicast data. It is recommended that the multicast device with many traffic flows be configured as an RP, for example, device C in Figure 1-5.
- Configure a BSR border on the interface connected to the Internet. BSMs must be transmitted in the PIM-SM domain within the border to improve the multicast controllability.

4. Procedure

(1) Configure IP addresses and enable unicast routing protocols on interfaces of different devices.

Configure IPv4 addresses on interfaces of device A, and configure the OSPF protocol to implement internetworking.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 10.100.10.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/21)# ip address 10.100.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)#interface gigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/23)# ip address 10.100.3.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# interface gigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/24)# ip address 10.100.2.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.1.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.10.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.2.0 0.0.0.255 area 0
```

```
DeviceA(cfg-router)# network-advertise 10.100.3.0 0.0.0.255 area 0
DeviceA(cfg-router)# exit
```

Configure IPv4 addresses on interfaces of device B, and configure the OSPF protocol to implement internetworking.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/30)# ip address 10.100.8.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/31)# ip address 10.100.2.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/32)# ip address 10.100.5.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface loopback 49
DeviceB(cfg-if-Loopback 49)# ip address 49.49.49.49 255.255.255.255
DeviceB(cfg-if-Loopback 49)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.8.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.2.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 49.49.49.49 0.0.0.0 area 0
DeviceB(cfg-router)# exit
```

Configure IPv4 addresses on interfaces of device C, and configure the OSPF protocol to implement internetworking.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/19
DeviceC(cfg-if-GigabitEthernet 0/19)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/19)# ip address 10.100.6.1 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/19)# exit
DeviceC(cfg)#interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/20)# ip address 10.100.5.1 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ip address 10.100.3.1 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface GigabitEthernet 0/22
DeviceC(cfg-if-GigabitEthernet 0/22)# unset switch-port
```

```
DeviceC(cfg-if-GigabitEthernet 0/22)# ip address 10.100.4.1 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/22)# exit
DeviceC(cfg)# interface loopback 50
DeviceC(cfg-if-Loopback 50)# ip address 50.50.50.50 255.255.255.255
DeviceC(cfg-if-Loopback 50)# exit
DeviceC(cfg)# router-protocol ospf 49
DeviceC(cfg-router)# network-advertise 10.100.6.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.3.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.4.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 50.50.50.50 0.0.0.0 area 0
DeviceC(cfg-router)# exit
```

Configure IPv4 addresses on interfaces of device D, and configure the OSPF protocol to implement internetworking.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# interface GigabitEthernet 0/6
DeviceD(cfg-if-GigabitEthernet 0/6)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/6)# ip address 10.100.7.1 255.255.255.0
DeviceD(cfg-if-GigabitEthernet 0/6)# exit
DeviceD(cfg)# interface GigabitEthernet 0/7
DeviceD(cfg-if-GigabitEthernet 0/7)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/7)# ip address 10.100.6.2 255.255.255.0
DeviceD(cfg-if-GigabitEthernet 0/7)# exit
DeviceD(cfg)# router-protocol ospf 49
DeviceD(cfg-router)# network-advertise 10.100.6.0 0.0.0.255 area 0
DeviceD(cfg-router)# network-advertise 10.100.7.0 0.0.0.255 area 0
DeviceD(cfg-router)# exit
```

Configure IPv4 addresses on interfaces of device E, and configure the OSPF protocol to implement internetworking.

```
DeviceE> privilege
DeviceE# config-system terminal
DeviceE(cfg)# interface GigabitEthernet 0/6
DeviceE(cfg-if-GigabitEthernet 0/6)# unset switch-port
DeviceE(cfg-if-GigabitEthernet 0/6)# ip address 10.100.9.1 255.255.255.0
DeviceE(cfg-if-GigabitEthernet 0/6)# interface GigabitEthernet 0/7
DeviceE(cfg-if-GigabitEthernet 0/7)# unset switch-port
DeviceE(cfg-if-GigabitEthernet 0/7)# ip address 10.100.4.2 255.255.255.0
DeviceE(cfg-if-GigabitEthernet 0/7)# exit
DeviceE(cfg)# router-protocol ospf 49
DeviceE(cfg-router)# network-advertise 10.100.9.0 0.0.0.255 area 0
DeviceE(cfg-router)# network-advertise 10.100.4.0 0.0.0.255 area 0
DeviceE(cfg-router)# exit
```

(2) Enable the multicast and PIM-SM functions on all multicast devices, and enable the IGMP function on all multicast devices connected to hosts.

Enable the multicast and PIM-SM functions on device A.

```
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface gigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface gigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# interface gigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
```

Enable the multicast, PIM-SM, and IGMP functions on device B.

```
DeviceB(cfg)# ip mrouting-enable
DeviceB(cfg)# interface gigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface gigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface gigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface loopback 49
DeviceB(cfg-if-Loopback 49)# ip pim sparse-mode
DeviceB(cfg-if-Loopback 49)# exit
```

Enable the multicast and PIM-SM functions on device C.

```
DeviceC(cfg)# ip mrouting-enable
DeviceC(cfg)# interface gigabitEthernet 0/19
DeviceC(cfg-if-GigabitEthernet 0/19)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/19)# exit
DeviceC(cfg)# interface gigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface gigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface gigabitEthernet 0/22
DeviceC(cfg-if-GigabitEthernet 0/22)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/22)# exit
DeviceC(cfg)# interface loopback 50
DeviceC(cfg-if-Loopback 50)# ip pim sparse-mode
DeviceC(cfg-if-Loopback 50)# exit
```

Enable the multicast, PIM-SM, and IGMP functions on device D.

```
DeviceD(cfg)# ip mrouting-enable
```

```
DeviceD(cfg)# interface gigabitEthernet 0/6
DeviceD(cfg-if-GigabitEthernet 0/6)# ip pim sparse-mode
DeviceD(cfg-if-GigabitEthernet 0/6)# exit
DeviceD(cfg)# interface gigabitEthernet 0/7
DeviceD(cfg-if-GigabitEthernet 0/7)# ip pim sparse-mode
DeviceD(cfg-if-GigabitEthernet 0/7)# exit
```

Enable the multicast, PIM-SM, and IGMP functions on device E.

```
DeviceE(cfg)# ip mrouting-enable
DeviceE(cfg)# interface gigabitEthernet 0/6
DeviceE(cfg-if-GigabitEthernet 0/6)# ip pim sparse-mode
DeviceE(cfg-if-GigabitEthernet 0/6)# exit
DeviceE(cfg)# interface gigabitEthernet 0/7
DeviceE(cfg-if-GigabitEthernet 0/7)# ip pim sparse-mode
DeviceE(cfg-if-GigabitEthernet 0/7)# exit
```

(3) Configure RPs.

Note

There are two methods to configure RPs: dynamic RPs and static RPs. You can configure either or both types of RPs. If static and dynamic RPs are available at the same time, the dynamic RPs are preferred by default.

Configure dynamic RPs. One or multiple multicast devices in the PIM-SM domain must be configured as C-RPs and C-BSRs. In this example, device B and device C are configured as C-RP and C-BSR. The address range of the served groups of the RP, as well as the interfaces of the C-RP and C-BSR are configured on device B and device C. The loopback interface is configured as the interface of the C-RP and C-BSR.

Configure C-RP and C-BSR and address range of the served groups of the RP on device B.

```
DeviceB(cfg)# ip pim c-bsr loopback 49
DeviceB(cfg)# ip access-control-list standard 49
DeviceB(cfg-stdacl)# permit 224.0.0.0 0.255.255.255
DeviceB(cfg)# ip pim c-rp loopback 49 group-list 49
```

Configure C-RP and C-BSR and address range of the served groups of the RP on device C.

```
DeviceC(cfg)# ip pim c-bsr loopback 50
DeviceC(cfg)# ip access-control-list standard 50
DeviceC(cfg-stdacl)# permit 225.0.0.0 0.255.255.255
DeviceC(cfg-stdacl)# exit
DeviceC(cfg)# ip pim c-rp loopback 50 group-acl-list 50
```

Configure static RPs. The addresses of static RPs must be specified on all multicast devices.

Configure a static RP on device A.

```
DeviceA(cfg)# ip pim static-rp-addr 50.50.50.50
```

Configure a static RP on device B.

```
DeviceB(cfg)# ip pim static-rp-addr 50.50.50.50
```

Configure a static RP on device C.

```
DeviceC(cfg)# ip pim static-rp-addr 50.50.50.50
```

Configure a static RP on device D.

```
DeviceD(cfg)# ip pim static-rp-addr 50.50.50.50
```

Configure a static RP on device E.

```
DeviceE(cfg)# ip pim static-rp-addr 50.50.50.50
```

(4) Configure a BSR border on the device A interface connected to the Internet.

```
DeviceA(cfg-if-GigabitEthernet 0/1)# ip pim bsr-boundary
```

5. Verification

(1) Run the **display ip pim-info sparse-mode interface** command to display the PIM-SM running information on the interfaces.

The PIM-SM running information on the interface of device A is as follows:

```
DeviceA# display ip pim-info sparse-mode interface
Address          Interface                    VIFindex Ver/Mode Nbr-Count DR-Prior
DR
10.100.1.1       GigabitEthernet 0/21         1        v2/S     0         1
10.100.1.1
10.100.3.2       GigabitEthernet 0/23         2        v2/S     1         1
10.100.3.2
10.100.2.1       GigabitEthernet 0/24         3        v2/S     0         1
10.100.2.1
DeviceA#
```

The PIM-SM running information on the interface of device B is as follows:

```
DeviceB# display ip pim-info sparse-mode interface
Address          Interface                    VIFindex Ver/Mode Nbr-Count DR-Prior
DR
10.100.8.1       GigabitEthernet 0/30         1        v2/S     0         1
10.100.8.1
10.100.2.2       GigabitEthernet 0/31         4        v2/S     1         1
10.100.2.2
10.100.5.2       GigabitEthernet 0/32         2        v2/S     1         1
10.100.5.2
49.49.49.49      Loopback 49                  3        v2/S     0         1
49.49.49.49
DeviceB#
```

The PIM-SM running information on the interface of device C is as follows:

```
DeviceC# display ip pim-info sparse-mode interface
Address          Interface                    VIFindex Ver/Mode Nbr-Count DR-Prior
DR
10.100.6.1       GigabitEthernet 0/19         1        v2/S     1         1
10.100.6.2
10.100.5.1       GigabitEthernet 0/20         2        v2/S     1         1
10.100.5.2
```

```
10.100.3.1      GigabitEthernet 0/21          3          v2/S    1          1
10.100.3.2
10.100.4.1      GigabitEthernet 0/22          4          v2/S    1          1
10.100.4.2
50.50.50.50     Loopback 50                   5          v2/S    0          1
50.50.50.50
DeviceC#
```

The PIM-SM running information on the interface of device D is as follows:

```
DeviceD# display ip pim-info sparse-mode interface
Address          Interface                    VIFindex Ver/Mode Nbr-Count DR-Prior
DR
10.100.7.1       GigabitEthernet 0/6          1        v2/S     0         1
10.100.7.1
10.100.6.2       GigabitEthernet 0/7          2        v2/S     1         1
10.100.6.2
DeviceD#
```

The PIM-SM running information on the interface of device E is as follows:

```
DeviceE# display ip pim-info sparse-mode interface
Address          Interface                    VIFindex Ver/Mode Nbr-Count DR-Prior
DR
10.100.9.1       GigabitEthernet 0/6          1        v2/S     0         1
10.100.9.1
10.100.4.2       GigabitEthernet 0/7          2        v2/S     1         1
10.100.4.2
DeviceE#
```

(2) Run the **display ip pim-info sparse-mode bsr-router** command to display the election information of the BSRs.

The election information on device A is as follows:

```
DeviceA# display ip pim-info sparse-mode bsr-router
PIMv2 Bootstrap information
  BSR address: 50.50.50.50
  Uptime:       00:48:33, BSR Priority: 64, Hash mask length: 10
  Expires:      00:01:32
  Role: Non-candidate BSR     Priority: 0, Hash mask length: 10
  State: Accept Preferred
DeviceA#
```

The election information on device B is as follows:

```
DeviceB# display ip pim-info sparse-mode bsr-router
PIMv2 Bootstrap information
  BSR address: 50.50.50.50
  Uptime:       00:49:26, BSR Priority: 64, Hash mask length: 10
  Expires:      00:01:39
  Role: Candidate BSR     Priority: 64, Hash mask length: 10
  State: Candidate BSR
```

```
  Candidate RP: 49.49.49.49(Loopback 49)
    Advertisement interval 60 seconds
    Next Cand_RP_advertisement in 00:00:19
DeviceB#
```

The election information on device C is as follows:

```
DeviceC# display ip pim-info sparse-mode bsr-router
PIMv2 Bootstrap information
This system is the Bootstrap Router (BSR)
  BSR address: 50.50.50.50
  Uptime:        00:51:04, BSR Priority: 64, Hash mask length: 10
  Next bootstrap message in 00:00:56
  Role: Candidate BSR     Priority: 64, Hash mask length: 10
  State: Elected BSR
DeviceC#
```

The election information on device D is as follows:

```
DeviceD# display ip pim-info sparse-mode bsr-router
PIMv2 Bootstrap information
  BSR address: 50.50.50.50
  Uptime:        00:51:38, BSR Priority: 64, Hash mask length: 10
  Expires:       00:01:28
  Role: Non-candidate BSR     Priority: 0, Hash mask length: 10
  State: Accept Preferred
DeviceD#
```

The election information on device E is as follows:

```
DeviceE# display ip pim-info sparse-mode bsr-router
PIMv2 Bootstrap information
  BSR address: 50.50.50.50
  Uptime:        00:52:24, BSR Priority: 64, Hash mask length: 10
  Expires:       00:01:42
  Role: Non-candidate BSR     Priority: 0, Hash mask length: 10
  State: Accept Preferred
DeviceE#
```

(3) Run the **display ip pim-info sparse-mode rp mapping** command to display the RP information on the multicast devices.

The RP information on device A is as follows:

```
DeviceA# display ip pim-info sparse-mode rp mapping
PIM Group-to-RP Mappings
Group(s): 224.0.0.0/8
  RP: 49.49.49.49(Not self)
    Info source: 50.50.50.50, via bootstrap, priority 192
          Uptime: 00:55:25, expires: 00:01:52
Group(s): 225.0.0.0/8
  RP: 50.50.50.50(Not self)
```

```
    Info source: 50.50.50.50, via bootstrap, priority 192
        Uptime: 00:00:48, expires: 00:01:52
Group(s): 224.0.0.0/4, Static
    RP: 50.50.50.50(Not self) , Static
        Uptime: 00:50:13
DeviceA#
```

The RP information on device B is as follows:

```
DeviceB# display ip pim-info sparse-mode rp mapping
PIM Group-to-RP Mappings
Group(s): 224.0.0.0/8
  RP: 49.49.49.49(Self)
    Info source: 50.50.50.50, via bootstrap, priority 192
        Uptime: 00:56:12, expires: 00:02:05
Group(s): 225.0.0.0/8
  RP: 50.50.50.50(Not self)
    Info source: 50.50.50.50, via bootstrap, priority 192
        Uptime: 00:01:35, expires: 00:02:05
Group(s): 224.0.0.0/4, Static
    RP: 50.50.50.50(Not self) , Static
        Uptime: 00:49:36
DeviceB#
```

The RP information on device C is as follows:

```
DeviceC# display ip pim-info sparse-mode rp mapping
PIM Group-to-RP Mappings
This system is the Bootstrap Router (v2)
Group(s): 224.0.0.0/8
  RP: 49.49.49.49(Not self)
    Info source: 49.49.49.49, via bootstrap, priority 192
        Uptime: 00:56:17, expires: 00:02:13
Group(s): 225.0.0.0/8
  RP: 50.50.50.50(Self)
    Info source: 50.50.50.50, via bootstrap, priority 192
        Uptime: 00:01:40, expires: 00:01:50
Group(s): 224.0.0.0/4, Static
    RP: 50.50.50.50(Self) , Static
        Uptime: 00:49:39
DeviceC#
```

The RP information on device D is as follows:

```
DeviceD# display ip pim-info sparse-mode rp mapping
PIM Group-to-RP Mappings
Group(s): 224.0.0.0/8
  RP: 49.49.49.49(Not self)
    Info source: 50.50.50.50, via bootstrap, priority 192
        Uptime: 00:56:22, expires: 00:01:55
Group(s): 225.0.0.0/8
```

```
  RP: 50.50.50.50(Not self)
    Info source: 50.50.50.50, via bootstrap, priority 192
          Uptime: 00:01:45, expires: 00:01:55
Group(s): 224.0.0.0/4, Static
    RP: 50.50.50.50(Not self) , Static
          Uptime: 00:49:40
DeviceD#
```

The RP information on device E is as follows:

```
DeviceE# display ip pim-info sparse-mode rp mapping
PIM Group-to-RP Mappings
Group(s): 224.0.0.0/8
  RP: 49.49.49.49(Not self)
    Info source: 50.50.50.50, via bootstrap, priority 192
          Uptime: 00:56:27, expires: 00:01:50
Group(s): 225.0.0.0/8
  RP: 50.50.50.50(Not self)
    Info source: 50.50.50.50, via bootstrap, priority 192
          Uptime: 00:01:50, expires: 00:01:50
Group(s): 224.0.0.0/4, Static
    RP: 50.50.50.50(Not self) , Static
          Uptime: 00:49:43
DeviceE#
```

(4) Run the **display ip pim-info sparse-mode mroute** command to display the multicast routing tables of PIM-SM. The multicast source (10.100.1.2/24) sends traffic to the multicast groups (225.0.0.1/24) and (224.0.1.1/24). Host A joins the multicast group (225.0.0.1), and host B joins the multicast group (224.0.1.1).

The routing table information on device A is as follows:

```
DeviceA# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 2
(S,G,rpt) Entries: 2
FCR Entries: 0
REG Entries: 2

(10.100.1.2, 224.0.1.1)
RPF nbr: 0.0.0.0
RPF idx: None
SPT bit: 1
Upstream State: JOINED
kat expires in 53 seconds
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . j . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(10.100.1.2, 224.0.1.1, rpt)
RP: 50.50.50.50
RPF nbr: 10.100.3.1
RPF idx: GigabitEthernet 0/23
Upstream State: RPT NOT JOINED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
```

```
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Pruned
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(10.100.1.2, 225.0.0.1)
RPF nbr: 0.0.0.0
RPF idx: None
SPT bit: 1
Upstream State: JOINED
kat expires in 100 seconds
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . j . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
```

```
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(10.100.1.2, 225.0.0.1, rpt)
RP: 50.50.50.50
RPF nbr: 10.100.3.1
RPF idx: GigabitEthernet 0/23
Upstream State: RPT NOT JOINED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Pruned
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
```

```
0  . . . . . .
1  . . . . . .
2  . . . . . .
3  . . . . . .
4  . . . . . .
5  . . . . . .
6  . . . . . .
7  . . . . . .

DeviceA#
```

The routing table information on device B is as follows (only key information is displayed):

```
DeviceB> privilege
DeviceB# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 1
(S,G) Entries: 0
(S,G,rpt) Entries: 0
FCR Entries: 1
REG Entries: 0

(*, 225.0.0.1)
RP: 50.50.50.50
RPF nbr: 10.100.5.1
RPF idx: GigabitEthernet 0/32
Upstream State: JOINED
jt_timer expires in 52 seconds
FCR:
Source: 10.100.1.2
KAT timer running, 166 seconds remaining
  Packet count 1

DeviceB#
```

The routing table information on device C is as follows (only key information is displayed):

```
DeviceC> privilege
DeviceC# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 2
(S,G) Entries: 2
(S,G,rpt) Entries: 2
FCR Entries: 0
REG Entries: 0
```

```
(*, 224.0.1.1)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: JOINED
FCR:

(10.100.1.2, 224.0.1.1)
RPF nbr: 10.100.3.2
RPF idx: GigabitEthernet 0/21
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 15 seconds
kat expires in 203 seconds

(10.100.1.2, 224.0.1.1, rpt)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: PRUNED

(*, 225.0.0.1)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: JOINED
FCR:

(10.100.1.2, 225.0.0.1)
RPF nbr: 10.100.3.2
RPF idx: GigabitEthernet 0/21
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 23 seconds
kat expires in 176 seconds

(10.100.1.2, 225.0.0.1, rpt)
RP: 50.50.50.50
RPF nbr: 0.0.0.0
RPF idx: None
Upstream State: PRUNED

DeviceC#
```

The routing table information on device E is as follows (only key information is displayed):

```
DeviceE> privilege
```

```
DeviceE# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 1
(S,G) Entries: 0
(S,G,rpt) Entries: 0
FCR Entries: 1
REG Entries: 0

(*, 224.0.1.1)
RP: 50.50.50.50
RPF nbr: 10.100.4.1
RPF idx: GigabitEthernet 0/7
Upstream State: JOINED
jt_timer expires in 34 seconds
FCR:
Source: 10.100.1.2
KAT timer running, 156 seconds remaining
  Packet count 1

DeviceE#
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
  unset switch-port
  ip address 10.100.1.1 255.255.255.0
  ip pim sparse-mode
  ip pim bsr-border
!
interface GigabitEthernet 0/21
  unset switch-port
  ip address 10.100.1.1 255.255.255.0
  ip pim sparse-mode
!
interface GigabitEthernet 0/23
  unset switch-port
  ip address 10.100.3.2 255.255.255.0
  ip pim sparse-mode
!
interface GigabitEthernet 0/24
```

```
 unset switch-port
 ip address 10.100.2.1 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.1.0 0.0.0.255 area 0
 network-advertise 10.100.2.0 0.0.0.255 area 0
 network-advertise 10.100.3.0 0.0.0.255 area 0
 network-advertise 10.100.10.0 0.0.0.255 area 0
!
ip pim static-rp-addr 50.50.50.50
!
```

- Device B configuration file

```
hostname DeviceB
!
ip access-control-list standard 49
 20 permit 224.0.0.0 0.255.255.255
!
ip mrouting-enable
!
interface GigabitEthernet 0/30
 unset switch-port
 ip address 10.100.8.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/31
 unset switch-port
 ip address 10.100.2.2 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/32
 unset switch-port
 ip address 10.100.5.2 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 49
 ip address 49.49.49.49 255.255.255.255
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.2.0 0.0.0.255 area 0
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 10.100.8.0 0.0.0.255 area 0
 network-advertise 49.49.49.49 0.0.0.0 area 0
```

```
!
ip pim static-rp-addr 50.50.50.50
ip pim c-bsr Loopback 49
ip pim c-rp Loopback 49 group-acl 49
!
```

- Device C configuration file

```
hostname DeviceC
!
ip access-control-list standard 50
 10 permit 225.0.0.0 0.255.255.255
!
ip mrouting-enable
!
interface GigabitEthernet 0/19
 unset switch-port
 ip address 10.100.6.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/20
 unset switch-port
 ip address 10.100.5.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.3.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/22
 unset switch-port
 ip address 10.100.4.1 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 50
 ip address 50.50.50.50 255.255.255.255
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.3.0 0.0.0.255 area 0
 network-advertise 10.100.4.0 0.0.0.255 area 0
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 10.100.6.0 0.0.0.255 area 0
 network-advertise 50.50.50.50 0.0.0.0 area 0
!
ip pim static-rp-addr 50.50.50.50
```

```
ip pim c-bsr Loopback 50
ip pim c-rp Loopback 50 group-list 50
!
```

- Device D configuration file

```
hostname DeviceD
!
ip mrouting-enable
!
interface GigabitEthernet 0/6
 unset switch-port
 ip address 10.100.7.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/7
 unset switch-port
 ip address 10.100.6.2 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.6.0 0.0.0.255 area 0
 network-advertise 10.100.7.0 0.0.0.255 area 0
!
ip pim static-rp-addr 50.50.50.50
!
```

- Device E configuration file

```
hostname DeviceE
!
ip mrouting-enable
!
interface GigabitEthernet 0/6
 unset switch-port
 ip address 10.100.9.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/7
 unset switch-port
 ip address 10.100.4.2 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.4.0 0.0.0.255 area 0
 network-advertise 10.100.9.0 0.0.0.255 area 0
!
```

```
ip pim static-rp-addr 50.50.50.50
!
```

7. Common Errors

- The multicast routing function is not enabled on multicast devices.
- The multicast function is not enabled on any interface.
- No BSR border is configured between PIM-SM domains.
- No RP is configured.
- No static RP is configured on a multicast device, or the configured static RP on this multicast device is not consistent with the configuration on other multicast devices.
- C-RPs are configured, but no C-BSR is configured.
- Static RP, C-RP or C-BSR is unreachable via unicast routes.

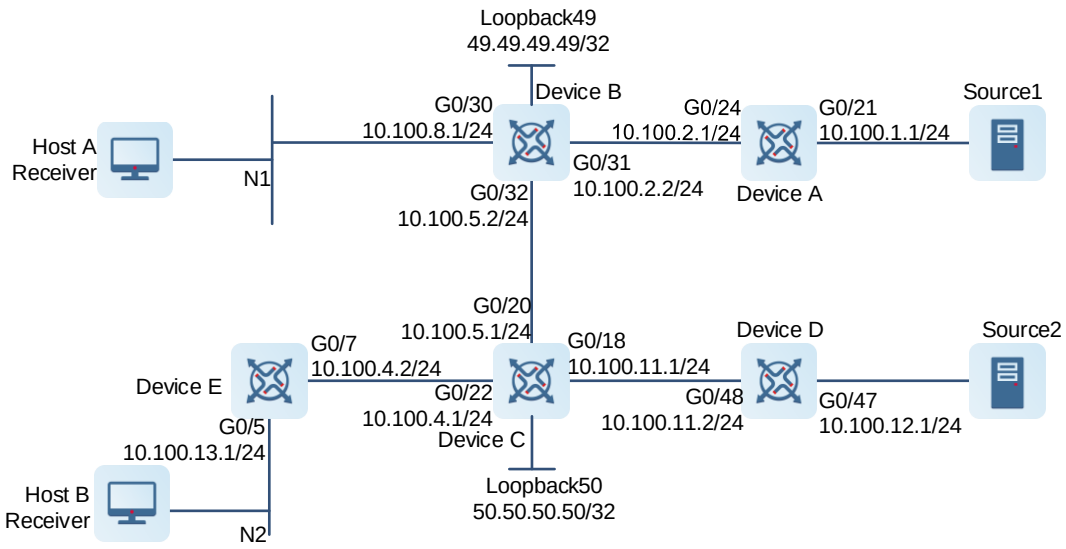
1.14.2 Configuring Basic Functions of PIM-SM in the SSM Model

1. Requirements

As shown in Figure 1-6, the PIM-SM protocol is enabled on multicast devices to provide SSM services for hosts in N1 and N2 so that the hosts can receive multicast data from specified multicast sources to the multicast group that the hosts join, for example, video-on-demand.

2. Topology

Figure 1-6 PIM-SM Multicast Network Topology in the SSM Model



3. Notes

- Configure IP addresses and unicast routing protocols for multicast devices. The PIM routing protocol in a multicast domain depends on unicast routing protocols. Normal communication of unicast routing lays a foundation of multicast routing. Ensure that each PIM node is reachable via unicast routes.
- Enable the multicast function on all multicast devices. Enabling the multicast function is a prerequisite of

PIM-SM configuration.

- Enable the basic functions of PIM-SM on all interfaces of the multicast devices. The other functions of PIM-SM can be enabled only after the basic functions of PIM-SM are enabled.
- Enable the IGMP function on the interface connected to hosts and set the IGMP version to v3 so that the hosts can join or leave a multicast group by sending IGMP messages based on the specified source. The multicast devices acting as leaf nodes maintain group membership based on the IGMP protocol.
- Configure the same address range of SSM on all multicast devices. The SSM services provided by the multicast devices in the PIM-SM domain are dedicated to specified groups, which realizes multicast controllability.

4. Procedure

(1) Configure IP addresses and enable unicast routing protocols on interfaces of different devices.

Configure IP addresses and enable unicast routing protocols on interfaces of device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/21)# ip address 10.100.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/24)# ip address 10.100.2.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.1.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.2.0 0.0.0.255 area 0
DeviceA(cfg-router)# exit
```

Configure IP addresses and enable unicast routing protocols on interfaces of device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/30)# ip address 10.100.8.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/31)# ip address 10.100.2.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)#interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/32)# ip address 10.100.5.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ip address 49.49.49.49 255.255.255.255
```

```
DeviceB(cfg-if-Loopback 49)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.2.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.8.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 49.49.49.49 0.0.0.0 area 0
DeviceB(cfg-router)# exit
```

Configure IP addresses and enable unicast routing protocols on interfaces of device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/18
DeviceC(cfg-if-GigabitEthernet 0/18)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/18)# ip address 10.100.11.1 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/18)# exit
DeviceC(cfg)#interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/20)# ip address 10.100.5.1 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/22
DeviceC(cfg-if-GigabitEthernet 0/22)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/22)# ip address 10.100.4.1 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/22)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ip address 50.50.50.50 255.255.255.255
DeviceC(cfg-if-Loopback 50)# exit
DeviceC(cfg)# router-protocol ospf 49
DeviceC(cfg-router)# network-advertise 10.100.4.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.5.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 50.50.50.50 0.0.0.0 area 0
DeviceC(cfg-router)# exit
```

Configure IP addresses and enable unicast routing protocols on interfaces of device D.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# interface GigabitEthernet 0/47
DeviceD(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/47)# ip address 10.100.12.1 255.255.255.0
DeviceD(cfg-if-GigabitEthernet 0/47)# exit
DeviceD(cfg)# interface GigabitEthernet 0/48
DeviceD(cfg-if-GigabitEthernet 0/48)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/48)# ip address 10.100.11.2 255.255.255.0
DeviceD(cfg-if-GigabitEthernet 0/48)# exit
DeviceD(cfg)# router-protocol ospf 49
DeviceD(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceD(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
```

```
DeviceD(cfg-router)# exit
DeviceD(cfg)# router-protocol ospf 49
DeviceD(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceD(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
DeviceD(cfg-router)# exit
```

Configure IP addresses and enable unicast routing protocols on interfaces of device E.

```
DeviceE> privilege
DeviceE# config-system terminal
DeviceE(cfg)# interface GigabitEthernet 0/5
DeviceE(cfg-if-GigabitEthernet 0/5)# unset switch-port
DeviceE(cfg-if-GigabitEthernet 0/5)# ip address 10.100.13.1 255.255.255.0
DeviceE(cfg-if-GigabitEthernet 0/5)# exit
DeviceE(cfg)# interface GigabitEthernet 0/7
DeviceE(cfg-if-GigabitEthernet 0/7)# unset switch-port
DeviceE(cfg-if-GigabitEthernet 0/7)# ip address 10.100.4.2 255.255.255.0
DeviceE(cfg-if-GigabitEthernet 0/7)# exit
DeviceE(cfg)# router-protocol ospf 49
DeviceE(cfg-router)# network-advertise 10.100.4.0 0.0.0.255 area 0
DeviceE(cfg-router)# network-advertise 10.100.13.0 0.0.0.255 area 0
DeviceE(cfg-router)# exit
```

(2) Enable the multicast and PIM-SM functions on all multicast devices, and enable the IGMP function on all multicast devices connected to hosts.

Enable the multicast and PIM-SM functions on device A.

```
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
```

Enable the multicast, PIM-SM, and IGMP functions on device B, and set the IGMP version to v3.

```
DeviceB(cfg)# ip mrouting-enable
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/30)# ip igmp specify-version 3
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# ip pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ip pim sparse-mode
```

```
DeviceB(cfg-if-Loopback 49)# exit
```

Enable the multicast and PIM-SM functions on device C.

```
DeviceC(cfg)# ip mrouting-enable
DeviceC(cfg)# interface GigabitEthernet 0/18
DeviceC(cfg-if-GigabitEthernet 0/18)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/18)# exit
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/22
DeviceC(cfg-if-GigabitEthernet 0/22)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/22)# exit
DeviceC(cfg)#interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ip pim sparse-mode
DeviceC(cfg-if-Loopback 50)# exit
```

Enable the multicast and PIM-SM functions on device D.

```
DeviceD(cfg)# ip mrouting-enable
DeviceD(cfg)# interface GigabitEthernet 0/47
DeviceD(cfg-if-GigabitEthernet 0/47)# ip pim sparse-mode
DeviceD(cfg-if-GigabitEthernet 0/47)# exit
DeviceD(cfg)# interface GigabitEthernet 0/48
DeviceD(cfg-if-GigabitEthernet 0/48)# ip pim sparse-mode
DeviceD(cfg-if-GigabitEthernet 0/48)# exit
```

Enable the multicast, PIM-SM, and IGMP functions on device E, and set the IGMP version to v3.

```
DeviceE(cfg)# ip mrouting-enable
DeviceE(cfg)# interface GigabitEthernet 0/5
DeviceE(cfg-if-GigabitEthernet 0/5)# ip pim sparse-mode
DeviceE(cfg-if-GigabitEthernet 0/5)# ip igmp specify-version 3
DeviceE(cfg-if-GigabitEthernet 0/5)# exit
DeviceE(cfg)# interface GigabitEthernet 0/7
DeviceE(cfg-if-GigabitEthernet 0/7)# ip pim sparse-mode
DeviceE(cfg-if-GigabitEthernet 0/7)# exit
```

(3) Enable the SSM function and configure an address range of multicast groups on all multicast devices.

Set the address range of an SSM multicast group to 232.1.1.0/24 on device A.

```
DeviceA(cfg)# ip access-control-list standard 59
DeviceA(cfg-stdacl)#permit 232.1.1.0 0.0.0.255
DeviceA(cfg-stdacl)#exit
DeviceA(cfg)# ip pim enable-ssm range 59
```

Set the address range of an SSM multicast group to 232.1.1.0/24 on device B.

```
DeviceB(cfg)# ip access-control-list standard 59
DeviceB(cfg-stdacl)#permit 232.1.1.0 0.0.0.255
DeviceB(cfg-stdacl)#exit
DeviceB(cfg)# ip pim enable-ssm range 59
```

Set the address range of an SSM multicast group to 232.1.1.0/24 on device C.

```
DeviceC(cfg)# ip access-control-list standard 59
DeviceC(cfg-stdacl)#permit 232.1.1.0 0.0.0.255
DeviceC(cfg-stdacl)#exit
DeviceC(cfg)# ip pim enable-ssm range 59
```

Set the address range of an SSM multicast group to 232.1.1.0/24 on device D.

```
DeviceD(cfg)# ip access-control-list standard 59
DeviceD(cfg-stdacl)#permit 232.1.1.0 0.0.0.255
DeviceD(cfg-stdacl)#exit
DeviceD(cfg)# ip pim enable-ssm range 59
```

Set the address range of an SSM multicast group to 232.1.1.0/24 on device E.

```
DeviceE(cfg)# ip access-control-list standard 59
DeviceE(cfg-stdacl)#permit 232.1.1.0 0.0.0.255
DeviceE(cfg-stdacl)#exit
DeviceE(cfg)# ip pim enable-ssm range 59
```

5. Verification

(1) Run the **display ip pim-info sparse-mode interface** command to display PIM-SM running information on the interfaces. The PIM-SM running information on the interfaces of device A is as follows. The PIM-SM running information on the interfaces of devices B, C, D, and E is the same.

```
DeviceA# display ip pim-info sparse-mode interface
Address          Interface                      VIFindex Ver/Mode Nbr-Count DR-Prior
DR
10.100.1.1       GigabitEthernet 0/21           3        v2/S     0         1
10.100.1.1
10.100.2.1       GigabitEthernet 0/24           1        v2/S     0         1
10.100.2.1
DeviceA#
```

(2) Run the **display ip pim-info sparse-mode mroute** command to display the multicast routing tables of PIM-SM. Host A needs to receive packets sent from source 1 (10.100.1.2/24) and source 2 (10.100.12.2/24) to the multicast group (232.1.1.10/24). Host B needs to receive packets sent from source 1 (10.100.1.2/24) to the multicast group (232.1.1.10/24). The information output is as follows:

The routing table information on device A is as follows:

```
DeviceA# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 1
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(10.100.1.2, 232.1.1.10)
```

```
RPF nbr: 0.0.0.0
RPF idx: None
SPT bit: 1
Upstream State: JOINED
kat expires in 133 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  j  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  .  .  .  o  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

DeviceA#
```

The routing table information on device B is as follows:

```
DeviceB# display ip pim-info sparse-mode mroute
```

```
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 2
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(10.100.1.2, 232.1.1.10)
RPF nbr: 10.100.2.1
RPF idx: GigabitEthernet 0/31
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 40 seconds
kat expires in 190 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  i  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  j  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  .  o  .  o  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

(10.100.12.2, 232.1.1.10)
RPF nbr: 10.100.5.1
RPF idx: GigabitEthernet 0/32
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 40 seconds
kat expires in 125 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  i  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  .  .  .  o  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

DeviceB#
```

The routing table information on device C is as follows:

```
DeviceC# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 2
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(10.100.1.2, 232.1.1.10)
RPF nbr: 10.100.5.2
RPF idx: GigabitEthernet 0/20
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 12 seconds
kat expires in 72 seconds
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Joined
0 . . . . . j . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Asserted
```

```
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(10.100.12.2, 232.1.1.10)
RPF nbr: 10.100.11.2
RPF idx: GigabitEthernet 0/18
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 8 seconds
kat expires in 98 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . . . j . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
```

```
1  . . . . . .
2  . . . . . .
3  . . . . . .
4  . . . . . .
5  . . . . . .
6  . . . . . .
7  . . . . . .
Outgoing
0  . . . . o . . . . . .
1  . . . . . .
2  . . . . . .
3  . . . . . .
4  . . . . . .
5  . . . . . .
6  . . . . . .
7  . . . . . .

DeviceC#
```

The routing table information on device D is as follows:

```
DeviceD# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 1
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(10.100.12.2, 232.1.1.10)
RPF nbr: 0.0.0.0
RPF idx: None
SPT bit: 1
Upstream State: JOINED
kat expires in 62 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  . . . . . .
1  . . . . . .
2  . . . . . .
3  . . . . . .
4  . . . . . .
5  . . . . . .
6  . . . . . .
7  . . . . . .
Joined
```

```
0 . . j . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Asserted
0 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Outgoing
0 . . o . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .

DeviceD#
```

The routing table information on device E is as follows:

```
DeviceE# display ip pim-info sparse-mode mroute
IP Multicast Routing Table

(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 1
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(10.100.1.2, 232.1.1.10)
RPF nbr: 10.100.4.1
RPF idx: GigabitEthernet 0/7
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 39 seconds
kat expires in 189 seconds
```

```
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  i  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  o  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .

DeviceE#
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip access-control-list standard 59
 10 permit 232.1.1.0 0.0.0.255
```

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.1.1 255.255.255.0
ip pim sparse-mode
!
interface GigabitEthernet 0/24
 unset switch-port
 ip address 10.100.2.1 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.1.0 0.0.0.255 area 0
 network-advertise 10.100.2.0 0.0.0.255 area 0
!
ip pim enable-ssm range 59
!
```

- Device B configuration file

```
hostname DeviceB
!
ip access-control-list standard 59
 10 permit 232.1.1.0 0.0.0.255
!
ip mrouting-enable
!
interface GigabitEthernet 0/30
 unset switch-port
 ip address 10.100.8.1 255.255.255.0
 ip igmp specify-version 3
 ip pim sparse-mode
!
interface GigabitEthernet 0/31
 unset switch-port
 ip address 10.100.2.2 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/32
 unset switch-port
 ip address 10.100.5.2 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 49
 ip address 49.49.49.49 255.255.255.255
```

```
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.2.0 0.0.0.255 area 0
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 10.100.8.0 0.0.0.255 area 0
 network-advertise 49.49.49.49 0.0.0.0 area 0
!
ip pim enable-ssm range 59
!
```

- Device C configuration file

```
hostname DeviceC
!
ip access-control-list standard 59
 10 permit 232.1.1.0 0.0.0.255
!
ip mrouting-enable
!
interface GigabitEthernet 0/18
 unset switch-port
 ip address 10.100.11.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/20
 unset switch-port
 ip address 10.100.5.1 255.255.255.0
 ip pim sparse-mode
!
interface GigabitEthernet 0/22
 unset switch-port
 ip address 10.100.4.1 255.255.255.0
 ip pim sparse-mode
!
interface Loopback 50
 ip address 50.50.50.50 255.255.255.255
ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.4.0 0.0.0.255 area 0
 network-advertise 10.100.5.0 0.0.0.255 area 0
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 50.50.50.50 0.0.0.0 area 0
!
ip pim enable-ssm range 59
```

```
!
```

- Device D configuration file

```
hostname DeviceD
!
ip access-control-list standard 59
 10 permit 232.1.1.0 0.0.0.255
!
ip mrouting-enable
!
interface GigabitEthernet 0/47
 unset switch-port
 ip address 10.100.12.1 255.255.255.0
ip pim sparse-mode
!
interface GigabitEthernet 0/48
 unset switch-port
 ip address 10.100.11.2 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 10.100.12.0 0.0.0.255 area 0
!
ip pim enable-ssm range 59
!
```

- Device E configuration file

```
hostname DeviceE
!
ip access-control-list standard 59
 10 permit 232.1.1.0 0.0.0.255
!
ip mrouting-enable
!
interface GigabitEthernet 0/5
 unset switch-port
 ip address 10.100.13.1 255.255.255.0
 ip igmp specify-version 3
 ip pim sparse-mode
!
interface GigabitEthernet 0/7
 unset switch-port
 ip address 10.100.4.2 255.255.255.0
 ip pim sparse-mode
!
```

```
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.4.0 0.0.0.255 area 0
 network-advertise 10.100.13.0 0.0.0.255 area 0
!
ip pim enable-ssm range 59
!
```

7. Common Errors

- An IPv4 unicast route is incorrectly configured.
- IPv4 multicast routing function is not enabled on a multicast device.
- The SSM function is not enabled on a multicast device, or the SSM group address range on this multicast device is not consistent with the configuration on other multicast devices.
- The PIM-SM function is not enabled on an interface, for example, the gateway interface connecting to hosts or a multicast source.
- The IGMPv3 function is not enabled on an interface connected to hosts.

Contents

1 Configuring PIM-DM....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 PIM-DM Network Deployment....1
1.1.3 Principles....1
1.1.4 PIM-DM and MIB....4
1.1.5 Protocols and Standards....4
1.2 Restrictions and Guidelines....4
1.3 Configuration Task Summary....4
1.4 Configuring Basic Functions of PIM-DM....4
1.4.1 Overview....4
1.4.2 Procedure....4
1.5 Configuring PIM-DM Passive Mode....5
1.5.1 Overview....5
1.5.2 Procedure....5
1.6 Configuring PIM Neighbors....6
1.6.1 Overview....6
1.6.2 Procedure....6
1.7 Configuring PIM State Refresh....7
1.7.1 Overview....7
1.7.2 Restrictions and Guidelines....7
1.7.3 Procedure....7

1.8 Enabling the PIM-BFD Function 8
1.8.1 Overview 8
1.8.2 Procedure 8
1.9 Enabling the MIB Function 8
1.9.1 Overview 8
1.9.2 Procedure 8
1.10 Monitoring 9
1.11 Configuration Examples 9
1.11.1 Configuring Basic Functions of PIM-DM 9

1 Configuring PIM-DM

1.1 Introduction

1.1.1 Overview

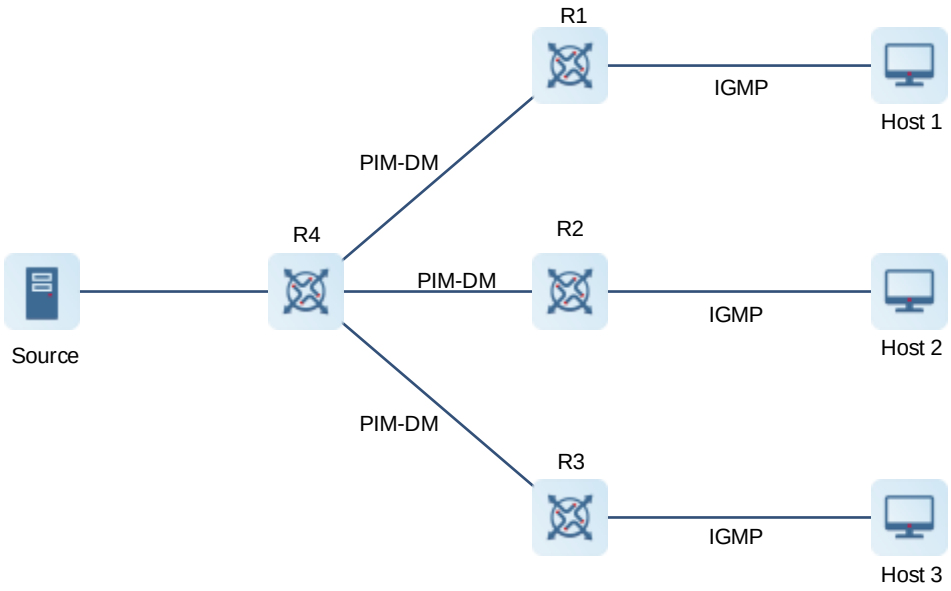
In multicast packet transmission, devices need to run multicast routing protocols to control multicast packet forwarding. A multicast network can be composed of multiple domains. The most typical multicast routing protocol is Protocol Independent Multicast (PIM). Based on implementation mechanism, PIM is categorized into Protocol Independent Multicast-Dense Mode (PIM-DM) and Protocol Independent Multicast-Sparse Mode (PIM-SM).

PIM-DM runs on multicast devices and is used to discover and maintain multicast routing information and determine a forwarding route of multicast packets.

Compared with PIM-SM, PIM-DM applies to small networks where multicast group members densely reside. Ideally, multicast group members on every path need to receive packets in the PIM network, and each forwarding node pushes and distributes the multicast packets in the entire network.

1.1.2 PIM-DM Network Deployment

Figure 1-1 PIM-DM Deployment Topology



As shown in Figure 1-1, when a multicast device is connected to other multicast devices, PIM must be deployed on this multicast device. Here, PIM-DM is deployed.

1.1.3 Principles

The basic principles of PIM-DM are as follows:

(1) Assume that all paths in a network have multicast group members and the multicast source sends multicast packets throughout the entire network.

(2) The entire path can be considered a tree with the source as the root node and group members as leaf nodes. This tree is the shortest path tree (SPT). Initially, some leaf nodes are not group members and they must be pruned to get a correct tree structure.

(3) If a new host joins the tree or an existing node leaves the tree, the tree structure must be readjusted. Multicast packets are forwarded on the paths of this tree, and the forwarding path is the shortest path from the source to the group members.

The working mechanisms of PIM-DM include neighbor creation, flooding, pruning, state refresh, graft, and assertion.

1. Neighbor Creation

Neighbors are the basis of PIM-DM. Neighbor relationships must be created between PIM routers before PIM control messages can be exchanged or multicast packets can be forwarded.

Neighbor creation is implemented based on the hello messages sent by the PIM interfaces and is composed of neighbor discovery, protocol parameter negotiation, and neighbor relationship maintenance.

- Discover neighbors

  A PIM interface periodically sends hello messages with the destination address being 224.0.0.13 (ff02::d for IPv6), source address being the IP address of the interface, and time to live (TTL) value being **1**. PIM routers in the same network segment receive the packets with the destination address 224.0.0.13 (ff02::d for IPv6). In this way, the PIM routers obtain neighbor information and create neighbor relationships.

  When a PIM interface is enabled or detects a new neighbor, a Triggered-Hello-Delay message is used to generate a random time. Within the time, the interface sends hello messages.

- Negotiate protocol parameters

  Protocol parameters are negotiated by hello messages. The key parameters are described as follows:

  - **DR_Priority**: The priority of a router for designated router (DR) election. The router with a higher priority is selected.
  - **Holdtime**: The neighbor is considered reachable within the hold time.
  - **Propagation Delay**: Duration between the time when a PIM device receives a prune packet and the time when the PIM device sends the prune packet to the upstream devices.
  - **Override-Interval**: The interval to override the prune.

- Maintain neighbor relationship

  A hello message is sent periodically between PIM routers. If a hello message is not received from a PIM neighbor within the hold time, the neighbor is considered unreachable and is deleted from the neighbor list. Any change of PIM neighbors will cause change of the multicast topology in the network. If an upstream or a downstream neighbor in a multicast distribution tree (MDT) is unreachable, multicast routes converge again and the MDT is reshaped.

2. Flooding

Multicast packets flooded to an entire network from a source is referred to flooding. Flooding enables all nodes in a network to receive these packets.

Multicast devices check reverse path forwarding (RPF) in the packets. If the checking fails, the packets are discarded. If the checking is passed and an outbound interface is detected, the multicast packets are forwarded through the outbound interface. This enables all node members in the network to receive these multicast packets.

3. Pruning

If no group member exists on a path branch, the path branch is pruned. This enables packets to be forwarded to paths with group members available.

(1) Multicast devices check RPF in the packets. If the checking is passed but no outbound interface exists, no group member exists on this path. A prune packet is sent to the upstream device.

(2) Upon receiving the prune packet, the upstream device marks the packet sending interface as pruned and sets a prune state timer.

(3) After all unnecessary paths are pruned, the remaining paths are necessary and they form an SPT.

- Prune override

After a PIM router sends a prune packet to the upstream device, the PIM router does not prune the downstream path immediately. If the downstream device has multicast requirements, it can send a prune override message to the PIM router to override the prune.

If the PIM router receives a prune override message from the downstream device in the prune-pending timer (PPT), the PIM router cancels pruning. PPT = Propagation delay + Override interval

4. State Refresh

A prune state timer is maintained on a device during pruning. When the timer times out, the device continues to forward packets. PIM-DM refreshes the timer state using a state refresh message (SRM) to avoid unnecessary prune state changes.

The multicast device nearest to the source periodically sends the SRM to the downstream devices. Upon receiving the SRM, the downstream devices add the local topology state information to the SRM by modifying some fields in the SRM, and continue to forward the SRM to downstream devices. Finally, the SRM arrives leaf devices and the states of all devices in the network get updated.

5. Graft

If a group member sends data to a group corresponding to a path being pruned, the path must be restored to the forwarding state. This is referred to as graft.

(1) When a group member sends data to a group corresponding to a path being pruned, the other devices on this path receive the join message and send a graft packet to the upstream device to apply for joining the SPT.

(2) Upon receiving the graft packet, the upstream device replies a Graft-Ack packet and starts to forward multicast packets to the downstream devices on this path.

(3) The downstream devices receive the Graft-Ack packet. If the downstream devices do not receive the Graft-Ack packet, they send the graft packet again and again until they receive the Graft-Ack packet.

6. Assertion

In a network segment where multiple PIM routers reside, the same multicast packet may be transmitted to the routers in the network segment multiple times. In this case, a special PIM router can be specified to forward multicast packets and avoid duplicate forwarding. This is referred to as assertion.

1.1.4 PIM-DM and MIB

Connected to other agents, the Simple Network Management Protocol (SNMP) manager uses information in the Management Information Base (MIB) to directly manage the PIM-DM function.

The MIB specifies variables (namely information that can be queried and set by the management process) maintained by network elements and directly manages the PIM-DM function.

1.1.5 Protocols and Standards

- RFC 3973: Protocol Independent Multicast - Dense Mode (PIM-DM)
- RFC 2715: Interoperability Rules for Multicast Routing Protocols

1.2 Restrictions and Guidelines

- If PIM-DM neighbor relationships between devices are created through links and downstream devices have no or few needs to receive multicast packets, the system performance is excessively consumed. In this scenario, PIM-SM is recommended for the environment deployment.
- Before PIM-DM is used, ensure that unicast routes are reachable in the network.

1.3 Configuration Task Summary

The PIM-DM configuration includes the following tasks:

(1) Configuring Basic Functions of PIM-DM

(2) (Optional) Configuring PIM-DM Passive Mode

(3) (Optional) Configuring PIM Neighbors

(4) (Optional) Configuring PIM State Refresh

(5) (Optional) Enabling the PIM-BFD Function

(6) (Optional) Enabling the MIB Function

1.4 Configuring Basic Functions of PIM-DM

1.4.1 Overview

A PIM network is created to provide IPv4 multicast services for sources and hosts in the network. Neighbor relationships must be created between PIM routers before PIM control messages can be exchanged or multicast packets can be forwarded.

1.4.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IPv4 multicast routing.

ip mrouting-enable

The multicast routing function is disabled by default.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(5) Enable the PIM-DM function.

ip pim dense-mode

The PIM-DM function is disabled on an interface by default.

The function can be enabled on the interconnection interfaces of routers, and interfaces connecting to multicast sources and user hosts.

(6) (Optional) Enable the PIM-DM function on a super VLAN interface.

ip pim dense-mode subvlan [*subvlan-id* | **all**]

The PIM-DM function is disabled on a super VLAN interface by default.

In most scenarios, the PIM-DM protocol is disabled by default and not needed on a super VLAN interface. This interface does not send or receive PIM packets. To enable PIM-DM to run on the super VLAN interface, specify a sub VLAN or all sub VLANs to receive packets. Note that if all sub VLANs are specified to receive packets, the transmission performance may be reduced, causing neighbor flapping.

1.5 Configuring PIM-DM Passive Mode

1.5.1 Overview

In a PIM network, if you want an interface, for example, a stub network device interface connecting to hosts, to receive multicast packets without involving the PIM network topology construction, you can set the interface to the PIM-DM passive mode. After the PIM-SM passive mode is configured, the interface does not process PIM-DM packets.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Enable the PIM-DM passive mode on an interface.

ip pim dense-mode passive

The PIM-DM passive mode is disabled on an interface by default.

1.6 Configuring PIM Neighbors

1.6.1 Overview

PIM devices support adjustment to protocol parameters in hello messages and neighbor filtering to improve network security.

1.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure PIM neighbors. The following configuration tasks are optional. Select tasks for configuration according to actual condition.

- Configure hello message sending interval.

  ip pim query-period *query-period*

  Hello messages are sent at an interval of **30** seconds by default.

  Every time when the hello message sending interval is updated, the hello message hold time value is automatically updated as 3.5 times the interval. If the product of the hello message sending interval and 3.5 is greater than 65,535, the hello message hold time value is updated as 65,535.

- Configure propagation delay of hello messages on an interface.

  ip pim propagation-delay *propagation-delay-time*

  The default propagation delay of hello messages on an interface is 500 ms.

- Configure prune override interval of hello messages.

  ip pim override-interval *override-interval*

  The default prune override interval of hello messages is 2500 ms.

- Configure PIM neighbor filtering.

  ip pim peering-filter { *acl-name* | *acl-number* }

  The neighbor filtering function is disabled on an interface by default.

  Only addresses that meet ACL filtering conditions can be used as PIM neighbors of the current interface. Otherwise, the addresses filtered out cannot be neighbors.

1.7 Configuring PIM State Refresh

1.7.1 Overview

PIM devices support enabling or disabling PIM state refresh and adjusting PIM SRM sending interval.

1.7.2 Restrictions and Guidelines

- If the state refresh function is disabled, SRMs are not processed or forwarded.
- Disabling the state refresh function may cause the converged PIM-DM MDT to re-converge, which leads to unnecessary bandwidth waste and multicast routing table flapping. Therefore, you are not advised to disable this function in general conditions.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the PIM-DM state refresh function. The following configuration tasks are optional. Select tasks for configuration according to actual condition.

- Disable the PIM-DM state refresh function on an interface.

  ip pim refresh-state disable

  The PIM-DM SRMs are processed and forwarded by default.

- Configure the PIM-DM SRM sending interval on an interface.

  ip pim refresh-state origination-period *origination-period*

  The PIM-DM SRMs are sent at an interval of **60** seconds by default.

1.8 Enabling the PIM-BFD Function

1.8.1 Overview

PIM-DM supports correlation with the bidirectional forwarding detection (BFD) function. After the PIM-BFD function is enabled on the interconnection interfaces of routers, the interfaces are used as BFD interfaces to quickly detect neighbor changes and specify a router to forward multicast packets in a specified network segment based on the assertion rule.

1.8.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Enable the PIM-BFD function on an interface.

ip pim bfd

The PIM-BFD function is disabled on an interface by default.

1.9 Enabling the MIB Function

1.9.1 Overview

The MIB function allows PIM devices to manage PIM-DM using the elements in the MIB.

1.9.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the MIB function to manage PIM-DM.

ip pim mib dense-mode

PIM-DM is not managed by the MIB function by default.

1.10 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup ip pim-info dense-mode track	Clears statistical information about PIM-DM protocol packets.
display ip pim-info dense-mode interface [*interface-type interface-number*] [**detail**]	Displays PIM-DM interface information.
display ip pim-info dense-mode neighbor [*interface-type interface-number*]	Displays the PIM-DM neighbor information.
display ip pim-info dense-mode nexthop	Displays the PIM-DM next hop information.
display ip pim-info dense-mode mroute [*ipv4-group-or-source-address-1* [*ipv4-group-or-source-address-2*]] [**summary**]	Displays information about the PIM-DM routing table.
display ip pim-info dense-mode track	Displays statistical information about PIM-DM protocol packets.

1.11 Configuration Examples

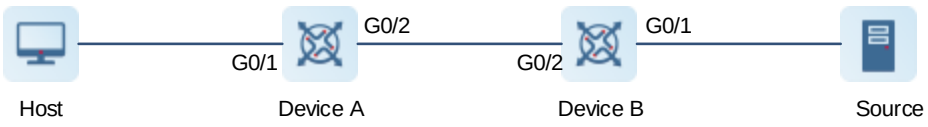
1.11.1 Configuring Basic Functions of PIM-DM

1. Requirements

Configure the basic functions of PIM-DM on multiple devices to form a PIM-DM network, so that a host can receive multicast traffic from a multicast source.

2. Topology

Figure 1-2 Topology of Basic PIM-DM



3. Notes

- Configure an IPv4 unicast routing protocol, for example, OSPF, on routers.
- Enable the IPv4 multicast routing function on all routers.
- Configure basic functions of PIM-DM on the interconnection interfaces of routers and interfaces connecting to multicast sources and user hosts.

4. Procedure

(1) Configure IP addresses for interfaces and enable unicast routing protocols (omitted).

(2) Configure the IPv4 multicast routing function.

Configure the IPv4 multicast routing function on Device A.

```
Device A> privilege
Device A#config-system terminal
Device A(cfg)#ip mrouting-enable
```

Configure the IPv4 multicast routing function on Device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ip mrouting-enable
```

(3) Configure basic functions of PIM-DM.

Configure basic functions of PIM-DM on Device A.

```
Device A(cfg)# interface GigabitEthernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip pim dense-mode
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface GigabitEthernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# ip pim dense-mode
```

Configure basic functions of PIM-DM on Device B.

```
Device B(cfg)# interface GigabitEthernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip pim dense-mode
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface GigabitEthernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# ip pim dense-mode
```

5. Verification

Verify that multicast source (192.168.1.10) can send packets to G (229.1.1.1), and that Receiver A joins G, so the host can receive packets from G.

Check the PIM-DM routing tables on Device A and Device B.

```
DeviceA# display ip pim-info dense-mode mroute
PIM-DM Multicast Routing Table
(192.168.1.10, 229.1.1.1)
  MRT lifetime expires in 182 seconds
  Source directly connected on GigabitEthernet 0/1
  State-Refresh Originator State: Originator
   SRT:57, SAT:147
```

```
  Upstream IF: GigabitEthernet 0/1
    Upstream State: Forwarding
    Assert State: NoInfo
  Downstream IF List:
    GigabitEthernet 0/2, in 'olist':
      Downstream State: NoInfo
      Assert State: NoInfo
```

6. Configuration Files

- Device A configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim dense-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ip pim dense-mode
!
```

- Device B configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim dense-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ip pim dense-mode
!
```

7. Common Errors

- An IPv4 unicast route is incorrectly configured.
- IPv4 multicast routing is not enabled on a router.
- PIM-DM function is not enabled on an interface.

Contents

1 Configuring IGMP Snooping 1
1.1 Introduction 1
1.1.2 Principles 1
1.1.3 Port Types 3
1.1.4 Working Mechanism 3
1.1.5 Protocols and Standards 4
1.2 Configuration Task Summary 4
1.3 Configuring Basic IGMP Snooping Functions 5
1.3.1 Overview 5
1.3.2 Restrictions and Guidelines 5
1.3.3 Configuration Tasks 6
1.3.4 Notes 6
1.3.5 Configuring Basic IGMP Snooping Functions in IVGL Mode 7
1.3.6 Configuring Basic IGMP Snooping Functions in SVGL Mode 7
1.3.7 Configuring Basic IGMP Snooping Functions in IVGL-SVGL Mode 8
1.4 Configuring Protocol Packet Processing Parameters 9
1.4.1 Overview 9
1.4.2 Restrictions and Guidelines 9
1.4.3 Configuration Tasks 9
1.4.4 Configuring a Static Multicast Router Port 10
1.4.5 Configuring a Static Member Port 10
1.4.6 Configuring Report Packet Suppression 11

1.4.7 Configuring Port Fast Leave 12

1.1.1 Configuring Dynamic Multicast Router Port Learning 12

1.4.8 Configuring the Aging Time for Dynamic Multicast Router Ports 13

1.4.9 Configuring the Aging Time for Dynamic Member Ports 13

1.5 Configuring IGMP Snooping Querier 14

1.5.1 Overview 14

1.5.2 Restrictions and Guidelines 14

1.5.3 Prerequisites 14

1.5.4 Configuration Tasks 14

1.5.5 Enabling IGMP Snooping Querier 15

1.5.6 Configuring the Source IP Address of a Querier 15

1.5.7 Configuring the IGMP Snooping Querier Version 16

1.5.8 Configuring the Maximum Response Time for Query Packets 16

1.5.9 Configuring the Interval for a Querier to Send Query Packets 17

1.5.10 Configuring the Querier Aging Time 17

1.6 Configuring Multicast Security Control 18

1.6.1 Overview 18

1.6.2 Prerequisites 18

1.6.3 Configuration Tasks 18

1.6.4 Configuring a Profile 18

1.6.5 Configuring a Range of Multicast Group Addresses for a Profile 19

1.6.6 Configuring Multicast Preview 20

1.6.7 Configuring the Maximum Number of Multicast Groups Allowed for Concurrent Request 21

1.6.8 Configuring Source Port Check 22

1.6.9 Configuring Source IP Address Check 22
1.7 Configuring QinQ Processing 23
1.7.1 Overview 23
1.7.2 Restrictions and Guidelines 23
1.7.3 Prerequisites 24
1.7.4 Procedure 24
1.8 Monitoring 24
1.9 Configuration Examples 25
1.9.1 Configuring Basic IGMP Snooping Functions in IVGL Mode 25
1.9.2 Configuring Basic IGMP Snooping Functions in SVGL Mode 30
1.9.3 Configuring Basic IGMP Snooping Functions in IVGL-SVGL Mode 35
1.9.4 Configuring Static Ports to Implement L2 Multicast 43
1.9.5 Configuring IGMP Snooping Querier 47

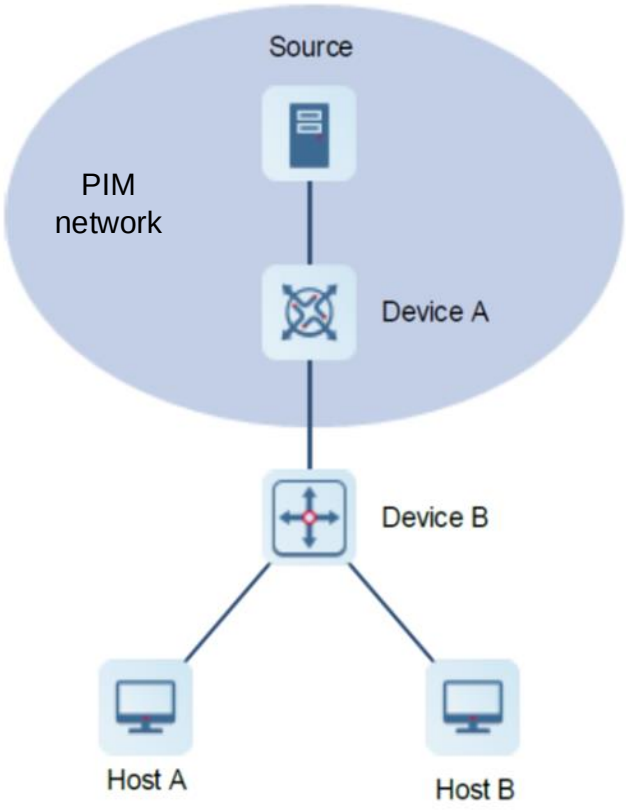
1 Configuring IGMP Snooping

1.1 Introduction

Internet Group Management Protocol (IGMP) snooping is a process of snooping protocol packets between an L3 multicast device and hosts to manage and control forwarding of IP multicast traffic at the data link layer, implementing L2 multicast.

Generally, multicast packets need to pass through L2 switches, especially in some local area networks (LANs). As shown in Figure 1-1, the Protocol Independent Multicast (PIM) multicast device sends packets to hosts through an L2 switch. Because multicast group addresses cannot be learned on L2 devices, the packets will be broadcast in the VLAN. This wastes bandwidth and affects security. After the IGMP snooping function is configured, the L2 switch can listen to IGMP packets between hosts and the uplink PIM multicast device and establish L2 multicast forwarding entries to ensure that multicast data is forwarded only to specific hosts. This prevents multicast data from being broadcast in the L2 network.

Figure 1-1 Network Topology



1.1.2 Principles

IGMP snooping is a basic L2 multicast function that forwards and controls multicast data at the data link layer. By listening to and analyzing protocol packets between hosts and the L3 multicast device, the L2 device

running IGMP snooping establishes and maintains L2 multicast forwarding entries to ensure on-demand forwarding of multicast data at the link layer.

As shown in Figure 1-2, when IGMP snooping is not configured on the L2 device, IP multicast packets are broadcast in the VLAN. As shown in Figure 1-3, when IGMP snooping is enabled on the L2 device, IP multicast packets are forwarded only to the member hosts of the multicast group. This prevents multicast data from being broadcast in the L2 network.

Figure 1-2 Multicast Data Transmission When IGMP Snooping Is Disabled

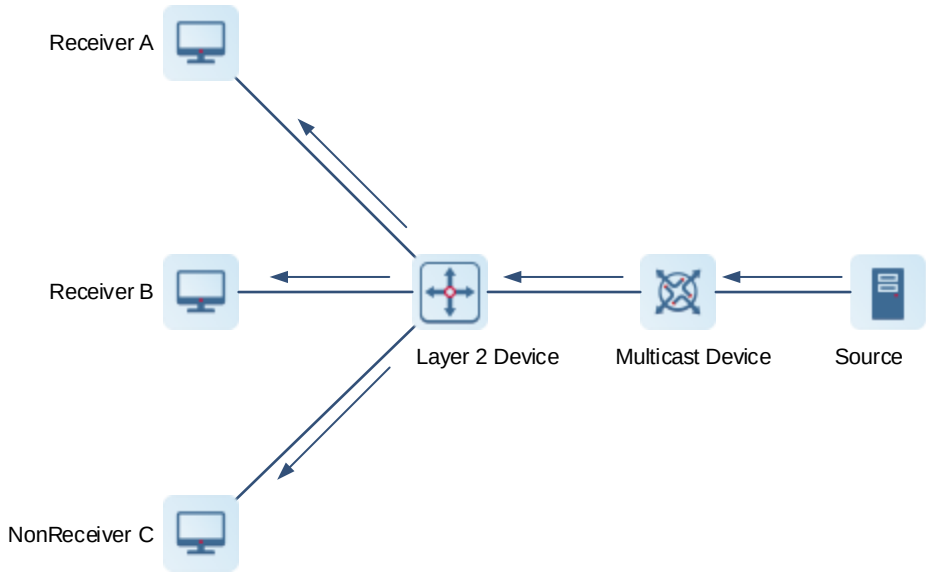
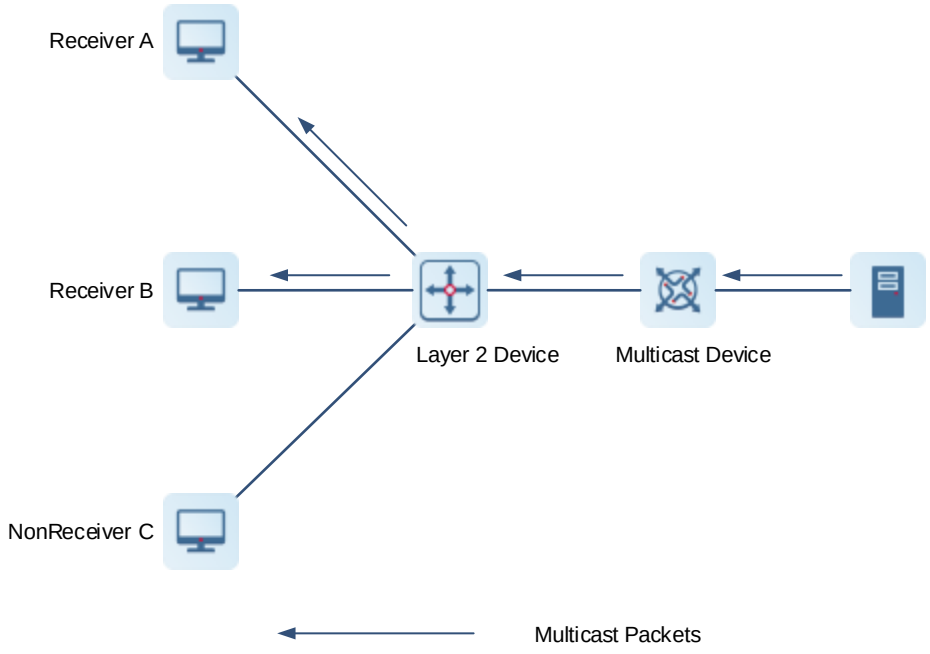


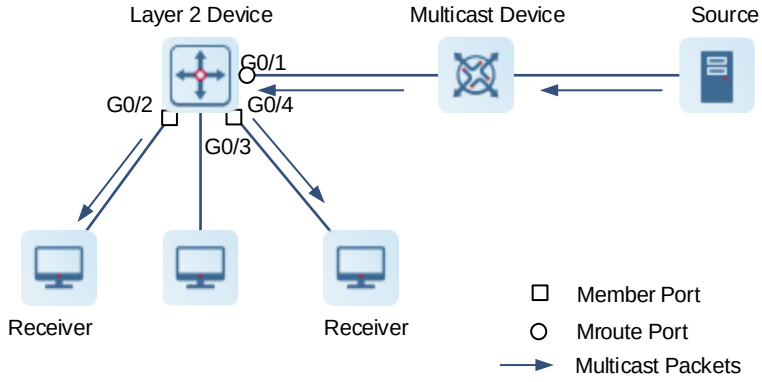
Figure 1-3 Multicast Data Transmission When IGMP Snooping Is Enabled



1.1.3 Port Types

IGMP snooping is enabled on a per-VLAN basis. IGMP snooping ports are ports in a VLAN. The device running IGMP snooping identifies ports in a VLAN as multicast router ports or member ports to manage and control forwarding of IP multicast data in the VLAN.

Figure 1-4 IGMP Snooping Ports



- A multicast router port connects the L2 multicast device to an L3 multicast device, and indicates the direction of the multicast source. By listening to IGMP packets, the L2 multicast device can automatically discover and maintain dynamic multicast router ports. You can also configure static multicast router ports.
- A member port connects the L2 multicast device to a host in a multicast group, and indicates the direction of a multicast group member. It is also called a listener port. By listening to IGMP packets, the L2 multicast device can automatically discover and maintain dynamic member ports. You can also configure static member ports.

As shown in Figure 1-4, IGMP snooping is enabled on the L2 multicast device, and the multicast data is received from a multicast router port and sent out from a member port.

1.1.4 Working Mechanism

A device running IGMP snooping analyzes received IGMP packets to discover and identify multicast router ports and member ports and establish and maintain IGMP snooping forwarding entries. The device can identify and process the following types of IGMP packets:

1. Query Packets

The IGMP querier periodically sends a general Query packet, and sends a group-specific Query packet when it receives a Leave packet. When the device running IGMP snooping receives a Query packet, the device performs the following actions in the VLAN:

- The device forwards the IGMP Query packet through all ports (except the port receiving the packet).
- If the port receiving the Query packet is a dynamic multicast router port, the device resets the aging timer of the port. After the timer times out, the port is no longer used as a dynamic multicast router port.
- If the port receiving the Query packet is not a dynamic multicast router port, the device uses the port as the dynamic multicast router port and starts the aging timer. After the timer times out, the port is no longer used as a dynamic multicast router port.

- If the function of dynamical multicast router port learning is disabled, the device will not learn dynamic multicast router ports.

2. Report Packets

When a member host receives a Query packet, the host will respond with a Report packet. If a host wants to join a multicast group, the host will proactively send a Report packet. The device running IGMP snooping can process IGMPv1 and IGMPv2 packets by default. For Report packets of IGMPv3, the device only processes the group information and does not process the carried source information.

When the device running IGMP snooping receives a Report packet, the device performs the following actions in the VLAN:

- The device forwards the Report packet through all multicast router ports. If Report packet suppression is enabled on the device, the device forwards only the first Report packet of each group in an IGMP query interval.
- If the port receiving the Report packet is a dynamic member port, the device resets the aging timer of the port. After the timer times out, the port is no longer used as a dynamic member port.
- If the port receiving the Report packet is not a member port, the device uses the port as the dynamic member port and starts the aging timer. After the timer times out, the port is no longer used as a dynamic member port.

3. Leave Packets

If a host wants to leave a multicast group, the host will proactively send a Leave packet. When the device running IGMP snooping receives a Leave packet, the device performs the following actions in the VLAN:

- The device forwards the Leave packet through all multicast router ports.
- If the port receiving the Leave packet is a dynamic member port and the fast leave function is configured, the device immediately deletes the port from the IGMP snooping forwarding entry of the corresponding multicast group. The port is no longer used as a dynamic member port of the group.
- If the port receiving the Leave packet is a dynamic member port but the fast leave function is not configured, the port status remains unchanged. When the aging timer of the port times out, the device deletes the port from the IGMP snooping forwarding entry of the corresponding multicast group. The port is no longer used as a dynamic member port of the group.

1.1.5 Protocols and Standards

- RFC 4541: Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches

1.2 Configuration Task Summary

IGMP snooping configuration includes the following tasks:

(1) Configuring Basic IGMP Snooping Functions. Select either of the following configuration tasks to configure.

- Configuring Basic IGMP Snooping Functions in IVGL Mode
- Configuring Basic IGMP Snooping Functions in SVGL Mode
- Configuring Basic IGMP Snooping Functions in IVGL-SVGL Mode

(2) Configuring Protocol Packet Processing Parameters. All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Static Multicast Router Port
- Configuring a Static Member Port
- Configuring Report Packet Suppression
- Configuring Port Fast Leave
- Configuring Dynamic Multicast Router Port Learning
- Configuring the Aging Time for Dynamic Multicast Router Ports
- Configuring the Aging Time for Dynamic Member Ports

(3) (Optional) Configuring IGMP Snooping Querier. The tasks are as follows:

- Enabling IGMP Snooping Querier
- Configuring the Source IP Address of a Querier
- (Optional) Configuring the IGMP Snooping Querier Version
- (Optional) Configuring the Maximum Response Time for Query Packets
- (Optional) Configuring the Interval for a Querier to Send Query Packets
- (Optional) Configuring the Querier Aging Time

(4) Configuring Multicast Security Control. All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Profile
- Configuring a Range of Multicast Group Addresses for a ProfileConfiguring Multicast Preview
- Configuring the Maximum Number of Multicast Groups Allowed for Concurrent Request
- Configuring Source Port Check
- Configuring Source IP Address Check

(5) (Optional) Configuring QinQ Processing

1.3 Configuring Basic IGMP Snooping Functions

1.3.1 Overview

The device running IGMP snooping can provide independent or shared multicast services for user VLANs when operated in the following modes:

- Independent VLAN Group Learning (IVGL): provides independent multicast services for each user VLAN.
- Shared VLAN Group Learning (SVGL): provides shared multicast services for multiple user VLANs.
- IVGL-SVGL: provides both shared and independent multicast services for user VLANs.

1.3.2 Restrictions and Guidelines

- The SVGL mode and the IP multicast function are mutually exclusive. If the IP multicast function needs to be enabled on the device running IGMP snooping, the device can operate only in IVGL mode.
- If the device running IGMP snooping operates in SVGL or IVGL-SVGL mode, the multicast groups associated with the SVGL mode must be configured. First, define the multicast groups applied in the SVGL

mode in a profile. Then, apply this profile in the command for configuring the group range in SVGL mode.

- PIM snooping depends on the IVGL mode of IGMP snooping. If the PIM snooping function needs to be enabled on the device running IGMP snooping, the device can operate only in IVGL mode.

1.3.3 Configuration Tasks

Basic IGMP snooping function configuration includes the following tasks. Please configure only one task.

- Configuring Basic IGMP Snooping Functions in IVGL Mode
- Configuring Basic IGMP Snooping Functions in SVGL Mode
- Configuring Basic IGMP Snooping Functions in IVGL-SVGL Mode

Basic IGMP snooping function configuration includes the following tasks.

1.3.4 Notes

The following table lists products supporting different basic IGMP snooping functions.

Table 1-1 Products Supporting Different Basic IGMP Snooping Functions

Function	Product Supporting the Function
Configuring Basic IGMP Snooping Functions in IVGL Mode	S6980-64QC S6920-4C S6580-24DC8QC S6510-32CQ S6510-48VS8CQ S6250-48XS8CQ
Configuring Basic IGMP Snooping Functions in SVGL Mode	S6980-64QC S6920-4C S6580-24DC8QC S6510-32CQ S6510-48VS8CQ S6250-48XS8CQ
Configuring Basic IGMP Snooping Functions in IVGL-SVGL Mode	S6980-64QC S6920-4C S6580-24DC8QC S6510-32CQ S6510-48VS8CQ S6250-48XS8CQ

1.3.5 Configuring Basic IGMP Snooping Functions in IVGL Mode

1. Overview

In IVGL mode, the device running IGMP snooping provides independent multicast services for each user VLAN. Multicast traffic can be forwarded only in the belonged VLAN, and hosts can request multicast traffic only in the belonged VLAN.

2. Restrictions and Guidelines

- Unless otherwise specified, you are advised to enable IGMP snooping globally on all L2 access devices connecting to hosts.
- After IGMP snooping is enabled globally, it takes effect to all VLANs. You can disable IGMP snooping on any VLAN. When the multicast service is disabled on a VLAN, the multicast services on other VLANs are not affected.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IGMP snooping globally and run the IVGL mode.

ip igmp snooping-mode ivgl

IGMP snooping is disabled by default.

(4) (Optional) Disable IGMP snooping on a VLAN.

unset ip igmp snooping-mode vlan *vlan-id*

After IGMP snooping is enabled globally, it takes effect to all VLANs by default.

If IGMP snooping is disabled globally, it is ineffective on all VLANs.

You can disable IGMP snooping on a VLAN only after IGMP snooping is enabled globally.

1.3.6 Configuring Basic IGMP Snooping Functions in SVGL Mode

1. Overview

In SVGL mode, the device running IGMP snooping can provide shared multicast services for multiple user VLANs. Shared multicast services are usually used to provide the same video on demand (VOD) service for users in multiple VLANs. Compared with independent multicast services, shared multicast services save bandwidth.

VLANs for shared multicast services are classified into a shared VLAN and sub VLANs. Multicast traffic of multicast groups applied in SVGL mode on the shared VLAN is forwarded from the shared VLAN to sub VLANs, and hosts on sub VLANs request multicast traffic of multicast groups applied in SVGL mode from the shared VLAN.

2. Restrictions and Guidelines

- To share multicast services among VLANs, you need to configure multicast groups applied in SVGL mode after configuring the SVGL mode.

- Shared multicast services apply only to the shared VLAN and sub VLANs and use group addresses applied in SVGL mode.

3. Prerequisites

Before configuring multicast groups associated with the SVGL mode, ensure that the corresponding profile is created and a range of multicast groups that are permitted or denied by the filter is specified. For details about profile configuration, see 1.6.4 Configuring a Profile.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IGMP snooping globally and run the SVGL mode.

ip igmp snooping-mode svgl

IGMP snooping is disabled by default.

(4) Configure multicast groups associated with the SVGL mode.

ip igmp snooping-mode svgl group-profile *profile-number*

No multicast group is associated with the SVGL mode by default.

(5) (Optional) Specify a shared VLAN.

ip igmp snooping-mode svgl shared-vlan *vlan-id*

The default shared VLAN in SVGL mode is VLAN 1.

(6) (Optional) Specify sub VLANs in SVGL mode.

ip igmp snooping-mode svgl subvlan *vlan-range*

By default, all VLANs except the shared VLAN are sub VLANs in SVGL mode.

1.3.7 Configuring Basic IGMP Snooping Functions in IVGL-SVGL Mode

1. Overview

The IVGL-SVGL mode is also called the hybrid mode. In IVGL-SVGL mode, the device running IGMP snooping can provide both shared and independent multicast services for user VLANs.

- In the shared VLAN and sub VLANs, shared multicast services apply to the multicast traffic of multicast groups applied in SVGL mode, and independent multicast services apply to other multicast traffic.
- In other VLANs (except the shared VLAN and sub VLANs), independent multicast services are provided.

2. Restrictions and Guidelines

- After configuring the IVGL-SVGL mode, you must specify the multicast groups associated with the SVGL mode.
- When a user VLAN is configured as a shared VLAN or sub VLAN, the user VLAN enjoys both shared and independent multicast services. When a user VLAN is configured as other VLANs, it enjoys only independent multicast services.

3. Prerequisites

Before configuring multicast groups associated with the SVGL mode, ensure that the corresponding profile is created and a range of multicast groups that are permitted or denied by the filter is specified. For details about profile configuration, see 1.6.4 Configuring a Profile.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IGMP snooping globally and run the IVGL-SVGL mode.

ip igmp snooping-mode ivgl-svgl

IGMP snooping is disabled by default.

(4) Configure multicast groups associated with the SVGL mode.

ip igmp snooping-mode svgl group-profile *profile-number*

No multicast group is associated with the SVGL mode by default.

(5) (Optional) Specify a shared VLAN.

ip igmp snooping-mode svgl shared-vlan *vlan-id*

The default shared VLAN in SVGL mode is VLAN 1.

(6) (Optional) Specify sub VLANs in SVGL mode.

ip igmp snooping-mode svgl subvlan *vlan-range*

All VLANs except the shared VLAN are sub VLANs in SVGL mode by default.

1.4 Configuring Protocol Packet Processing Parameters

1.4.1 Overview

By controlling protocol packet processing, an L2 multicast device can establish static or dynamic multicast forwarding entries. In addition, the device can adjust parameters to refresh dynamic multicast forwarding entries and IGMP snooping membership quickly.

1.4.2 Restrictions and Guidelines

Configuring Basic IGMP Snooping Functions Related configuration functions take effect only after basic IGMP snooping functions are configured.

1.4.3 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Static Multicast Router Port
- Configuring a Static Member Port
- Configuring Report Packet Suppression
- Configuring Port Fast Leave

- Configuring Dynamic Multicast Router Port Learning
- Configuring the Aging Time for Dynamic Multicast Router Ports
- Configuring the Aging Time for Dynamic Member Ports

1.4.4 Configuring a Static Multicast Router Port

1. Overview

A multicast router port is used to receive uplink multicast data and forward IGMP Report and Leave packets. Static multicast router ports never age and can forward IGMP Report and Leave packets to the uplink IGMP querier stably.

2. Restrictions and Guidelines

- In SVGL mode, if no sub VLAN is configured, the configuration of static multicast router ports is valid only in the shared VLAN. In other VLANs, static multicast router ports can be configured but do not take effect. If sub VLANs are configured, the configuration of static multicast router ports is valid in the shared VLAN and non-sub VLANs. In sub VLANs, static multicast router ports can be configured but do not take effect.
- In SVGL-IVGL mode, if no sub VLAN is configured, the configuration of static multicast router ports is valid in all VLANs. If sub VLANs are configured, the configuration of static multicast router ports is valid in the shared VLAN and non-sub VLANs. In sub VLANs, static multicast router ports can be configured but do not take effect.
- In IVGL mode, the configuration of static multicast router ports is valid in all VLANs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static multicast router port.

ip igmp snooping-mode vlan *vlan-id* **multicast-route interface** *interface-type interface-number*

No static multicast router port is configured by default.

1.4.5 Configuring a Static Member Port

1. Overview

When a member port connecting to a member host is configured as a static member port, the host can receive multicast traffic from the specified multicast group no matter whether the host joins the multicast group. Static member ports never age.

2. Restrictions and Guidelines

- In SVGL mode, if no sub VLAN is configured, the configuration of static member ports is valid only in the shared VLAN. In other VLANs, static member ports can be configured but do not take effect. If sub VLANs are configured, the configuration of static member ports is valid in the shared VLAN and non-sub VLANs. In sub VLANs, static member ports can be configured but do not take effect.
- In SVGL-IVGL mode, if no sub VLAN is configured, the configuration of static member ports is valid in all

VLANs. If sub VLANs are configured, the configuration of static member ports is valid in the shared VLAN and non-sub VLANs. In sub VLANs, static member ports can be configured but do not take effect.

- In IVGL mode, the configuration of static member ports is valid in all VLANs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static member port.

ip igmp snooping-mode vlan *vlan-id* **static-member** *ipv4-group-address* **interface** *interface-type interface-number*

No static member port is configured by default.

1.4.6 Configuring Report Packet Suppression

1. Overview

IGMP Query and Report packets are exchanged periodically to maintain the group membership. When many hosts in a network join the same multicast group, a large number of IGMP packets will be sent to the IGMP multicast device, which wastes network bandwidth and affects the performance of the IGMP multicast device. Report packet suppression can optimize this situation.

When Report packet suppression is configured, the IGMP multicast device only forwards the first Report packet from a specific VLAN for a multicast group to the multicast router port and suppresses subsequent Report packets for the same multicast group during one query interval. This function reduces the number of packets in the network.

2. Restrictions and Guidelines

Only IGMPv1 and IGMPv2 Report packets can be suppressed, and IGMPv3 Report packets cannot be suppressed.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure Report packet suppression.

ip igmp snooping-mode suppress enable

Report packet suppression is disabled by default.

1.4.7 Configuring Port Fast Leave

1. Overview

When the port fast leave function is enabled and a port receives a Leave packet (including the IGMPv2 Leave packet and IGMPv3 Report packet of the INCLUDE type without carrying any source address), the port is directly deleted from the member port list of the corresponding multicast forwarding entry. When receiving group-specific Query packets and multicast data, the device does not forward the packets to this port.

2. Restrictions and Guidelines

The port fast leave function is applicable when only one host is connected to each port. This function helps save bandwidth and resources.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the port fast leave function.

ip igmp snooping-mode inf-fast-leave enable

The port fast leave function is disabled by default.

1.1.1 Configuring Dynamic Multicast Router Port Learning

1. Overview

In some application scenarios, a user host may send Query packets and PIM neighbor messages. When the L2 device receives such Query packets or PIM neighbor messages, the L2 device will set the port receiving the packets as a dynamic multicast router port. All multicast packets in the VLAN will be forwarded to this port, and the host receives a large number of useless multicast packets. In addition, Query packets and PIM neighbor messages sent by user hosts may affect the status of the L3 multicast routing protocol, such as the querier and designated router (DR) election. In this case, you can disable dynamic multicast router port learning to resolve this problem and improve network security.

2. Restrictions and Guidelines

Disabling dynamic multicast router port learning and enabling a static multicast router port do not affect each other.

3. Procedure

(4) Enter the privileged EXEC mode.

privilege

(5) Enter the global configuration mode.

config-system terminal

(6) Configure dynamic multicast router port learning.

ip igmp snooping-mode [**vlan** *vlan-id*] **multicast-route learn dvmrp-pim**

Dynamic multicast router port learning is enabled by default.

1.4.8 Configuring the Aging Time for Dynamic Multicast Router Ports

1. Overview

A multicast router port is used to receive uplink multicast data and forward IGMP Report and Leave packets. After enabling IGMP snooping, an L2 device can dynamically learn multicast router ports to forward multicast packets for the uplink device. If a dynamic multicast router port does not receive a Query packet or neighbor message from the uplink device within the aging time due to unstable network or packet congestion, the L2 device will delete the dynamic multicast router port, which may cause multicast data interruption.

2. Restrictions and Guidelines

- You can adjust the aging time of dynamic multicast router ports based on the network load. When the network load is heavy, set the aging time to a larger value. A too short aging time may cause dynamic multicast router ports to be added and deleted frequently, resulting in data interruption.
- You can set the aging time based on the interval for the connected multicast router to send IGMP Query packets. Generally, the aging time is set to twice the interval plus the response time of the last IGMP Query packet.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the aging time for dynamic multicast router ports.

ip igmp snooping-mode dyn-mr-aging-interval *dynamic-mroute-aging-time*

The default aging time of dynamic multicast router ports is 300s.

1.4.9 Configuring the Aging Time for Dynamic Member Ports

1. Overview

When the device running IGMP snooping receives an IGMP Join packet from a host to join an IP multicast group, the device adds the port receiving the packet to the member port list and sets an aging time for the port. If the port is already in the member port list, the device resets the aging timer of the port. After the timer times out, it is deemed that no user host receives multicast packets through this port, and the multicast device deletes the port from the IGMP snooping member port list.

2. Restrictions and Guidelines

- You can adjust the aging time of dynamic member ports based on the network load. When the network load is heavy, set the aging time to a larger value. A too short aging time may cause dynamic member ports to be added and deleted frequently, resulting in data interruption.
- After the aging time is configured, the aging timer value of dynamic member ports is *host-aging-interval* for subsequent IGMP Join packets. The configured aging time takes effect after the next Join packet is received, and the started member port aging timers are not updated.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the aging time for dynamic member ports.

ip igmp snooping-mode host-aging-interval *host-aging-time*

The default aging time of dynamic member ports is 260s.

1.5 Configuring IGMP Snooping Querier

1.5.1 Overview

In a three-layer multicast network, the L3 multicast device serves as the querier and runs IGMP to maintain group membership. The L2 multicast device only needs to listen to IGMP packets to establish and maintain multicast forwarding entries, implementing L2 multicast. When a multicast source and user host are in the same L2 network, the query function is unavailable because the L2 device does not support IGMP. To resolve this problem, you can enable the IGMP snooping querier function on the L2 device so that the L2 device sends IGMP Query packets to user hosts on behalf of the L3 multicast device, and listens to and maintains IGMP Report packets responded by user hosts to establish L2 multicast forwarding entries. When multiple devices are configured as the querier in the L2 network and a device receives IGMP Query packets from other devices, an IGMP snooping querier is elected.

1.5.2 Restrictions and Guidelines

- After the querier function is enabled, you must specify a source IP address for the querier so that the querier function can take effect.
- When a device with the querier function enabled receives a PIM or Distance Vector Multicast Routing Protocol (DVMRP) packet, it regards that a multicast router exists in the network. To prevent the IGMP routing function of the multicast router from being affected, the querier function of the device becomes invalid and the multicast router in the network serves as the IGMP snooping querier.
- The query interval configured for the querier must be greater than the maximum response time of Query packets. Otherwise, the configuration fails.

1.5.3 Prerequisites

Configuring Basic IGMP Snooping Functions

1.5.4 Configuration Tasks

IGMP snooping querier configuration includes the following tasks:

- Enabling IGMP Snooping Querier
- Configuring the Source IP Address of a Querier
- (Optional) Configuring the IGMP Snooping Querier Version
- (Optional) Configuring the Maximum Response Time for Query Packets

- (Optional) Configuring the Interval for a Querier to Send Query Packets
- (Optional) Configuring the Querier Aging Time

1.5.5 Enabling IGMP Snooping Querier

1. Overview

After the IGMP snooping querier function is enabled on an L2 device, the L2 device sends IGMP Query packets to user hosts on behalf of the L3 device.

2. Restrictions and Guidelines

- The IGMP snooping querier function must be enabled globally and on a specific VLAN. The IGMP snooping querier function on a specific VLAN takes effect only after the function is enabled globally.
- You can disable the IGMP snooping querier function for a specific VLAN.
- If the IGMP snooping querier function is disabled globally, the function is disabled on all VLANs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IGMP snooping querier function.

ip igmp snooping-mode [**vlan** *vlan-id*] **querier**

The IGMP snooping querier function is disabled by default.

1.5.6 Configuring the Source IP Address of a Querier

1. Overview

You can configure the source IP address of a querier globally or on a specific VLAN. If the source IP address of the querier is not configured, the querier function does not take effect.

2. Restrictions and Guidelines

The querier source IP address configured on a specific VLAN, if any, prevails.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the source IP address of a querier.

ip igmp snooping-mode [**vlan** *vlan-id*] **querier source-address** *ipv4-source-address*

By default, no VLAN is specified, and the source IP address configuration applies to all VLANs.

1.5.7 Configuring the IGMP Snooping Querier Version

1. Overview

You can configure the IGMP snooping querier version to specify the version of IGMP Query packets sent by the querier. The supported IGMP versions include IGMPv1, and IGMPv2.

2. Restrictions and Guidelines

- An IGMP snooping querier supports IGMPv1, and IGMPv2, IGMPv2 is used by default. You can run the corresponding command to enable an IGMP snooping querier to run IGMPv1.
- The IGMP snooping querier version configured on a specific VLAN, if any, prevails.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the IGMP snooping querier version.

ip igmp snooping-mode [**vlan** *vlan-id*] **querier version 1**

An IGMP snooping querier runs IGMPv2 by default.

1.5.8 Configuring the Maximum Response Time for Query Packets

1. Overview

When hosts in the directly-connected network segment of the device serving as the IGMP snooping querier receive a Query packet from the device, the hosts need to return a Report packet within the maximum response time. This function allows you to configure the maximum response time for Query packets on the device. If no host returns a Report packet within the maximum response time, the device considers that no group member exists in the directly-connected network segment and deletes the group information.

2. Restrictions and Guidelines

- You can adjust the maximum response time for Query packets based on the network load. When the network load is light, you are advised to set the maximum response time for Query packets to a small value to enable hosts to make quick responses to IGMP Query packets. When the network load is heavy, set a large maximum response time to avoid network congestion due to flooding of Report packets sent by hosts after the timer times out.
- Because IGMPv1 does not support carrying the maximum response time in packets, this configuration does not take effect to a querier running IGMPv1.
- You can configure different maximum response time values for queriers in different VLANs. The maximum response time configured on a specific VLAN, if any, prevails.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum response time for Query packets.

ip igmp snooping-mode [**vlan** *vlan-id*] **querier max-response-time** *max-response-time*

The maximum response time for Query packets is 10s by default.

1.5.9 Configuring the Interval for a Querier to Send Query Packets

1. Overview

An IGMP snooping querier periodically sends Query packets. This parameter is used to define the interval for sending Query packets.

2. Restrictions and Guidelines

You can configure the interval for a querier to send Query packets globally or on a specific VLAN. The query interval specified on a VLAN, if any, prevails.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the interval for a querier to send Query packets.

ip igmp snooping-mode [**vlan** *vlan-id*] **querier query-period** *query-interval*

The default interval for an IGMP snooping querier to send Query packets is 60s.

1.5.10 Configuring the Querier Aging Time

1. Overview

When multiple queriers are configured in an L2 network, querier election will be performed, and the device elected as the querier will send Query packets periodically. Other candidate queriers receive Query packets from the elected querier. If a candidate querier does not receive Query packets from the elected querier within a specific period of time, the candidate querier regards that it is the only querier in the directly-connected network segment and initiates a new round querier election. This parameter is used to define the timeout time for candidate queriers to receive Query packets from the elected querier.

2. Restrictions and Guidelines

The querier aging time configured on a specific VLAN, if any, prevails.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the querier aging time.

ip igmp snooping-mode [**vlan** *vlan-id*] **querier other-querier-timer present** *timeout*

The default aging time of IGMP snooping queriers is 125s.

1.6 Configuring Multicast Security Control

1.6.1 Overview

After multicast security control is configured, the device running IGMP snooping can control the multicast service scope and load to prevent invalid multicast traffic and improve L2 multicast network security.

1.6.2 Prerequisites

Configuring Basic IGMP Snooping Functions

1.6.3 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Profile
- Configuring a Range of Multicast Group Addresses for a ProfileConfiguring Multicast Preview
- Configuring the Maximum Number of Multicast Groups Allowed for Concurrent Request
- Configuring Source Port Check
- Configuring Source IP Address Check

1.6.4 Configuring a Profile

1. Overview

A profile is used to define a range of multicast groups that permit or deny user access for reference by other functions.

- When the SVGL mode is enabled, a profile is used to define a range of multicast groups applied in SVGL mode.
- When multicast filtering is configured on a port, a profile is used to define a range of multicast groups that permit or deny user access through the port.
- When multicast filtering is configured on a VLAN, a profile is used to define a range of multicast groups that permit or deny user access on the VLAN.
- When the preview function is enabled, a profile is used to define a range of multicast groups to be previewed.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a profile.

ip igmp group-profile *profile-number*

No profile is created by default.

(4) Define a range of multicast groups for a profile.

range *start-ipv4-address* [*end-ipv4-address*]

No multicast group range is defined for a profile by default.

Multiple multicast group ranges can be configured at the same time. If two multicast group ranges overlap, the ranges will be merged automatically.

(5) Configure the filtering action for a profile. The configuration steps below are mutually exclusive. Please configure only one task.

- Set the filtering action of a profile to Deny.

  deny

  If only the Deny action is configured and no multicast group range is configured, no group is denied. The effect is the same as permitting all groups.

- (Optional) Set the filtering action of a profile to Permit.

  permit

  If only the Permit action is configured and no multicast group range is configured, no group is permitted. The effect is the same as denying all groups.

The Deny action is performed for a profile by default.

1.6.5 Configuring a Range of Multicast Group Addresses for a Profile

1. Overview

Generally, the device running ports and VLANs can join any multicast group. By configuring a range of multicast groups that permit or deny user access, you can customize the multicast service scope for users to guarantee the interest of operators and prevent invalid multicast traffic.

2. Restrictions and Guidelines

- By applying a profile to a VLAN, you can restrict the multicast groups that users can join in the VLAN.
- By applying a profile to a port, you can restrict the multicast groups that users can join on the port.

3. Prerequisites

Before configuring a range of multicast group addresses for a profile, ensure that the corresponding profile is created and a range of multicast groups is specified for the profile. For details about profile configuration, see 1.6.4 Configuring a Profile.

4. Procedure

Configure a range of multicast groups for a profile. Select at least one of them to configure.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

(1) (Optional) Configure the limit of VLAN group address range.

ip igmp snooping-mode vlan *vlan-id* **group-filter** *profile-number*

By default, a VLAN is allowed to join any group.

(2) (Optional) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

(3) (Optional) Configure the limit of interface group address range.

ip igmp snooping-mode group-filter *profile-number*

By default, an interface is allowed to join any group.

1.6.6 Configuring Multicast Preview

1. Overview

In video applications, the administrator configures a VLAN filter or port filter to control some premium channels. Users who do not subscribe to these channels cannot watch these channels on demand. The multicast preview function allows unpaid users to preview premium channels before they decide whether to pay for watching. This function can be used together with multicast permission control.

2. Restrictions and Guidelines

After the preview duration ends, unpaid users can no longer watch premium channels on demand. The previewed content can be previewed again after 300s.

3. Prerequisites

Before configuring multicast preview, ensure that the corresponding profile is created and a range of multicast groups allowed for preview is specified. For details about profile configuration, see 1.6.4 Configuring a Profile.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure multicast preview.

ip igmp snooping-mode preview *profile-number*

The multicast preview function is disabled by default.

(4) (Optional) Configure the preview duration.

ip igmp snooping-mode preview interval preview-*interval*

The default multicast preview duration is 60s.

1.6.7 Configuring the Maximum Number of Multicast Groups Allowed for Concurrent Request

1. Overview

If too much multicast traffic is requested concurrently, the multicast device will be severely burdened. Configuring the maximum number of multicast groups allowed for concurrent request can guarantee the bandwidth. You can limit the number of multicast groups allowed for concurrent request globally or on a port.

- Global maximum number of multicast groups allowed for concurrent request: This configuration contains statically configured and dynamically learned multicast groups.
- Maximum number of multicast groups allowed for concurrent request on a port: This configuration contains only dynamically learned multicast groups on a port and does not include statically configured multicast groups.

2. Restrictions and Guidelines

- To limit the number of multicast groups allowed for concurrent request globally, configure the global maximum number of multicast groups allowed for concurrent request.
- To limit the number of multicast groups allowed for concurrent request on a port, configure the maximum number of multicast groups that can be dynamically learned by the port.
- The number of multicast groups that can be dynamically learned by a port is counted based on the VLANs to which the port belongs. For example, if a port belongs to three VLANs and the port receives requests of multicast group 224.1.1.1 from each VLAN, the multicast group count on the port is 3 instead of 1.

3. Procedure

Configure the maximum number of multicast groups allowed for concurrent request. Select at least one of them to configure.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c (Optional) Configure the maximum number of multicast groups allowed for concurrent request globally.

ip igmp snooping-mode l2-entry-limit *l2-entry-limit*

The maximum number of multicast groups allowed for concurrent request globally is 64,000 by default.

d (Optional) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 2 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*

e Configure the maximum number of multicast groups that can be dynamically learned by a port.

ip igmp snooping-mode max-member-record *max-member-record*

The maximum number of multicast groups that can be dynamically learned by a port is 64,000 by default.

1.6.8 Configuring Source Port Check

1. Overview

The source port check function restricts users to receive only multicast traffic on multicast router ports to prevent users from sending invalid multicast traffic.

- After the source port check function is enabled, only multicast traffic received on multicast router ports is valid. Multicast traffic received on other ports is invalid and will be discarded.
- When the source port check function is disabled, multicast traffic received on any port is valid.

2. Restrictions and Guidelines

- To restrict users to receive multicast traffic only on multicast router ports, configure this function.
- If no multicast router port exists after the source port check function is enabled, the device will discard received multicast traffic.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure source port check.

ip igmp snooping-mode source-inf port

Source port check is disabled by default.

1.6.9 Configuring Source IP Address Check

1. Overview

The source IP address check function strictly restricts the source IP address of multicast traffic to prevent invalid multicast traffic.

Source IP address check contains the following two configurations:

- Default source IP address check: A source IP address is specified for each multicast group in all VLANs. Only multicast traffic with the source IP address the same as the preset one is valid.
- Group-specific source IP address check: A source IP address is specified for a specific multicast group in a specific VLAN. The multicast device forwards only data traffic of the specific multicast group received in the VLAN that has the same source IP address as the preset one and discards other multicast traffic from the specific multicast group.

2. Restrictions and Guidelines

- When source IP address check is enabled or disabled, all multicast forwarding entries on the L2 device will be deleted and the device needs to learn entries in the next query interval.
- Group-specific source IP address check can be configured only after default source IP address check is configured.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure default source IP address check.

ip igmp snooping-mode source-inf default-server *ipv4-source-address*

Source IP address check is disabled by default.

(4) Configure group-specific source IP address check.

ip igmp snooping-mode limit-ipmc vlan *vlan-id* **address** *ipv4-group-address* **server** *ipv4-source-address*

Only the multicast source address is checked by default.

1.7 Configuring QinQ Processing

Note

This command is not supported by the S6980-64QC, or S6580-24DC8QC.

1.7.1 Overview

On a device with IGMP snooping enabled and a dot1q-tunnel (QinQ) port configured, IGMP snooping will process the IGMP packets received on the QinQ port using the following two modes:

- Mode 1: Create multicast entries on the VLAN where the IGMP packets are located. Forwarding IGMP packets on the VLAN where these packets are located is called transparent transmission. For example, IGMP snooping is enabled for a device, Port A on the device is designated as the QinQ port, the default VLAN of this port is VLAN 1, and Port A allows the passage of VLAN 1 and VLAN 10 packets. When an IGMP Report packet is sent from VLAN 10 to Port A, IGMP snooping establishes a multicast entry for VLAN 10 and forwards the IGMP Report packet to the multicast router port of VLAN 10.
- Mode 2: Create multicast entries on the default VLAN of the QinQ port. Encapsulate the IGMP packets with the VLAN tag of the default VLAN where the QinQ port is located and forward the packets within the default VLAN. For example, IGMP snooping is enabled for a device, Port A on the device is designated as the QinQ port, the default VLAN of this port is VLAN 1, and Port A allows the passage of VLAN 1 and VLAN 10 packets. When an IGMP Report packet is sent from VLAN 10 to Port A, IGMP snooping establishes a multicast entry for VLAN 1, encapsulates the IGMP Report packet with the tag of VLAN 1, and forwards the packet to the multicast router port of VLAN 1.

IGMP snooping works in mode 2 by default. You can run the ip igmp snooping-mode tunnel command to enable IGMP snooping to work in mode 1.

1.7.2 Restrictions and Guidelines

If the QinQ port needs to forward multicast packets on the VLANs specified by the VLAN IDs (VIDs) of the packets, transparent packet transmission needs to be configured on the QinQ port.

1.7.3 Prerequisites

- Configuring Basic IGMP Snooping Functions
- Configure basic QinQ functions. For details about QinQ configuration, see *Configuring QinQ*.

1.7.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure transparent packet transmission on the QinQ port.

ip igmp snooping-mode tunnel

The function of transparent IGMP packet transmission on the QinQ port is disabled by default.

1.8 Monitoring

Run the display command to check the running status of a configured function to verify the configuration effect.

Run the debugging command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the cleanup commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-2 Monitoring

Command	Purpose
cleanup ip igmp-cache snooping-mode stats	Clears IGMP snooping statistics.
cleanup ip igmp-cache snooping-mode gda-table	Clears dynamic multicast router ports and member ports.
display ip igmp-info snooping [**vlan** *vlan-id*]	Displays basic IGMP snooping configurations.
display ip igmp-info snooping statistics [**vlan** *vlan-id*]	Displays IGMP snooping statistics.
display ip igmp-info snooping multicast-router	Displays multicast router ports.

Command	Purpose
display ip igmp-info snooping gda-table	Displays IGMP snooping forwarding entries.
display ip igmp-info group-profile [*profile-number*]	Displays a profile.
display ip igmp-info snooping interface [*interface-type interface-number*]	Displays IGMP snooping configurations on a port.
display ip igmp-info snooping querier [**detail**]	Displays the IGMP snooping querier.
display ip igmp-info snooping thread-info	Displays detailed user information of the IGMP snooping process.
debugging igmp-snp	Debugs all IGMP snooping functions.
debugging igmp-snp event	Debugs IGMP snooping events.
debugging igmp-snp packet	Debugs IGMP snooping packets.
debugging igmp-snp msf	Debugs communication between IGMP snooping and Multicast Source Filter (MSF).
debugging igmp-snp warning	Debugs IGMP snooping warnings.

1.9 Configuration Examples

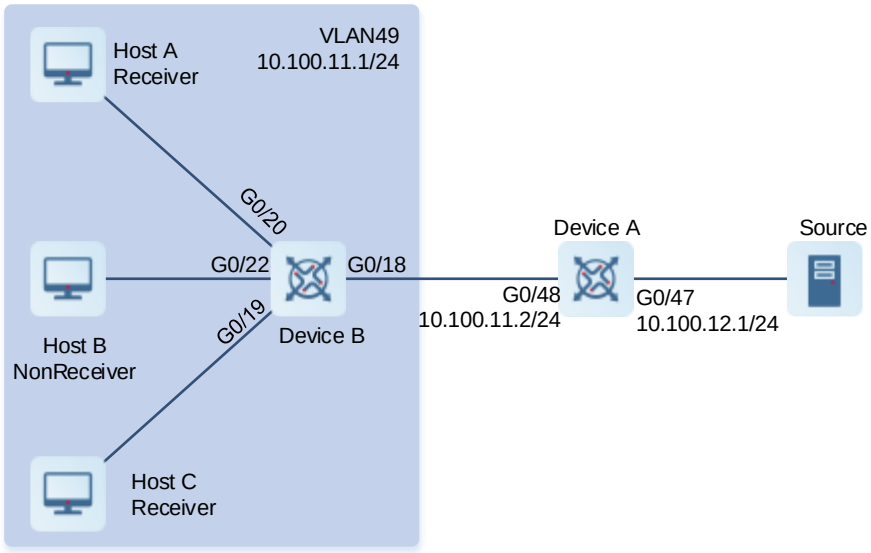
1.9.1 Configuring Basic IGMP Snooping Functions in IVGL Mode

1. Requirements

In the multicast network shown in Figure 1-5, the router (Device A) connects to the user network through the switch (Device B), and IGMP snooping is run on Device A. The multicast source sends data to multicast group 225.0.0.10. There are three hosts (Host A, Host B, and Host C) in the network. Host A and Host C join multicast group 225.0.0.10. Only Host A and Host C can receive multicast data that the multicast source sends to multicast group 225.0.0.10.

2. Topology

Figure 1-5 Topology of Basic IGMP Snooping Functions in IVGL Mode



3. Notes

Configure basic IGMP snooping functions on Device B.

- Configure the IP addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.
- Enable multicast routing and Protocol Independent Multicast-Sparse Mode (PIM-SM) related functions on Device A.
- Enable IGMP snooping and run the IVGL mode on Device B.

4. Procedure

(1) Configure the IP addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.

Configure the IP address and unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/47)# ip address 10.100.12.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/48)# ip address 10.100.11.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
```

```
DeviceA(cfg-router)# exit
```

Configure the IP address, VLAN, and unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan 49
DeviceB(cfg-vlan)# exit
DeviceB(cfg)# interface vlan 49
DeviceB(cfg-if-VLAN 49)#ip address 10.100.11.1 255.255.255.0
DeviceB(cfg-if-VLAN 49)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/19)# exit
DeviceB(cfg)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/20)# exit
DeviceB(cfg)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/22)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/22)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceB(cfg-router)# exit
```

(2) Enable multicast routing and PIM-SM related functions on Device A.

```
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# ip pim c-bsr GigabitEthernet 0/47
DeviceA(cfg)# ip pim c-rp GigabitEthernet 0/47
```

(3) Enable IGMP snooping and run the IVGL mode on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip igmp snooping-mode ivgl
```

5. Verification

Send packets from the multicast source (10.100.12.2) to multicast group G (225.0.0.10). Enable Host A and Host C to join G.

Run the **display ip igmp-info snooping gda-table** command on Device B to display the IGMP snooping forwarding entry and check whether the member port list contains only GigabitEthernet 0/19 and GigabitEthernet 0/20.

```
DeviceB# display ip igmp-info snooping gda-table
Multicast Switching Cache Table
   D: DYNAMIC
   S: STATIC
   M: MROUTE
(*, 225.0.0.10, 49):
   VLAN(49) 3 OPORTS:
      GigabitEthernet 0/18(M)
      GigabitEthernet 0/19(DM)
      GigabitEthernet 0/20(DM)
```

Run the **display ip igmp-info snooping** command on Device B and check whether the IGMP snooping working mode is IVGL.

```
DeviceB# display ip igmp-info snooping
IGMP Snooping running mode: IVGL
IGMP Snooping L2-entry-limit: 64000
Source port check: Disable
Source ip check: Disable
IGMP Fast-Leave: Disable
IGMP Report suppress: Disable
IGMP Global Querier: Disable
IGMP Preview: Disable
IGMP Tunnel: Disable
IGMP Snooping version: 2
IGMP Preview group aging time: 60(Seconds)
Dynamic Mroute Aging Time: 300(Seconds)
Dynamic Host Aging Time: 260(Seconds)

vlan 49
----------
IGMP Snooping state: Enable
Multicast router learning mode: pim-dvmrp
IGMP Fast-Leave: Enabled
IGMP VLAN querier: Disable
IGMP VLAN Mode: STATIC
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
interface GigabitEthernet 0/47
 unset switch-port
 ip address 10.100.12.1 255.255.255.0
```

```
ip pim sparse-mode
!
interface GigabitEthernet 0/48
 unset switch-port
 ip address 10.100.11.2 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 10.100.12.0 0.0.0.255 area 0
!
ip pim c-bsr GigabitEthernet 0/48
ip pim c-rp GigabitEthernet 0/48
!
```

- Device B configuration file

```
hostname DeviceB
!
vlan range 1,49
!
interface GigabitEthernet 0/18
 switch-port access vlan 49
!
interface GigabitEthernet 0/19
 switch-port access vlan 49
!
interface GigabitEthernet 0/20
 switch-port access vlan 49
!
interface GigabitEthernet 0/22
 switch-port access vlan 49
!
interface VLAN 49
 ip address 10.100.11.1 255.255.255.0
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
!
ip igmp snooping-mode ivgl
!
```

7. Common Errors

- The working mode of IGMP snooping is improper.

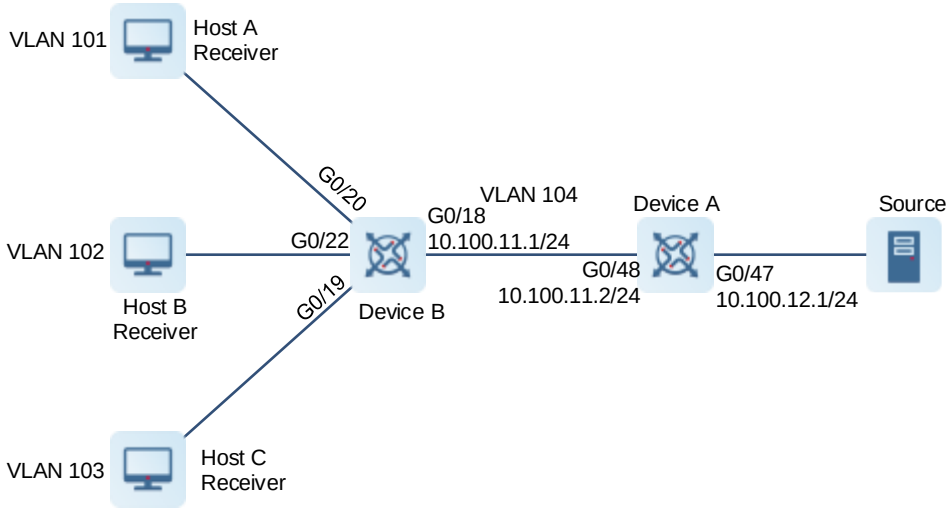
1.9.2 Configuring Basic IGMP Snooping Functions in SVGL Mode

1. Requirements

In the multicast network shown in Figure 1-6, the router (Device A) connects to the user network through the switch (Device B), IGMP snooping is run on Device A, and Device A directly connects to the multicast source. The multicast source sends data to multicast group 225.0.0.10. Host A, Host B, and Host C are connected to VLAN 101, VLAN 102, and VLAN 103, respectively, and the three hosts join multicast group 225.0.0.10. Host A, Host B, and Host C can receive multicast data that the multicast source sends to multicast group 225.0.0.10.

2. Topology

Figure 1-6 Topology of Basic IGMP Snooping Functions in SVGL Mode



3. Notes

Configure basic IGMP snooping functions on Device B.

- Configure the IP addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.
- Enable multicast routing and PIM-SM related functions on Device A.
- Enable IGMP snooping and run the SVGL mode on Device B.
- Configure the multicast groups associated with the IGMP snooping SVGL mode on Device B and specify the shared VLAN and sub VLANs.

4. Procedure

(1) Configure the IP addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.

Configure the IP address and unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
```

```
DeviceA(cfg-if-GigabitEthernet 0/47)# ip address 10.100.12.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# vlan 104
DeviceA(cfg)# interface vlan 104
DeviceA(cfg-if-VLAN 104)# ip address 10.100.11.2 255.255.255.0
DeviceA(cfg-if-VLAN 104)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet 0/48)# switch-port trunk native vlan 104
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
DeviceA(cfg-router)# exit
```

Configure the IP address, VLAN, and unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan range 101-104
DeviceB(cfg-vlan-range)# exit
DeviceB(cfg)# interface VLAN 101
DeviceB(cfg-if-VLAN 101)# ip address 10.100.101.1 255.255.255.0
DeviceB(cfg-if-VLAN 101)# exit
DeviceB(cfg)# interface VLAN 102
DeviceB(cfg-if-VLAN 102)# ip address 10.100.102.1 255.255.255.0
DeviceB(cfg-if-VLAN 102)# exit
DeviceB(cfg)# interface VLAN 103
DeviceB(cfg-if-VLAN 103)# ip address 10.100.103.1 255.255.255.0
DeviceB(cfg-if-VLAN 103)# exit
DeviceB(cfg)# interface VLAN 104
DeviceB(cfg-if-VLAN 104)# ip address 10.100.11.1 255.255.255.0
DeviceB(cfg-if-VLAN 104)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port trunk native vlan 104
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 103
DeviceB(cfg-if-GigabitEthernet 0/19)# exit
DeviceB(cfg)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 101
DeviceB(cfg-if-GigabitEthernet 0/20)# exit
DeviceB(cfg)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/22)# switch-port access vlan 102
DeviceB(cfg-if-GigabitEthernet 0/22)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
```

```
DeviceB(cfg-router)# network-advertise 10.100.101.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.102.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.103.0 0.0.0.255 area 0
DeviceB(cfg-router)# exit
```

(2) Enable multicast routing and PIM-SM related functions on Device A.

```
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface vlan 104
DeviceA(cfg-if-VLAN 104)# ip pim sparse-mode
DeviceA(cfg-if-VLAN 104)# exit
DeviceA(cfg)# ip pim c-bsr GigabitEthernet 0/47
DeviceA(cfg)# ip pim c-rp GigabitEthernet 0/47
```

(3) Enable IGMP snooping and run the IVGL mode on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip igmp snooping-mode svgl
```

(4) Configure the multicast groups associated with the IGMP snooping SVGL mode on Device B and specify the shared VLAN and sub VLANs.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip igmp group-profile 1
DeviceB(cfg-profile)# permit
DeviceB(cfg-profile)# range 224.1.1.1 238.1.1.1
DeviceB(cfg-profile)# exit
DeviceB(cfg)# ip igmp snooping-mode svgl group-profile 1
DeviceB(cfg)# ip igmp snooping-mode svgl shared-vlan 104
DeviceB(cfg)# ip igmp snooping-mode svgl subvlan 101-103
```

5. Verification

Send packets from the multicast source (10.100.12.2) to multicast group G (225.0.0.10). Enable Host A, Host B, and Host C to join G.

Run the **display ip igmp-info snooping gda-table** command on Device B to display the IGMP snooping forwarding entry and check whether the member port list contains GigabitEthernet 0/19, GigabitEthernet 0/20, and GigabitEthernet 0/22.

```
DeviceB# display ip igmp-info snooping gda-table
Multicast Switching Cache Table
  D: DYNAMIC
  S: STATIC
  M: MROUTE
(*, 225.0.0.10, 104):
  VLAN(104) 1 OPORTS:
    GigabitEthernet 0/18(M)
```

```
VLAN(102) 1 OPORTS:
   GigabitEthernet 0/22(D)

VLAN(103) 1 OPORTS:
   GigabitEthernet 0/19(D)

VLAN(101) 1 OPORTS:
   GigabitEthernet 0/20(D)
```

Run the **display ip igmp-info snooping** command on Device B and check whether the IGMP snooping working mode is SVGL.

```
DeviceB# display ip igmp-info snooping
IGMP Snooping running mode: SVGL
IGMP Snooping L2-entry-limit: 64000
SVGL vlan: 104
SVGL profile number: 1
IGMP Snooping SVGL subvlan: 101-103
Source port check: Disable
Source ip check: Disable
IGMP Fast-Leave: Disable
IGMP Report suppress: Disable
IGMP Global Querier: Disable
IGMP Preview: Disable
IGMP Tunnel: Disable
IGMP Snooping version: 2
IGMP Preview group aging time: 60(Seconds)
Dynamic Mroute Aging Time: 300(Seconds)
Dynamic Host Aging Time: 260(Seconds)
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
vlan 104
!
interface GigabitEthernet 0/47
 unset switch-port
 ip address 10.100.12.1 255.255.255.0
ip pim sparse-mode
!
interface GigabitEthernet 0/48
 switch-port mode trunk
```

```
 switch-port trunk native vlan 104
!
interface VLAN 104
 ip address 10.100.11.2 255.255.255.0
ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 10.100.12.0 0.0.0.255 area 0
!
ip pim c-bsr GigabitEthernet 0/48
ip pim c-rp GigabitEthernet 0/48
!
```

- Device B configuration file

```
hostname DeviceB
!
ip igmp group-profile 1
 permit
 range 224.1.1.1 238.1.1.1
!
vlan range 1,101-104
!
interface GigabitEthernet 0/18
 switch-port mode trunk
 switch-port trunk native vlan 104
!
interface GigabitEthernet 0/19
 switch-port access vlan 103
!
interface GigabitEthernet 0/20
 switch-port access vlan 101
!
interface GigabitEthernet 0/22
 switch-port access vlan 102
!
interface VLAN 101
 ip address 10.100.101.1 255.255.255.0
!
interface VLAN 102
 ip address 10.100.102.1 255.255.255.0
!
interface VLAN 103
 ip address 10.100.103.1 255.255.255.0
!
interface VLAN 104
```

```
 ip address 10.100.11.1 255.255.255.0
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 10.100.101.0 0.0.0.255 area 0
 network-advertise 10.100.102.0 0.0.0.255 area 0
 network-advertise 10.100.103.0 0.0.0.255 area 0
!
ip igmp snooping-mode svgl shared-vlan 104
ip igmp snooping-mode svgl group-profile 1
ip igmp snooping-mode svgl subvlan 101-103
ip igmp snooping-mode svgl
!
```

7. Common Errors

- The multicast groups associated with the SVGL mode are not configured.
- The sent multicast traffic is not in the multicast groups associated with the SVGL mode.

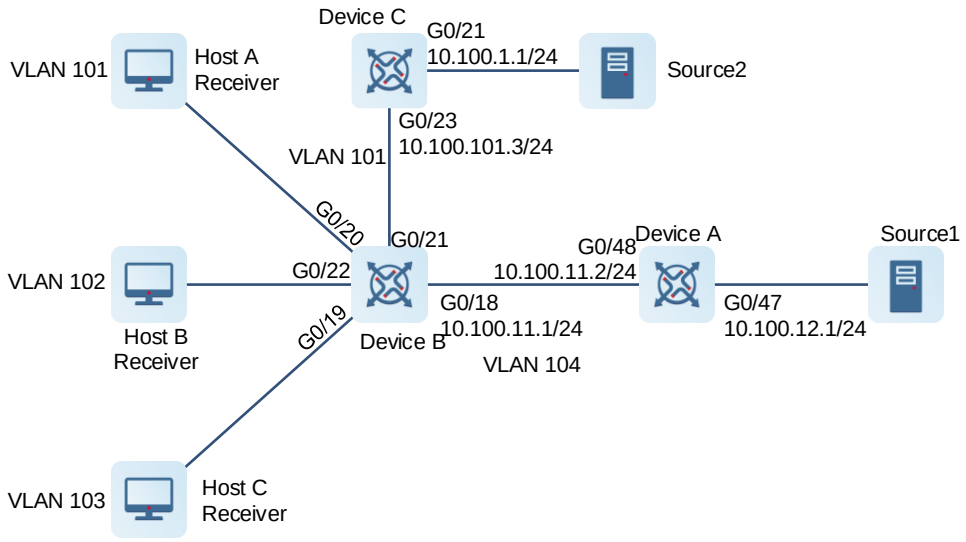
1.9.3 Configuring Basic IGMP Snooping Functions in IVGL-SVGL Mode

1. Requirements

In the multicast network shown in Figure 1-7, the routers (Device A and Device C) connect to the user network through the switch (Device B), IGMP snooping is run on Device A and Device C, and Device A and Device C directly connect to the multicast sources. Multicast source 1 sends data to multicast group 225.0.0.10, and multicast source 2 sends data to multicast group 239.0.0.11. Host A, Host B, and Host C are connected to VLAN 101, VLAN 102, and VLAN 103, respectively. Host B and Host C join multicast group 225.0.0.10, and Host A joins multicast groups 225.0.0.10 and 239.0.0.11. Host B and Host C can receive multicast data that source 1 sends to multicast group 225.0.0.10, and Host A can receive multicast data that source 1 sends to multicast group 225.0.0.10 and multicast data that source 2 sends to multicast group 239.0.0.11. Host A, Host B, and Host C enjoy shared multicast services, and Host A also enjoys independent multicast services.

2. Topology

Figure 1-7 Topology of Basic IGMP Snooping Functions in IVGL-SVGL Mode



3. Notes

Configure basic IGMP snooping functions in IVGL-SVGL mode on Device B.

- Configure the IP addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.
- Enable multicast routing and PIM-SM related functions on Device A and Device C.
- Enable IGMP snooping and run the IVGL-SVGL mode on Device B.
- Configure the multicast groups associated with the IGMP snooping SVGL mode on Device B and specify the shared VLAN and sub VLANs.

4. Procedure

(1) Configure the IP addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.

Configure the IP address and unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/47)# ip address 10.100.12.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# vlan 104
DeviceA(cfg)# interface vlan 104
DeviceA(cfg-if-VLAN 104)# ip address 10.100.11.2 255.255.255.0
DeviceA(cfg-if-VLAN 104)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# switch-port mode trunk
```

```
DeviceA(cfg-if-GigabitEthernet 0/48)# switch-port trunk native vlan 104
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
DeviceA(cfg-router)# exit
```

Configure the IP address, VLAN, and unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan range 101-104
DeviceB(cfg-vlan-range)# exit
DeviceB(cfg)# interface VLAN 101
DeviceB(cfg-if-VLAN 101)# ip address 10.100.101.1 255.255.255.0
DeviceB(cfg-if-VLAN 101)# exit
DeviceB(cfg)# interface VLAN 102
DeviceB(cfg-if-VLAN 102)# ip address 10.100.102.1 255.255.255.0
DeviceB(cfg-if-VLAN 102)# exit
DeviceB(cfg)# interface VLAN 103
DeviceB(cfg-if-VLAN 103)# ip address 10.100.103.1 255.255.255.0
DeviceB(cfg-if-VLAN 103)# exit
DeviceB(cfg)# interface VLAN 104
DeviceB(cfg-if-VLAN 104)# ip address 10.100.11.1 255.255.255.0
DeviceB(cfg-if-VLAN 104)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port trunk native vlan 104
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 103
DeviceB(cfg-if-GigabitEthernet 0/19)# exit
DeviceB(cfg)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 101
DeviceB(cfg-if-GigabitEthernet 0/20)# exit
DeviceB(cfg)# interface GigabitEthernet 0/21
DeviceB(cfg-if-GigabitEthernet 0/21)# switch-port access vlan 101
DeviceB(cfg-if-GigabitEthernet 0/21)# exit
DeviceB(cfg)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/22)# switch-port access vlan 102
DeviceB(cfg-if-GigabitEthernet 0/22)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.101.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.102.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 10.100.103.0 0.0.0.255 area 0
DeviceB(cfg-router)# exit
```

Configure the IP address and unicast routing protocol on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ip address 10.100.1.1 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface GigabitEthernet 0/23
DeviceC(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/23)# ip address 10.100.101.3 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/23)# exit
DeviceC(cfg)# router-protocol ospf 49
DeviceC(cfg-router)# network-advertise 10.100.1.0 0.0.0.255 area 0
DeviceC(cfg-router)# network-advertise 10.100.101.0 0.0.0.255 area 0
DeviceC(cfg-router)# exit
```

(2) Enable multicast routing and PIM-SM related functions on Device A and Device C.

Enable multicast routing and PIM-SM related functions on Device A.

```
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface vlan 104
DeviceA(cfg-if-VLAN 104)# ip pim sparse-mode
DeviceA(cfg-if-VLAN 104)# exit
DeviceA(cfg)# ip pim c-bsr GigabitEthernet 0/47
DeviceA(cfg)# ip pim c-rp GigabitEthernet 0/47
```

Enable multicast routing and PIM-SM related functions on Device C.

```
DeviceC(cfg)# ip mrouting-enable
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface GigabitEthernet 0/23
DeviceC(cfg-if-GigabitEthernet 0/23)# ip pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/23)# exit
```

(3) Enable IGMP snooping and run the IVGL mode on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip igmp snooping-mode ivgl-svgl
```

(4) Configure the multicast groups associated with the IGMP snooping SVGL mode on Device B and specify the shared VLAN and sub VLANs.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip igmp group-profile 1
DeviceB(cfg-profile)# permit
```

```
DeviceB(cfg-profile)# range 224.1.1.1 238.1.1.1
DeviceB(cfg-profile)# exit
DeviceB(cfg)# ip igmp snooping-mode svgl group-profile 1
DeviceB(cfg)# ip igmp snooping-mode svgl shared-vlan 104
DeviceB(cfg)# ip igmp snooping-mode svgl subvlan 101-103
```

5. Verification

Run the **display ip igmp-info snooping gda-table** command on Device B to display the IGMP snooping forwarding entries and check whether the member port list of entry (*, 225.0.0.10, 104) contains GigabitEthernet 0/22, GigabitEthernet 0/20, and GigabitEthernet 0/19 and the member port list of entry (*, 239.0.0.11, 101) contains GigabitEthernet 0/20.

```
DeviceB# display ip igmp-info snooping gda-table
Multicast Switching Cache Table
  D: DYNAMIC
  S: STATIC
  M: MROUTE
(*, 239.0.0.11, 101):
  VLAN(101) 2 OPORTS:
    GigabitEthernet 0/20(D)
    GigabitEthernet 0/21(M)

(*, 225.0.0.10, 104):
  VLAN(104) 1 OPORTS:
    GigabitEthernet 0/18(M)

  VLAN(102) 1 OPORTS:
    GigabitEthernet 0/22(D)

  VLAN(101) 1 OPORTS:
    GigabitEthernet 0/20(D)

  VLAN(103) 1 OPORTS:
    GigabitEthernet 0/19(D)
```

Run the **display ip igmp-info snooping** command on Device B and check whether the IGMP snooping working mode is IVGL-SVGL.

```
DeviceB# display ip igmp-info snooping
IGMP Snooping running mode: IVGL-SVGL
IGMP Snooping L2-entry-limit: 64000
SVGL vlan: 104
SVGL profile number: 1
IGMP Snooping SVGL subvlan: 101-103
Source port check: Disable
Source ip check: Disable
IGMP Fast-Leave: Disable
```

```
IGMP Report suppress: Disable
IGMP Global Querier: Disable
IGMP Preview: Disable
IGMP Tunnel: Disable
IGMP Snooping version: 2
IGMP Preview group aging time: 60(Seconds)
Dynamic Mroute Aging Time: 300(Seconds)
Dynamic Host Aging Time: 260(Seconds)

vlan 1
----------
IGMP Snooping state: Enable
Multicast router learning mode: pim-dvmrp
IGMP Fast-Leave: Enabled
IGMP VLAN querier: Disable
IGMP VLAN Mode: STATIC

vlan 101
----------
IGMP Snooping state: Enable
Multicast router learning mode: pim-dvmrp
IGMP Fast-Leave: Enabled
IGMP VLAN querier: Disable
IGMP VLAN Mode: STATIC

vlan 102
----------
IGMP Snooping state: Enable
Multicast router learning mode: pim-dvmrp
IGMP Fast-Leave: Enabled
IGMP VLAN querier: Disable
IGMP VLAN Mode: STATIC

vlan 103
----------
IGMP Snooping state: Enable
Multicast router learning mode: pim-dvmrp
IGMP Fast-Leave: Enabled
IGMP VLAN querier: Disable
IGMP VLAN Mode: STATIC

vlan 104
----------
IGMP Snooping state: Enable
Multicast router learning mode: pim-dvmrp
IGMP Fast-Leave: Enabled
```

```
IGMP VLAN querier: Disable
IGMP VLAN Mode: STATIC
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
vlan 104
!
interface GigabitEthernet 0/47
 unset switch-port
 ip address 10.100.12.1 255.255.255.0
ip pim sparse-mode
!
interface GigabitEthernet 0/48
 switch-port mode trunk
 switch-port trunk native vlan 104
!
interface VLAN 104
 ip address 10.100.11.2 255.255.255.0
ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 10.100.12.0 0.0.0.255 area 0
!
ip pim c-bsr GigabitEthernet 0/48
ip pim c-rp GigabitEthernet 0/48
!
```

- Device B configuration file

```
hostname DeviceB
!
ip igmp group-profile 1
 permit
 range 224.1.1.1 238.1.1.1
!
vlan range 1,101-104
!
interface GigabitEthernet 0/18
 switch-port mode trunk
 switch-port trunk native vlan 104
!
```

```
interface GigabitEthernet 0/19
 switch-port access vlan 103
!
interface GigabitEthernet 0/20
 switch-port access vlan 101
!
interface GigabitEthernet 0/21
 switch-port access vlan 101
!
interface GigabitEthernet 0/22
 switch-port access vlan 102
!
interface VLAN 101
 ip address 10.100.101.1 255.255.255.0
!
interface VLAN 102
 ip address 10.100.102.1 255.255.255.0
!
interface VLAN 103
 ip address 10.100.103.1 255.255.255.0
!
interface VLAN 104
 ip address 10.100.11.1 255.255.255.0
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 10.100.101.0 0.0.0.255 area 0
 network-advertise 10.100.102.0 0.0.0.255 area 0
 network-advertise 10.100.103.0 0.0.0.255 area 0
!
ip igmp snooping-mode svgl shared-vlan 104
ip igmp snooping-mode svgl group-profile 1
ip igmp snooping-mode svgl subvlan 101-103
ip igmp snooping-mode ivgl-svgl
!
```

- Device C configuration file

```
hostname DeviceC
!
ip mrouting-enable
!
interface GigabitEthernet 0/21
 unset switch-port
 ip address 10.100.1.1 255.255.255.0
 ip pim sparse-mode
!
```

```
interface GigabitEthernet 0/23
 unset switch-port
 ip address 10.100.101.3 255.255.255.0
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.1.0 0.0.0.255 area 0
 network-advertise 10.100.101.0 0.0.0.255 area 0
!
```

7. Common Errors

- The multicast groups associated with the SVGL mode are not configured.
- The sent multicast traffic is not in the multicast groups associated with the SVGL mode.
- The group addresses of multicast traffic in IVGL mode are within the multicast group range in SVGL mode. As a result, IVGL forwarding cannot be implemented.

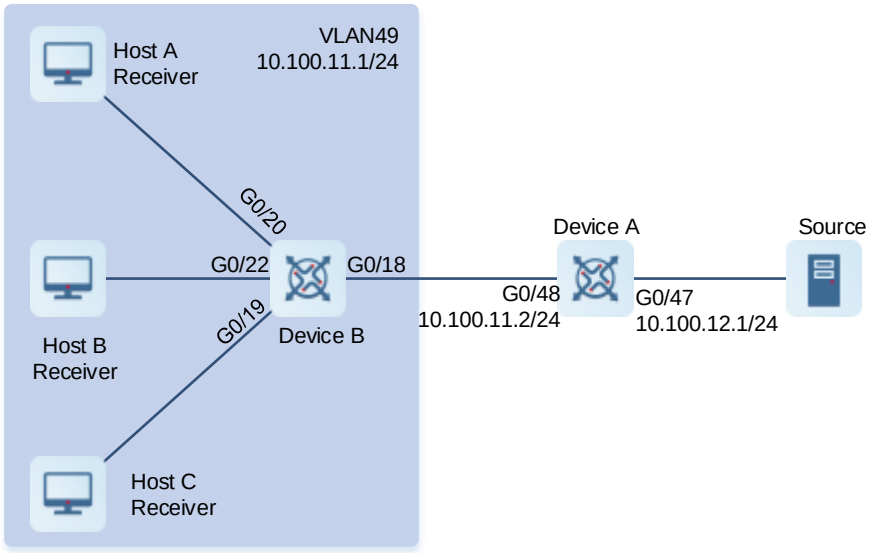
1.9.4 Configuring Static Ports to Implement L2 Multicast

1. Requirements

In the multicast network shown in Figure 1-8, Host A, Host B, and Host C need to receive data that a multicast source sends to a specific multicast group stably. The router (Device A) connects to the user network through the L2 device (Device B). The static IGMP multicast groups 225.0.0.10 to 225.0.0.12 are configured on the user-side L3 port on Device A. Host A and Host B want to receive data of multicast group 225.0.0.10 stably, and Host C wants to receive data of multicast group 225.0.0.11 stably.

2. Topology

Figure 1-8 Topology of Static Ports to Implement L2 Multicast



3. Notes

Configure IGMP snooping static multicast router ports and member ports on Device B.

- Configure the IP addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.
- Enable multicast routing and IGMP static multicast groups on Device A.
- Enable IGMP snooping on Device B.
- Configure static multicast router ports and member ports on Device B.

4. Procedure

(1) Configure the IP addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.

Configure the IP address and unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/47)# ip address 10.100.12.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/48)# ip address 10.100.11.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# router-protocol ospf 49
DeviceA(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 10.100.12.0 0.0.0.255 area 0
DeviceA(cfg-router)# exit
```

Configure the IP address, VLAN, and unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan 49
DeviceB(cfg-vlan)# exit
DeviceB(cfg)# interface vlan 49
DeviceB(cfg-if-VLAN 49)#ip address 10.100.11.1 255.255.255.0
DeviceB(cfg-if-VLAN 49)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/19)# exit
DeviceB(cfg)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/20)# exit
```

```
DeviceB(cfg)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/22)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/22)# exit
DeviceB(cfg)# router-protocol ospf 49
DeviceB(cfg-router)# network-advertise 10.100.11.0 0.0.0.255 area 0
DeviceB(cfg-router)# exit
```

(2) Enable IP multicast routing and PIM-SM related functions on Device A.

```
DeviceA(cfg)# ip mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# ip igmp static-group 225.0.0.10
DeviceA(cfg-if-GigabitEthernet 0/48)# ip igmp static-group 225.0.0.11
DeviceA(cfg-if-GigabitEthernet 0/48)# ip igmp static-group 225.0.0.12
DeviceA(cfg-if-GigabitEthernet 0/48)# ip pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# ip pim c-bsr GigabitEthernet 0/47
DeviceA(cfg)# ip pim c-rp GigabitEthernet 0/47
```

(3) Enable basic IGMP snooping functions on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip igmp snooping-mode ivgl
```

(4) Configure static multicast router ports and member ports on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip igmp snooping-mode vlan 49 static-member 225.0.0.11 interface GigabitEthernet 0/19
DeviceB(cfg)# ip igmp snooping-mode vlan 49 static-member 225.0.0.10 interface GigabitEthernet 0/22
DeviceB(cfg)# ip igmp snooping-mode vlan 49 static-member 225.0.0.10 interface GigabitEthernet 0/20
DeviceB(cfg)# ip igmp snooping-mode vlan 49 multicast-router interface GigabitEthernet 0/18
```

5. Verification

Run the **display ip igmp-info snooping multicast-router** command and check whether the static router port is GigabitEthernet 0/18.

```
DeviceB# display ip igmp-info snooping mrouter
Multicast Switching Mroute Port
  D: DYNAMIC
  S: STATIC
(*, *, 49):
  VLAN(49)   1 MROUTES:
    GigabitEthernet 0/18(DS)

(*, *, 101):
  VLAN(101)   1 MROUTES:
```

```
    GigabitEthernet 0/21(D)
```

Run the **display ip igmp-info snooping gda-table** command and check whether the static member ports of multicast group 225.0.0.10 are GigabitEthernet 0/20 and GigabitEthernet 0/22 and the static member port of multicast group 225.0.0.11 is GigabitEthernet 0/19.

```
DeviceB# display ip igmp-info snooping gda-table
Multicast Switching Cache Table
  D: DYNAMIC
  S: STATIC
  M: MROUTE
(*, 225.0.0.10, 49):
  VLAN(49) 3 OPORTS:
    GigabitEthernet 0/18(M)
    GigabitEthernet 0/20(S)
    GigabitEthernet 0/22(S)

(*, 225.0.0.11, 49):
  VLAN(49) 2 OPORTS:
    GigabitEthernet 0/18(M)
    GigabitEthernet 0/19(S)
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip mrouting-enable
!
interface GigabitEthernet 0/47
 unset switch-port
 ip address 10.100.12.1 255.255.255.0
ip pim sparse-mode
!
interface GigabitEthernet 0/48
 unset switch-port
 ip address 10.100.11.2 255.255.255.0
 ip igmp static-group-entries 225.0.0.10
 ip igmp static-group-entries 225.0.0.11
 ip igmp static-group-entries 225.0.0.12
 ip pim sparse-mode
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
 network-advertise 10.100.12.0 0.0.0.255 area 0
```

```
!
ip pim c-bsr GigabitEthernet 0/47
ip pim c-rp GigabitEthernet 0/47
!
```

- Device B configuration file

```
hostname DeviceB
!
vlan range 1,49
!
interface GigabitEthernet 0/18
 switch-port access vlan 49
!
interface GigabitEthernet 0/19
 switch-port access vlan 49
!
interface GigabitEthernet 0/20
 switch-port access vlan 49
!
interface GigabitEthernet 0/22
 switch-port access vlan 49
!
interface VLAN 49
 ip address 10.100.11.1 255.255.255.0
!
router-protocol ospf 49
 graceful-restart
 network-advertise 10.100.11.0 0.0.0.255 area 0
!
ip igmp snooping-mode ivgl
!
ip igmp snooping-mode vlan 49 static-member 225.0.0.11 interface GigabitEthernet 0/19
ip igmp snooping-mode vlan 49 static-member 225.0.0.10 interface GigabitEthernet 0/22
ip igmp snooping-mode vlan 49 static-member 225.0.0.10 interface GigabitEthernet 0/20
ip igmp snooping-mode vlan 49 multicast-router interface GigabitEthernet 0/18
!
```

7. Common Errors

- Basic IGMP snooping functions are not configured or fail to be configured.

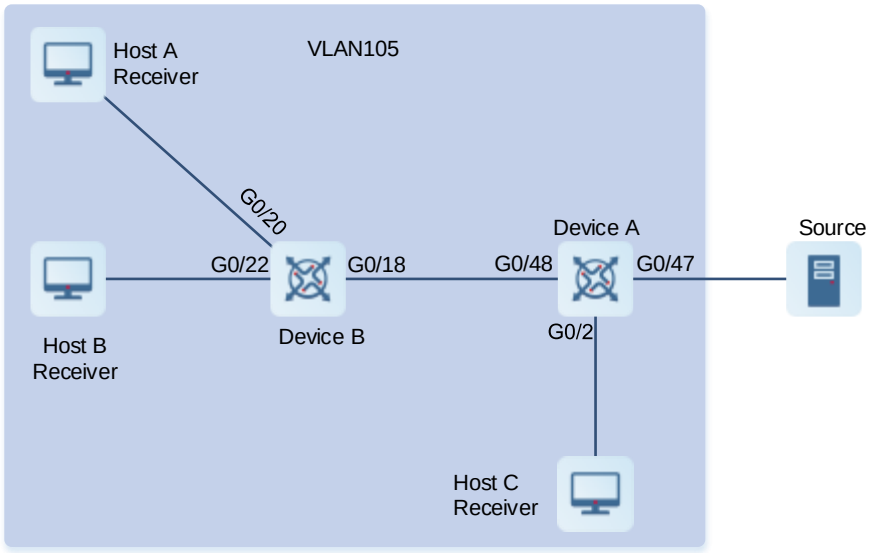
1.9.5 Configuring IGMP Snooping Querier

1. Requirements

In a two-layer multicast network, as shown in Figure 1-9, the multicast source sends multicast data to multicast groups 225.0.0.10 and 225.0.0.11. Host A and Host C join multicast group 225.0.0.10, and Host B joins multicast group 225.0.0.11. All receivers use IGMPv2. IGMP snooping is used to transfer multicast data in the two-layer network.

2. Topology

Figure 1-9 Topology of IGMP Snooping Querier



3. Notes

Enable IGMP snooping on Device A and Device B, and configure Device A as the IGMP snooping querier.

- Configure VLANs and IP addresses on Device A and Device B.
- Enable IGMP snooping and run the IVGL mode on Device A and Device B.
- Configure the IGMP snooping querier function for VLAN 105 on Device A.

4. Procedure

(1) Configure VLANs and IP addresses on Device A and Device B.

Configure the VLAN and IP address on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan range 1,105
DeviceA(cfg)# interface VLAN105
DeviceA(cfg-if-VLAN105)# ip address 10.100.12.1 255.255.255.0
DeviceA(cfg-if-VLAN105)# exit
DeviceA(cfg)# interface GigabitEthernet0/47
DeviceA(cfg-if-GigabitEthernet0/47)# switch-port access vlan 105
DeviceA(cfg-if-GigabitEthernet0/47)# interface GigabitEthernet0/48
DeviceA(cfg-if-GigabitEthernet0/48)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet0/48)# switch-port trunk native vlan 105
DeviceA(cfg-if-GigabitEthernet0/48)# exit
```

Configure the VLAN and IP address on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
```

```
DeviceB(cfg)# vlan range 1,105
DeviceB(cfg)# interface VLAN105
DeviceB(cfg-if-VLAN105)# ip address 10.100.12.3 255.255.255.0
DeviceB(cfg-if-VLAN105)# exit
DeviceB(cfg-if-GigabitEthernet 0/18)# interface GigabitEthernet 0/2
DeviceB(cfg-if-GigabitEthernet 0/2)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/2)# exit
DeviceB(cfg)#interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port trunk native vlan 105
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 105
DeviceB(cfg-if-GigabitEthernet 0/20)# exit
DeviceB(cfg)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/22)# switch-port access vlan 105
DeviceB(cfg-if-GigabitEthernet 0/22)# exit
```

(2) Configure basic IGMP snooping functions and run the IVGL mode on Device A and Device B.

Enable IGMP snooping and run the IVGL mode on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip igmp snooping-mode ivgl
```

Enable IGMP snooping and run the IVGL mode on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip igmp snooping-mode ivgl
```

(3) Configure the IGMP snooping querier function on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)#ip igmp snooping-mode querier
DeviceA(cfg)#ip igmp snooping-mode querier source-address 10.100.12.1
DeviceA(cfg)#ip igmp snooping-mode vlan 105 querier
```

5. Verification

Run the **display ip igmp-info snooping querier** command to check whether the querier of VLAN 1 takes effect.

```
DeviceA# display ip igmp-info snooping querier
Vlan        IP Address              IGMP Version            Port
------------------------------------------------------------
105         10.100.12.1                 2                    switch
DeviceA#
DeviceA# display ip igmp-info snooping querier vlan 105

Vlan 105:    IGMP switch querier status
```

```
------------------------------------------------------------
elected querier is 10.100.12.1              (this switch querier)
------------------------------------------------------------
admin state                          : Enable
admin version                        : 2
source IP address                    : 10.100.12.1
query-interval (sec)                 : 60
max-response-time (sec)              : 10
querier-timeout (sec)                : 125
operational state                    : Querier
operational version                  : 2
```

Run the **display ip igmp-info snooping gda-table** command on Device A and check whether the outbound ports of multicast group 225.0.0.10 are GigabitEthernet0/2 and GigabitEthernet0/48 and the outbound port of multicast group 225.0.0.11 is GigabitEthernet0/48.

```
DeviceA# display ip igmp-info snooping gda-table
Multicast Switching Cache Table
  D: DYNAMIC
  S: STATIC
  M: MROUTE
(*, 225.0.0.10, 105):
  VLAN(105) 2 OPORTS:
    GigabitEthernet0/2(D)
    GigabitEthernet0/48(D)

(*, 225.0.0.11, 105):
  VLAN(105) 1 OPORTS:
    GigabitEthernet0/48(D)
```

Run the **display ip igmp-info snooping gda-table** command on Device B and check whether the outbound port of multicast group 225.0.0.10 is GigabitEthernet0/20 and the outbound port of multicast group 225.0.0.11 is GigabitEthernet0/22.

```
DeviceB# display ip igmp-info snooping gda-table
Multicast Switching Cache Table
  D: DYNAMIC
  S: STATIC
  M: MROUTE
(*, 225.0.0.10, 105):
  VLAN(105) 2 OPORTS:
    GigabitEthernet 0/18(M)
    GigabitEthernet 0/20(D)

(*, 225.0.0.11, 105):
  VLAN(105) 2 OPORTS:
    GigabitEthernet 0/18(M)
    GigabitEthernet 0/22(D)
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
vlan range 1,105
!
interface GigabitEthernet0/2
 switch-port access vlan 105
!
interface GigabitEthernet0/47
 switch-port access vlan 105
!
interface GigabitEthernet0/48
 switch-port mode trunk
 switch-port trunk native vlan 105
!
interface VLAN105
 ip address 10.100.12.1 255.255.255.0
!
ip igmp snooping-mode ivgl
ip igmp snooping-mode querier
ip igmp snooping-mode querier source-address 10.100.12.1
ip igmp snooping-mode vlan 105 querier
!
```

- Device B configuration file

```
hostname DeviceA
!
vlan range 1,105
!
interface GigabitEthernet 0/18
 switch-port mode trunk
 switch-port trunk native vlan 105
!
interface GigabitEthernet 0/20
 description cyx
 switch-port access vlan 105
!
interface GigabitEthernet 0/22
 description cyx
 switch-port access vlan 105
!
interface VLAN 105
 ip address 10.100.12.3 255.255.255.0
```

```
!
ip igmp snooping-mode ivgl
!
```

7. Common Errors

- The source IP address of Query packets sent by the querier is not configured. As a result, the querier function does not take effect.

Contents

1 Configuring MSDP 1
1.1 Introduction 1
1.1.1 Principles 1
1.1.2 Protocols and Standards 8
1.2 Restrictions and Guidelines 8
1.3 Configuration Task Summary 8
1.4 Prerequisites 9
1.5 Managing MSDP Peers 9
1.5.1 Overview 9
1.5.2 Configuration Tasks 9
1.5.3 Configuring MSDP Peers 10
1.5.4 Configuring the Default MSDP Peer 10
1.5.5 Configuring an MSDP Mesh Group 11
1.5.6 Configuring the TCP Reconnection Interval of an MSDP Peer 11
1.5.7 Enabling MD5 Authentication on TCP Connections Between MSDP Peers 12
1.5.8 Configuring the Number of MSDP Peers Supported on a Device 12
1.6 Restricting SA Messages 13
1.6.1 Overview 13
1.6.2 Configuration Tasks 13
1.6.3 Prerequisites 13
1.6.4 Configuring the TTL Value of Multicast Packets in SA Messages 13
1.6.5 Enabling the Function of Filtering Source Information Released Locally 14

1.6.6 Enabling the Function of Filtering Received SA Requests 14
1.6.7 Enabling the Function of Filtering Received SA Messages 15
1.6.8 Enabling the Function of Filtering Sent SA Messages 16
1.6.9 Configuring SA Cache Capacity 17
1.7 Configuring an Anycast RP 17
1.7.1 Overview 17
1.7.2 Prerequisites 17
1.7.3 Procedure 17
1.8 Monitoring 18
1.9 Configuration Examples 19
1.9.1 Enabling Basic Functions of MSDP 19
1.9.2 Configuring an Anycast RP 25
1.9.3 Enabling MD5 Authentication 30
1.9.4 Managing MSDP Peers 33
1.9.5 Modifying Protocol Parameters 35

1 Configuring MSDP

1.1 Introduction

Multicast Source Discovery Protocol (MSDP) is used to connect multiple rendezvous points (RPs) on the multicast network and implement the following two functions.

- Use MSDP among multiple Protocol Independent Multicast Sparse-Mode (PIM-SM) domains to share the multicast source information of these PIM-SM domains to implement cross-domain multicast.

  With the increase of a network, a PIM network can be divided into multiple PIM-SM domains to facilitate management. The RPs in a domain cannot understand multicast source information in other domains. MSDP peers are created among the PIM-SM domains to interact source-active (SA) messages so that multicast source information can be shared and multicast packets can be forwarded to hosts in other domains.

- Use MSDP in a PIM-SM domain to share the multicast source information of multiple RPs to implement an anycast RP.

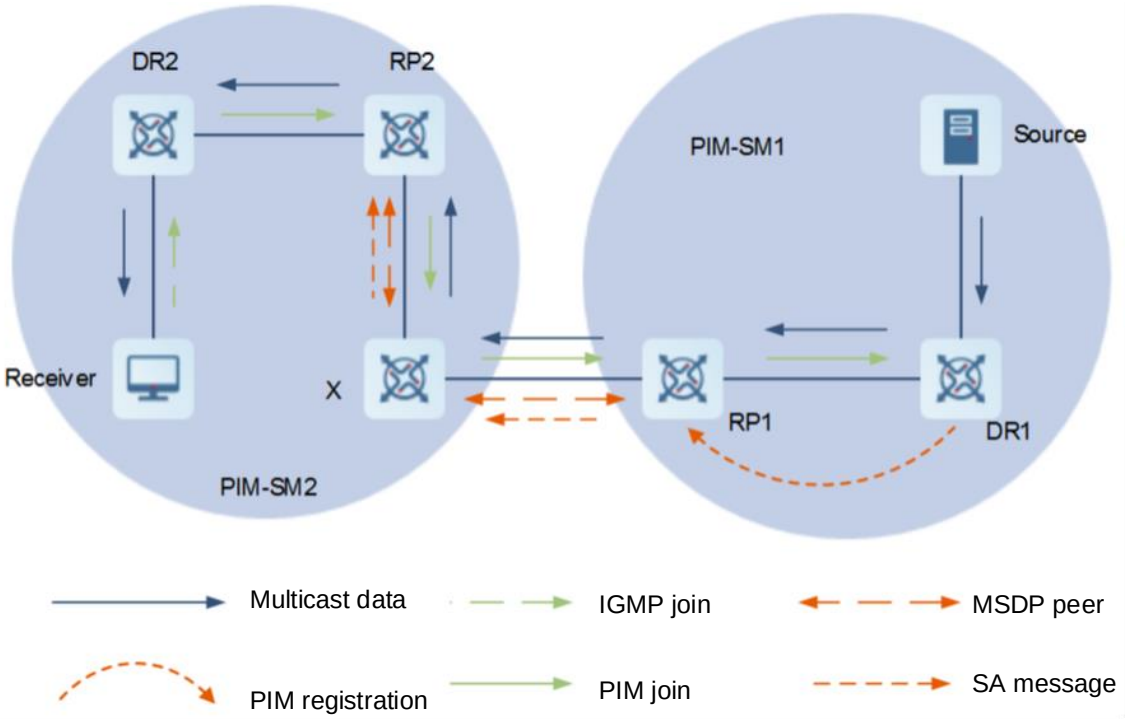
  After RPs in a PIM-SM domain are configured with the same address and MSDP peers are created on these RPs, the RPs can implement load sharing and backup in the PIM-SM domain.

1.1.1 Principles

1. MSDP Peer

Configure one or more pairs of MSDP peers on the network to connect RPs. The peers interact SA messages so that an RP can notify other RPs of the multicast source information on this RP. So far as the unicast route is reachable, the MSDP peer relationship can be established. MSDP peers can be configured on RPs as well as on any other PIM device, for example, X in Figure 1-1.

Figure 1-1 Positions of MSDP Peers



- RP connected to the multicast source

  Configure the MSDP peer on the RP connected to the multicast source. Then, this RP can use SA messages to send the local multicast source information to other RPs.

  As shown in Figure 1-1, DR 1 registers the multicast source information with RP 1. As a peer relationship is established between RP 1 and the PIM device X, RP 1 sends the multicast source information to the PIM device X.

- SA message forwarder

  Non-RPs (common PIM devices) can also act as MSDP peers, but only forward SA messages.

  As shown in Figure 1-1, X forwards SA messages sent from RP 1 to RP 2. In this way, the multicast source information is transferred to RP 2.

- RP connected to the multicast receiver

  Configure the MSDP peer on the RP connected to the multicast receiver. Then, this RP can trigger a join towards the multicast source based on the received SA message.

  As shown in Figure 1-1, DR 2 triggers a join towards RP 2. As RP 2 already obtains the multicast source information, RP 2 continues to trigger a join towards the multicast source, thus establishing a multicast distribution tree (MDT) from DR 1 to DR 2.

2. MSDP Packet

An MSDP packet is encapsulated in a TCP packet and follows the format of type-length-value (TLV), as shown in Figure 1-2.

Figure 1-2 Format of MSDP Packet

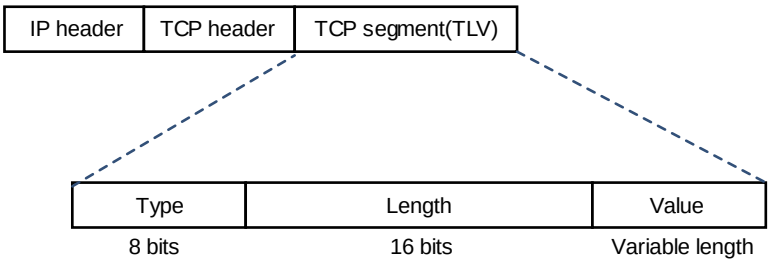


Table 1-1 Fields of MSDP Packet

Field	Length	Description
Type	8 bits	Specifies a message type. ● **1**: Source-Active. Messages of this type carry multiple pairs of (S, G) information and they are transmitted among RPs. PIM-SM multicast packets can be encapsulated in the messages. ● **2**: Source-Active Request. Messages of this type request an (S, G) list of a specified group G to reduce the source join delay. ● **3**: Source-Active Response. Messages of this type are responses to the Source-Active Request. ● **4**: Keep Alive. Messages of this type hold the connection relationship of MSDP peers. ● **5**: Reserved. This is the reserved type of messages. ● **6**: MSDP traceroute in progress. Messages of this type specify the Traceroute function of MSDP and detect the reverse path forwarding (RPF) of SA messages. ● **7**: MSDP traceroute reply. Messages of this type specify the Traceroute function of MSDP and detect the RPF of SA messages.
Length	16 bits	Specifies the length of a message in bytes. The length includes the Type, Length and Value fields. Except for the Keep Alive message, the length of other messages must be 4 bytes at least and 9192 bytes at most.
Value	Subject to the length of the variable	Specifies the message content, which varies with the message type.

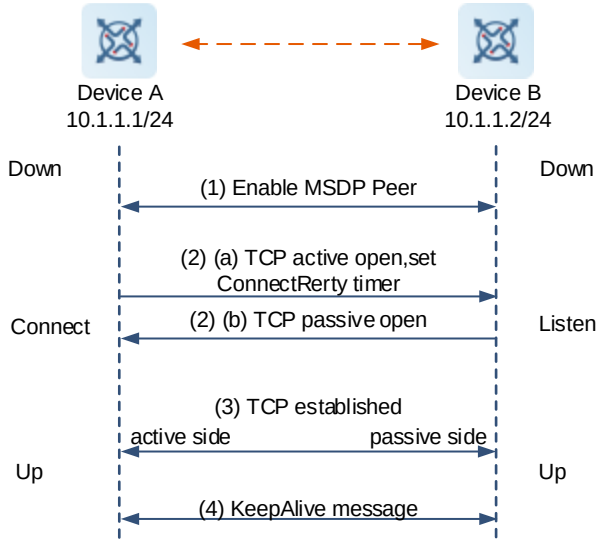
Based on Table 1-1, SA messages can carry multiple pairs of (S, G) information and they are transmitted among RPs. PIM-SM multicast packets can be encapsulated in the messages. The peers interact SA messages with each other to share multicast source information. Multicast packets are encapsulated in the SA messages to avoid the (S, G) entries in the SA messages timeout. If this is the case, hosts cannot receive the multicast packets of the source.

MSDP supports periodic sending of SA requests and SA responses to improve the source information update efficiency. With the mechanism, new hosts do not need to wait for an SA sending period to obtain valid (S, G) information, which reduces the time for the hosts to join a shortest path tree (SPT).

3. MSDP Peer Establishment Procedure

MSDP peers are connected over TCP through port 639. After a TCP connection is established between the peers, the connection is maintained through the keep alive messages.

Figure 1-3 MSDP Peer Connection Establishment Procedure



As shown in Figure 1-3, initially, the MSDP session state of device A and device B is down. The MSDP peer connection establishment procedure is as follows:

(1) Device A and device B specify the other party as a peer mutually and they compare their addresses.

a The IP address of device A is smaller than that of device B. Therefore, device A enters the connection status, initiates a connection request to device B, and starts a timer for reconnection.

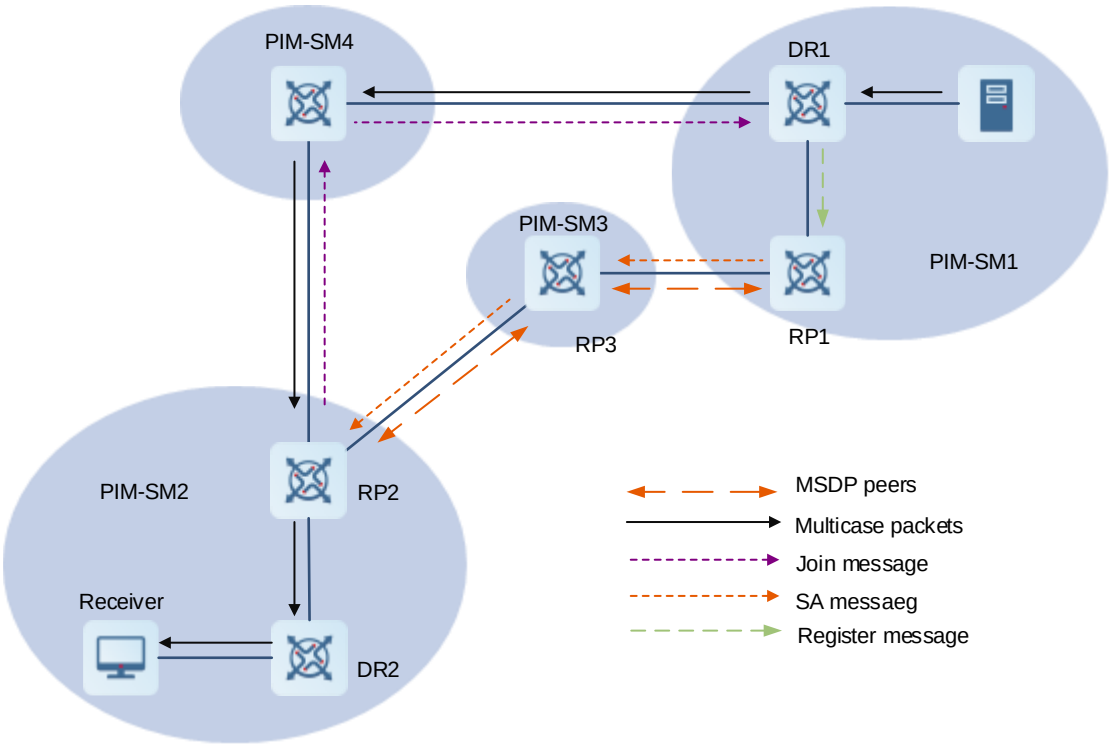
b The IP address of device B is greater. Therefore, device B enters the listening status and waits for the connection request.

(2) After the MSDP session is established between device A and device B, the connection status is up. The two devices send keep alive messages to each other to notify the other party of keeping the MSDP session.

4. Inter-domain Multicast of MSDP

As shown in Figure 1-4, the multicast network is divided into four PIM-SM domains. In PIM-SM 1, the active source sends multicast data to the multicast group and registers with RP 1. In PIM-SM 2, the receiver joins the group, expecting to receive multicast data from the source in PIM-SM 1. The devices in PIM-SM 2 and PIM-SM 3 need to know the position of the source so as to obtain the corresponding multicast data. To implement this purpose, MSDP peer relationship must be established between RP 1 and RP 3 and between RP 3 and RP 2.

Figure 1-4 Inter-domain Multicast of MSDP



The working process of PIM-SM inter-domain multicast is as follows:

(1) The multicast source in PIM-SM 1 sends multicast data to the group. The multicast data is sent to source DR 1 for registration and encapsulated in messages and sent to RP 1. RP 1 obtains the multicast source information.

(2) Source RP 1 sends SA messages to RP 3 periodically. An SA message contains the multicast source address S, multicast group address G, and RP 1 address.

(3) Upon receiving the SA messages, RP 3 performs RPF checking and filters forwarding policies. After the messages pass the checking, they are sent to RP 2. Then, PIM-SM 2 and PIM-SM 3 obtain the source information.

(4) Upon receiving the SA messages, RP 2 checks whether the receiver of the group exists in PIM-SM 2.

- If the receiver exists in PIM-SM 2, an RPT about the group is created between RP 2 and the receiver. RP 2 creates an (S, G) pair and sends an (S, G) join packet to the source hop by hop to generate an SPT. The multicast data is sent to RP 2 along the SPT and then forwarded to the receiver along the RPT. Upon receiving the multicast data, member DR 2 decides whether to perform SPT switchover based on configuration.
- If the receiver does not exist in PIM-SM 2, RP 2 does not create an (S, G) pair or send join packets to the source.

5. Receiving and Forwarding SA Messages

An SA message contains the multicast source address, multicast group address, and RP address. The RP address is the IP address of the RP with which the multicast source is registered.

- The RP encapsulates the locally registered multicast source information in an SA message, and sends the

message to all its MSDP peers.

- On receiving the SA message, each MSDP peer performs the peer RPF check, compares the SA message with the content in the SA cache, and matches the SA message against the SA incoming and outgoing filtering rules. If the SA message passes the peer RPF check and meets the outgoing filtering rules and does not exist in the SA cache, this SA message is forwarded to other MSDP peers.
- For any SA message coming from an MSDP peer (address: N), the following peer RPF check rules are used to prevent loops and avoid SA message flooding:

  a If N is a member of the mesh group, the SA message passes the peer RPF check; otherwise, go to step b.

  b If N is the only active MSDP peer on the local device, the SA message passes the peer RPF check; otherwise, go to step c.

  c If N is the RP address in the SA message, the SA message passes the peer RPF check; otherwise, go to step d.

  d If an EBGP route to the RP address in the SA message exists on the local device and the next hop of this route is N, the SA message passes the peer RPF check; otherwise, go to step e.

  e If an optimum route to the RP address in the SA message exists on the local device, check as follows:

  If this optimum route is a distance vector route (such as the BGP/RIP route), and this router is advertised by N, the SA message passes the peer RPF check.

  If this optimum route is a link status route (such as the OSPF/IS-IS route), and the next hop of this router is N, the SA message passes the peer RPF check.

  Otherwise, go to step f.

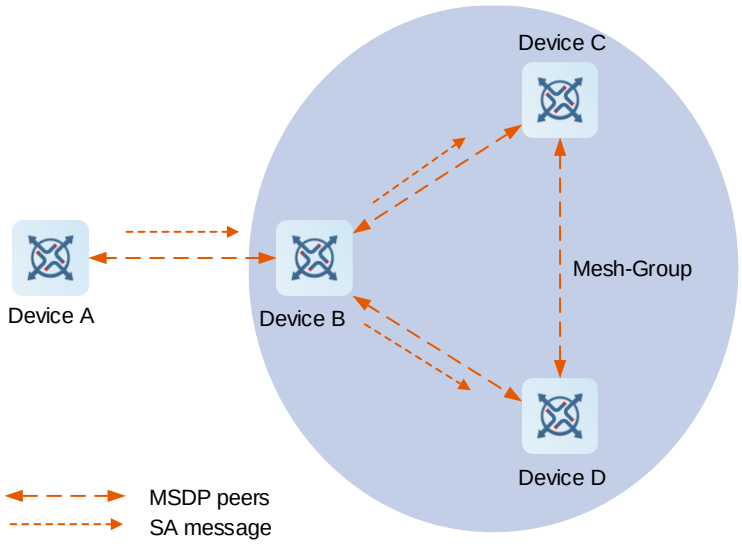
  f If an optimum route to the RP address in the SA message exists on the local device, and this route is an MBGP/BGP route, extract the nearest autonomous system (AS) of the AS path of this MBGP/BGP route. If the local device has multiple MSDP peers in this AS and N is the MSDP peer with the largest IP address, or N is the only MSDP peer in this AS, the SA message passes the peer RPF check; otherwise, go to step g.

  g If N is the default MSDP peer, the SA message passes the peer RPF check; otherwise, go to step h.

  h The SA message fails in the peer RPF check.

- When MSDP peer relationship is established on any two members in a group, this group is a mesh group. The mesh group helps reduce the number of SA messages. As shown in Figure 1-5, device B, device C, and device D belong to the same mesh group.
  - For SA messages from device A outside the mesh group and received by device B in the mesh group, after the messages pass the peer RPF check and SA cache comparison, these SA messages are forwarded to other members in the group.
  - For SA messages from device B to device C and device D, they are no longer forwarded to other members in the group.

Figure 1-5 Mesh Group Topology



- The SA cache buffers status of SA messages. Expired SA messages will be deleted.
  - When an MSDP peer receives an SA message

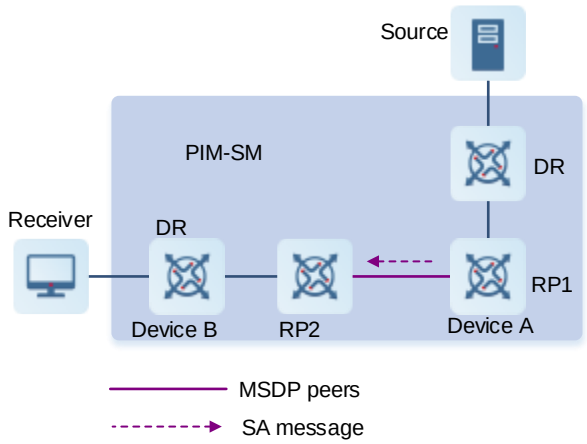
    If this message does not exist in the SA cache and passes the peer RPF check, the message is stored in the SA cache.

    If this message already exists in the SA cache, the message is ignored. This helps suppress SA message flooding.
  - When an MSDP peer receives an SA message, if this message already exists in the SA cache, the message is immediately responded. This helps improve the protocol efficiency.

6. Anycast RP based on MSDP

Generally, each multicast group in a PIM-SM network is served by only one RP. If the RP is faulty or fails, the multicast traffic may be interrupted. Anycast RP enables multiple RPs with the same address to serve a multicast group to implement redundant backup between the RPs. The multicast sources or group members register with the nearest RPs so that the RPs synchronize source information. Anycast RP is implemented by creating MSDP peer relationship between these RPs.

Figure 1-6 Anycast RP based on MSDP



As shown in Figure 1-6, device A and device B in a PIM-SM domain establish peer relationship. The source sends multicast data to the group, and the receiver receives the multicast data as a member of the group. The working process of Anycast RP based on MSDP is as follows:

(1) When the source DR receives the multicast data from the source, the source DR registers with the nearest RP 1.

(2) When the member DR receives a join request packet from the receiver, the member DR forwards the join packet to RP 2. In this way, an RPT with RP 2 as the root is formed.

(3) RP 1 and RP 2 interact SA messages and share the (S, G) information of the multicast source. Upon obtaining the multicast source information, RP 2 sends the join packet to the source hop by hop to generate an SPT.

(4) The multicast data is forwarded to RP 2 through the SPT and then forwarded to the receiver through the RPT. The member DR decides whether to perform SPT switchover based on configuration.

1.1.2 Protocols and Standards

- RFC 3618: Multicast Source Discovery Protocol (MSDP)
- RFC 3446: Anycast Rendevous Point (RP) mechanism using Protocol Independent Multicast (PIM) and Multicast Source Discovery Protocol (MSDP)

1.2 Restrictions and Guidelines

Establish the MSDP peer relationship between multiple ASs so that group members can apply for the multicast streams across ASs.

- The inter-AS unicast route must be reachable.
- Run PIM-SM within each AS, and configure the BSR border.
- The IP address and local interface of the MSDP peer must be the same as those of the EBGP peer.

1.3 Configuration Task Summary

The IP multicast configuration includes the following tasks:

(1) Managing MSDP Peers.

- Configuring MSDP Peers
- (Optional) Configuring the Default MSDP Peer
- (Optional) Configuring an MSDP Mesh Group
- (Optional) Configuring the TCP Reconnection Interval of an MSDP Peer
- (Optional) Enabling MD5 Authentication on TCP Connections Between MSDP Peers
- (Optional) Configuring the Number of MSDP Peers Supported on a Device

(2) Restricting SA Messages. All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the TTL Value of Multicast Packets in SA Messages
- Enabling the Function of Filtering Source Information Released Locally
- Enabling the Function of Filtering Received SA Requests
- Enabling the Function of Filtering Received SA Messages
- Enabling the Function of Filtering Sent SA Messages
- Configuring SA Cache Capacity

(3) Configuring an Anycast RP.

1.4 Prerequisites

- Unicast routing protocols are configured and the network in a domain is reachable.
- Basic PIM-SM functions are configured to implement multicast in a domain.

1.5 Managing MSDP Peers

1.5.1 Overview

MSDP peers in different PIM-SM domains interact SA messages to share inter-domain multicast source information.

1.5.2 Configuration Tasks

The MSDP peer management configuration includes the following tasks:

(1) Configuring MSDP Peers

(2) (Optional) Configuring the Default MSDP Peer

(3) (Optional) Configuring an MSDP Mesh Group

(4) (Optional) Configuring the TCP Reconnection Interval of an MSDP Peer

(5) (Optional) Enabling MD5 Authentication on TCP Connections Between MSDP Peers

(6) (Optional) Configuring the Number of MSDP Peers Supported on a Device

1.5.3 Configuring MSDP Peers

1. Overview

MSDP peers are identified based on addresses. The relationship between MSDP peers is bidirectional. Therefore, they must be created on both ends. You can manage MSDP peers by adding description information to a specified MSDP. A connection with a specified peer can be temporarily disabled by configuration.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Establish an MSDP peer relationship.

ip msdp add-peer *ipv4-peer-address* **source-interface** *interface-type interface-number*

No MSDP peer is added by default.

(4) (Optional) Configure description information for an MSDP peer.

ip msdp peer-description *ipv4-peer-address description*

No description information is added for an MSDP peer by default.

(5) (Optional) Disable an MSDP peer.

ip msdp peer-closing *ipv4-peer-address*

The connection with an MSDP peer is enabled by default.

Only the TCP connection with the MSDP peer is disabled, and this MSDP peer and the configuration of this MSDP peer are reserved.

1.5.4 Configuring the Default MSDP Peer

1. Overview

On an MSDP peer, if it is not necessary to perform peer RPF check on SA messages sent from a specified peer, configure this peer as the default peer.

2. Restrictions and Guidelines

- If the **prefix-list** keyword parameter is not specified, all SA messages are accepted.
- If an inexistent **prefix-list** is specified, all SA messages are accepted.
- If an existent **prefix-list** is specified, only the SA messages of RPs specified in this prefix list are accepted.

3. Prerequisites

- An MSDP peer relationship is established.
- If the **prefix-list** keyword parameter is specified, the prefix list must be created before a default MSDP peer is configured. For details about the prefix list configuration, see *Configuring Routing Policies*.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a default MSDP peer.

ip msdp peer-default-addr *ipv4-peer-address* [**prefix-list** *prefix-list-name*]

No default MSDP peer is configured by default.

1.5.5 Configuring an MSDP Mesh Group

1. Overview

When MSDP peer relationship is established on any two members in a group, this group is a mesh group. The mesh group helps reduce the number of SA messages and avoids SA message flooding between the peers. SA messages received by peers in a mesh group are processed in one of the following ways:

- For SA messages coming outside the mesh group, after the SA messages pass the peer RPF check and SA cache comparison, they are forwarded to the other members in the group.
- For SA messages from the internal mesh group, they are received without passing peer RPF check and are no longer forwarded to the other members in the group.

2. Restrictions and Guidelines

Among multiple MSDP peers, if SA messages from these peers pass RPF check by default, these peers can be added to a mesh group.

3. Prerequisites

An MSDP peer relationship is established.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a mesh group.

ip msdp peer-to-mesh-group *mesh-name ipv4-peer-address*

No mesh group is configured and no MSDP peer is added to any mesh group by default.

1.5.6 Configuring the TCP Reconnection Interval of an MSDP Peer

1. Overview

Within the TCP reconnection interval, the MSDP peer on the proactive connection side can initiate at most one TCP connection. In some application scenarios, for example, MSDP peer creation, restart, or failback, you can shorten the TCP reconnection interval to accelerate convergence of the MSDP peer relationship.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the TCP reconnection interval of an MSDP peer.

ip msdp retry-period *interval*

The default TCP reconnection interval is **30** seconds.

1.5.7 Enabling MD5 Authentication on TCP Connections Between MSDP Peers

1. Overview

MD5 authentication is enabled on TCP connections between MSDP peers to prevent invalid TCP connections and improve MSDP security. To authenticate the ID of an MSDP peer, you can enable MD5 authentication on the TCP connection established with this MSDP peer.

2. Restrictions and Guidelines

- The MSDP peers must have the consistent configuration, and the ciphers must be the same; otherwise, TCP connections fail.
- If the encryption level is set to 7, the cipher text length must be an even number equal to or greater than 4; otherwise, the configuration fails.
- If the configuration or cipher changes, the local device does not stop the current session. Instead, it attempts to use a new cipher to retain the current session until timeout.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable MD5 authentication on TCP connections between MSDP peers.

ip msdp peer-connection password *ipv4-peer-address* [**0** | **7**] *password-string*

The MD5 authentication function is disabled by default.

1.5.8 Configuring the Number of MSDP Peers Supported on a Device

1. Overview

You can configure the number of MSDP peers supported on a device. If the default number (64) cannot meet requirements, the number of the MSDP peers can be increased.

2. Restrictions and Guidelines

When the number of MSDP peers on the device exceeds the configured number, a notification is displayed to indicate configuration failure. In this case, you need to delete some peers to make the configuration take effect.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the number of MSDP peers supported on a device.

ip msdp peer-limit *peer-limit*

The maximum number of MSDP peers supported on a device is **64** by default.

1.6 Restricting SA Messages

1.6.1 Overview

MSDP peers share multicast source information by exchanging SA messages. Based on the actual network conditions, multicast packets encapsulated in SA messages can be defined, and SA requests and SA messages can be filtered.

1.6.2 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the TTL Value of Multicast Packets in SA Messages
- Enabling the Function of Filtering Source Information Released Locally
- Enabling the Function of Filtering Received SA Requests
- Enabling the Function of Filtering Received SA Messages
- Enabling the Function of Filtering Sent SA Messages
- Configuring SA Cache Capacity

1.6.3 Prerequisites

MSDP peer relationship is established between devices.

1.6.4 Configuring the TTL Value of Multicast Packets in SA Messages

1. Overview

If the interval of a multicast source for sending multicast data exceeds the timeout of an (S, G) entry, remote users cannot receive the multicast data sent by the multicast source. After the function of encapsulating multicast packets in SA messages is enabled on the source RP, the source RP encapsulates multicast packets in SA messages and sends them to the remote RPs. Upon receiving the messages, the remote RPs decapsulate the messages and transmit the packets in the local domain.

By configuring the time-to-live (TTL) value of multicast packets encapsulated in SA messages, you can control the sending of the multicast packets encapsulated in the SA messages.

- A multicast packet can be sent to MSDP peers only when the TTL value in the IP header of the multicast packet is equal to or greater than the preset TTL value.
- If the TTL value in the IP header of the multicast packet is smaller than the preset TTL value, the multicast packet is removed from the SA message and discarded before the SA message is sent to the MSDP peer.

Therefore, this function affects the sending of multicast packets in SA messages, and does not affect the sending of (S, G) information in the SA messages.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the TTL value of multicast packets in SA messages.

ip msdp sa-ttl-limit *ipv4-peer-address ttl-value*

The TTL value of the multicast packets in SA messages is not defined by default.

1.6.5 Enabling the Function of Filtering Source Information Released Locally

1. Overview

After source information registered on the local RP is filtered on an MSDP device, only allowed (S, G) information of the local or other domains can be released on the MSDP device.

2. Restrictions and Guidelines

- If the **list** keyword parameter is specified, only the (S, G) information matching this ACL is released.
- If the **route-map** keyword parameter is specified, only the (S, G) information matching this route map is released.
- If these two parameters are specified, only the (S, G) information matching the ACL and route map is released.
- If no parameter is specified, no (S, G) information is released.

3. Prerequisites

Before an IP multicast border is configured, ensure that the corresponding ACL or route map is created and an (S, G) range is specified.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Filter the source information released locally.

ip msdp redistribute-msg [{ **list** *acl-name* | **list** *acl-number* } | **route-map** *route-map-name*] *

All (S, G) information registered on the local RP is released on the MSDP device by default.

1.6.6 Enabling the Function of Filtering Received SA Requests

1. Overview

An MSDP device gives a reply by default when the device receives an SA request. If you want to control the SA request and reply range, you can enable the SA request filtering function on the MSDP device to filter received SA requests.

2. Restrictions and Guidelines

- If the **list** keyword parameter is not specified, all the SA request messages are ignored.
- If the **list** keyword parameter is specified but this ACL is not configured, all SA request messages are ignored.
- If the **list** keyword parameter is specified and this ACL is configured, only the SA requests allowed by the ACL are accepted, and other SA requests are ignored.

3. Prerequisites

Before a received SA request is filtered, ensure that the corresponding ACL is created and a multicast group range is specified.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Filter received SA requests.

ip msdp filter-peer-sa-query *ipv4-peer-address* [{ **list** *acl-name* | **list** *acl-number* }]

All SA request messages from MSDP peers are received and replied by default.

1.6.7 Enabling the Function of Filtering Received SA Messages

1. Overview

By default, SA messages are received and forwarded to other MSDP peers after they pass RPF check. You can configure filtering rules for incoming SA message on the MSDP device to filter received SA messages.

2. Restrictions and Guidelines

- If the function of filtering SA messages is enabled but no ACL or route map is configured, all incoming SA messages are filtered.
- If only one keyword (**list** or **route-map**) is specified and the multicast source (S, G) information in an SA message meets the rule specified by this keyword, this SA message is received.
- If only one keyword (**originate-rp-list** or **originate-rp-route-map**) is specified and the RP address contained in an SA message meets the rule specified by this keyword, this SA message is received.
- If two or more of the keywords (including **list**, **route-map, rp-list** and **rp-route-map**) are specified and the multicast source (S, G) information contained in an SA message meets the rules specified by these keywords, this SA message is received.

3. Prerequisites

Before an IP multicast border is configured, ensure that the corresponding ACL or route map is created and an (S, G) range is specified.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Filter received SA messages.

ip msdp sa-msg-filter inbound *ipv4-peer-address* [{ **list** *acl-name* | **list** *acl-number* } | **route-map** *route-map-name* | { **originate-rp-list** *acl-name* | **originate-rp-list** *acl-number* } | **originate-rp-route-map** *rp-route-map-name*] *

All SA messages from MSDP peers are received by default.

1.6.8 Enabling the Function of Filtering Sent SA Messages

1. Overview

By default, SA messages are received and forwarded to other MSDP peers after they pass RPF check. You can configure outgoing SA message filtering rules on the MSDP device to filter SA messages to be forwarded.

2. Restrictions and Guidelines

- If the function of filtering sent SA messages is enabled but no ACL or route map is specified, no SA message is sent to other MSDP peers.
- If only one of the keywords (including **list**, **route-map**, **originate-rp-list**, and **originate-rp-route-map**) is specified and the multicast source (S, G) information in an SA message meets the rule specified by this keyword, this SA message is forwarded to other MSDP peers.
- If two or more of the keywords (including **list**, **route-map**, **originate-rp-list**, and **originate-rp-route-map**) are specified and the multicast source (S, G) information in an SA message meets the rules specified by these keywords, this SA message is forwarded to other MSDP peers.

3. Prerequisites

Before an IP multicast border is configured, ensure that the corresponding ACL or route map is created and an (S, G) range is specified.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Filter sent SA messages.

ip msdp sa-msg-filter outbound *ipv4-peer-address* [{ **list** *acl-name* | **list** *acl-number* } | **route-map** *route-map-name* | { **originate-rp-list** *acl-name* | **originate-rp-list** *acl-number* } | **originate-rp-route-map** *rp-route-map*] *

By default, all SA messages are forwarded to MSDP peers.

1.6.9 Configuring SA Cache Capacity

1. Overview

If an SA cache is used to cache SA messages, the (S, G) entries contained in the SA messages are cached in the SA cache. When a device receives a new join message, the device searches for the SA cache to reduce the delay time of obtaining the multicast information. The more (S, G) entries cached, the greater the used memory space. To suppress SA message flooding, you can limit the number of SA cache entries from a specified MSDP peer or a device.

2. Restrictions and Guidelines

- You are advised to configure the SA cache capacity when you start a device.
- During MSDP running,
  - if the capacity is increased, the adjustment does not affect the SA cache entries that are originally learned.
  - If the capacity is decreased, all SA cache entries learned and initiated must be deleted and re-learned.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Limit the number of SA cache entries from a specified MSDP peer.

ip msdp peer-sa-cache-limit *ipv4-peer-address sa-limit*

The number of SA cache entries from a specified MSDP peer is not limited by default.

(4) (Optional) Configure the SA cache capacity supported on a device.

ip msdp global-sa-limit *sa-limit*

The default SA cache capacity supported on a device is **1024**.

1.7 Configuring an Anycast RP

1.7.1 Overview

RPs in a PIM-SM domain use the same RP address to serve the same group. MSDP peer relationship is established between RPs so that multicast sources and receivers can register with and join the nearest RP. The RPs implement redundancy and load sharing and help accelerate convergence of multicast routes.

1.7.2 Prerequisites

Basic multicast functions are enabled in the PIM-SM domain and multiple RPs are configured for the domain.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Establish an MSDP peer relationship.

ip msdp add-peer *ipv4-peer-address* **source-interface** *interface-type interface-number*

No MSDP peer is added by default.

(4) Modify the RP address in SA messages.

ip msdp originator-inf *interface-type interface-number*

By default, no primary address of an interface is specified as the initiator address in SA messages.

1.8 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-2 Monitoring

Command	Purpose
cleanup ip msdp-conn peer-addr *ipv4-peer-address*	Resets the TCP connection with a specified MSDP peer.
cleanup ip msdp-conn sa-cache [*ipv4-group-address*]	Clears the SA cache.
cleanup ip msdp-conn peer-stats [*ipv4-peer-address*]	Clears the statistics of MSDP peers.
display ip msdp-info count [*as-number*]	Displays the number of sources and the number of groups generated by SA messages.
display ip msdp-info peer-to-mesh-group	Displays information of a mesh group.
display ip msdp-info peer-detail [*ipv4-peer-address*]	Displays detailed information about MSDP peers.

Command	Purpose
display ip msdp-info rpf-peer *ipv4-address*	Displays information about the MSDP RPF peer corresponding to the address of a specified initiator.
display ip msdp-info sa-cache [*ipv4-group-address* **|** *ipv4-source-address*] [*ipv4-group-address* **|** *ipv4-source-address*] [*as-number*]	Displays learned (S, G) information.
display ip msdp-info sa-originated	Displays (S, G) information initiated by the local device.
display ip msdp-info peer-summary	Displays summary information of all MSDP peers.
debugging ip msdp-info peer	Debugs MSDP peers.

1.9 Configuration Examples

1.9.1 Enabling Basic Functions of MSDP

1. Requirements

MSDP peers are configured in different domains to connect the RPs. Multicast source information can be shared to other RPs.

2. Topology

Figure 1-7 Topology for Basic Functions of MSDP

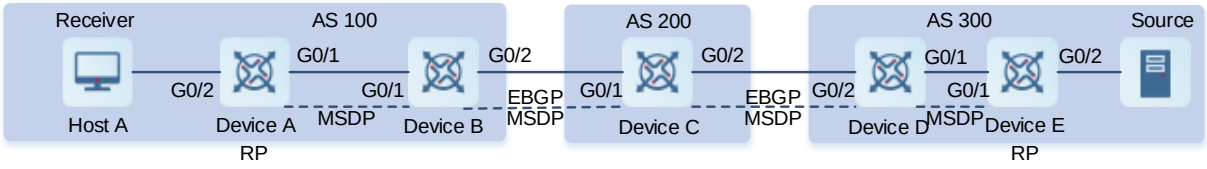


Table 1-3 Description of the Topology

Device	Interface	IP Address	Remarks
Device A	G0/1	10.110.1.1/24	-
	Loopback0	10.10.10.10/32	RP address, which is used to establish an MSDP connection.
Device B	G0/1	10.110.1.2/24	-
	G0/2	10.110.2.1/24	BSR border
	Loopback0	20.20.20.20/32	Used to establish the EBGP and MSDP connections.
Device C	G0/1	10.110.2.2/24	BSR border

Device	Interface	IP Address	Remarks
	G0/2	10.110.3.1/24	BSR border
	Loopback0	30.30.30.30/32	Used to establish the EBGP and MSDP connections.
Device D	G0/2	10.110.3.2/24	BSR border
	G0/1	10.110.4.1/24	-
	Loopback0	40.40.40.40/32	Used to establish the EBGP and MSDP connections.
Device E	G0/1	10.110.4.2/24	-
	Loopback0	50.50.50.50/32	RP address, which is used to establish an MSDP connection.

3. Notes

- Configure an IP address for an interface.
- Enable OSPF in each AS. Establish EBGP peer relationship between AS 200 and AS 100 and between AS 200 and AS 300. BGP and OSPF import learned routes to each other.
- Enable PIM-SM in each AS, configure C-BSRs and C-RPs, and configure the BSR border.
- Establish the MSDP peer relationship between EBGP peers and between the RP and EBGP peers.

Caution

The IP address and local interface of the MSDP peer must be the same as those of the EBGP peer.

4. Procedure

(1) Configure an IP address for each interface (omitted).

(2) Enable OSPF in each AS. Establish EBGP peer relationship between AS 200 and AS 100 and between AS 200 and AS 300. BGP and OSPF import learned routes to each other. For the omitted information, see *Configuring OSPFv2* and *Configuring BGP*

(3) Enable PIM-SM in each AS, configure C-BSRs and C-RPs, and configure the BSR border.

Configure device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ip mrouting-enable
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip pim sparse-mode
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface loopback 0
Device A(cfg-if-Loopback 0)# ip pim sparse-mode
Device A(cfg-if-Loopback 0)# exit
Device A(cfg)# ip pim c-rp loopback 0
Device A(cfg)# ip pim c-bsr loopback 0
```

Configure device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ip mrouting-enable
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip pim sparse-mode
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# ip pim sparse-mode
Device B(cfg-if-GigabitEthernet 0/2)# ip pim bsr-boundary
Device B(cfg-if-GigabitEthernet 0/2)# exit
Device B(cfg)# interface loopback 0
Device B(cfg-if-Loopback 0)# ip pim sparse-mode
Device B(cfg-if-Loopback 0)# exit
```

Configure device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# ip mrouting-enable
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# ip pim sparse-mode
Device C(cfg-if-GigabitEthernet 0/1)# ip pim bsr-boundary
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface gigabitethernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# ip pim sparse-mode
Device C(cfg-if-GigabitEthernet 0/2)# ip pim bsr-boundary
Device C(cfg-if-GigabitEthernet 0/2)# exit
Device C(cfg)# interface loopback 0
Device C(cfg-if-loopback 0)# ip pim sparse-mode
Device C(cfg-if-loopback 0)# exit
```

Configure device D.

```
Device D> privilege
Device D# config-system terminal
Device D(cfg)# ip mrouting-enable
Device D(cfg)# ip pim enable-ssm default
Device D(cfg)# interface gigabitethernet 0/1
Device D(cfg-if-GigabitEthernet 0/1)# ip pim sparse-mode
Device D(cfg-if-GigabitEthernet 0/1)# exit
Device D(cfg)# interface gigabitethernet 0/2
Device D(cfg-if-GigabitEthernet 0/2)# ip pim sparse-mode
Device D(cfg-if-GigabitEthernet 0/2)# ip pim bsr-boundary
Device D(cfg-if-GigabitEthernet 0/2)# exit
Device D(cfg)# interface loopback 0
Device D(cfg-if-Loopback 0)# ip pim sparse-mode
Device D(cfg-if-Loopback 0)# exit
```

Configure device E.

```
Device E> privilege
Device E# config-system terminal
Device E(cfg)# ip mrouting-enable
Device E(cfg)# interface gigabitethernet 0/1
Device E(cfg-if-GigabitEthernet 0/1)# ip pim sparse-mode
Device E(cfg-if-GigabitEthernet 0/1)# exit
Device E(cfg)# interface loopback 0
Device E(cfg-if-Loopback 0)# ip pim sparse-mode
Device E(cfg-if-Loopback 0)# exit
Device E(cfg)# ip pim c-rp loopback 0
Device E(cfg)# ip pim c-bsr loopback 0
```

(4) Establish the MSDP peer relationship between EBGP peers and between the RP and EBGP peers.

Configure device A.

```
Device A(cfg)# ip msdp add-peer 10.10.10.10 source-interface loopback 0
```

Configure device B.

```
Device B(cfg)# ip msdp add-peer 10.10.10.10 source-interface loopback 0
Device B(cfg)# ip msdp add-peer 30.30.30.30 source-interface loopback 0
```

Configure device C.

```
Device C(cfg)# ip msdp add-peer 20.20.20.20 source-interface loopback 0
Device C(cfg)# ip msdp add-peer 40.40.40.40 source-interface loopback 0
```

Configure device D.

```
Device D(cfg)# ip msdp add-peer 30.30.30.30 source-interface loopback 0
Device D(cfg)# ip msdp add-peer 50.50.50.50 source-interface loopback 0
```

Configure device E.

```
Device E(cfg)# ip msdp add-peer 50.50.50.50 source-interface loopback 0
```

5. Verification

(1) Use the multicast source to send a packet (200.200.200.200, 225.1.1.1). After a host joins the group 225.1.1.1, this host can receive this packet.

(2) Display the status and SA message of the MSDP peer on device D.

```
Device D# display ip msdp-info peer-summary
Msdp Peer Status Summary
Peer Address     As    State    Uptime/Downtime     Reset-Count    SA-Count    Peer-Description
30.30.30.30      200   Up       00:01:420                          1           No description

Device D# display ip msdp-info sa-cache
MSDP Source-Active Cache: 1 entries
  (200.200.200.200,225.1.1.1),RP:10.10.10.10,(M)BGP/AS 100, 00:00:18/00:01:57, Peer 30.30.30.30
Learned from peer 30.30.30.30, RPF peer 30.30.30.30,
SAs received: 1, Encapsulated data received: 1
```

6. Configuration Files

- Device A configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
!
interface Loopback 0
 ip pim sparse-mode
!
ip msdp add-peer 10.10.10.10 source-interface Loopback 0
!
ip pim c-bsr Loopback 0
ip pim c-rp Loopback 0
!
```

- Device B configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ip pim sparse-mode
 ip pim bsr-boundary
!
interface Loopback 0
 ip pim sparse-mode
!
ip msdp add-peer 10.10.10.10 source-interface Loopback 0
ip msdp add-peer 30.30.30.30 source-interface Loopback 0
!
```

- Device C configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
 ip pim bsr-boundary
!
interface GigabitEthernet 0/2
 unset switch-port
```

```
 ip pim sparse-mode
 ip pim bsr-boundary
!
interface Loopback 0
 ip pim sparse-mode
!
ip msdp add-peer 20.20.20.20 source-interface Loopback 0
ip msdp add-peer 40.40.40.40 source-interface Loopback 0
!
```

- Device D configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ip pim sparse-mode
 ip pim bsr-boundary
!
interface Loopback 0
 ip pim sparse-mode
!
ip pim enable-ssm default
!
ip msdp add-peer 30.30.30.30 source-interface Loopback 0
ip msdp add-peer 50.50.50.50 source-interface Loopback 0
!
```

- Device E configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
!
interface Loopback 0
 ip pim sparse-mode
!
ip msdp add-peer 50.50.50.50 source-interface Loopback 0
!
ip pim c-bsr Loopback 0
ip pim c-rp Loopback 0
```

```
!
```

7. Common Errors

- The BSR border is not configured, or is not configured on a correct interface.
- PIM-SM is not enabled on the local interface and the peer interface on which MSDP peer connection is established.
- SA messages fail the peer RPF check.

1.9.2 Configuring an Anycast RP

1. Requirements

There are multiple RPs in a PIM domain. After an anycast RP based on MSDP peers is configured, receivers can join the nearest RP, and the multicast source can register with the nearest RP to implement load sharing between the RPs.

2. Topology

Figure 1-8 Anycast RP Topology

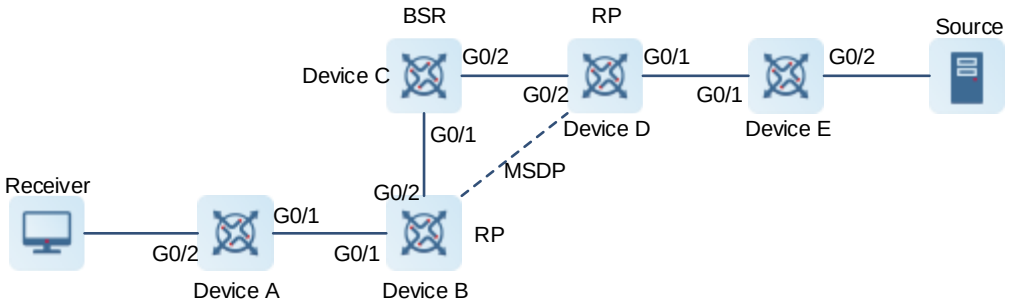


Table 1-4 Description of the Topology

Device	Interface	IP Address	Remarks
Device A	G0/1	10.110.1.1/24	
	G0/2	10.110.2.1/24	
Device B	G0/1	10.110.1.2/24	
	G0/2	10.110.3.1/24	
	Loopback0	10.10.10.10/32	A C-RP is configured.
	Loopback1	30.30.30.30/32	Used to establish an MSDP connection and modify the RP address in the SA message.
Device C	G0/1	10.110.3.2/24	
	G0/2	10.110.4.1/24	
	Loopback0	100.100.100.100/32	A C-BSR is configured.

Device	Interface	IP Address	Remarks
Device D	G0/1	10.110.5.1/24	
	G0/2	10.110.4.2/24	
	Loopback0	10.10.10.10/32	A C-RP is configured.
	Loopback1	20.20.20.20/32	Used to establish an MSDP connection and modify the RP address in the SA message.
Device E	G0/1	10.110.5.2/24	
	G0/2	10.110.6.1/24	

3. Notes

- Configure an IP address for each interface.
- Enable OSPF in each AS.
- Enable PIM-SM in each AS, and configure C-BSRs and C-RPs.
- Establish the MSDP peer relationship between RPs, and modify the RP address in the SA message.
- Configure a mesh group.

4. Procedure

(1) Configure an IP address for each interface (omitted).

(2) Enable OSPF in each AS (omitted). For the omitted information, see *Configuring OSPFv2*.

(3) Enable PIM-SM in each AS, and configure C-BSRs and C-RPs.

Configure device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ip mrouting-enable
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# ip pim sparse-mode
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface GigabitEthernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# ip pim sparse-mode
Device A(cfg-if-GigabitEthernet 0/2)# exit
```

Configure device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ip mrouting-enable
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# ip pim sparse-mode
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/2
```

```
Device B(cfg-if-GigabitEthernet 0/2)# ip pim sparse-mode
Device B(cfg-if-GigabitEthernet 0/2)# exit
Device B(cfg)# interface loopback 0
Device B(cfg-if-Loopback 0)# ip pim sparse-mode
Device B(cfg-if-Loopback 0)# exit
Device B(cfg)# interface loopback 1
Device B(cfg-if-Loopback 1)# ip pim sparse-mode
Device B(cfg-if-Loopback 1)# exit
Device B(cfg)# ip pim c-rp loopback 0
```

Configure device C.

```
Device C> privilege
Device C# config-system terminal
Device C(cfg)# ip mrouting-enable
Device C(cfg)# interface gigabitethernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# ip pim sparse-mode
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface gigabitethernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# ip pim sparse-mode
Device C(cfg-if-GigabitEthernet 0/2)# exit
Device C(cfg)# interface loopback 0
Device C(cfg-if-Loopback 0)# ip pim sparse-mode
Device C(cfg-if-Loopback 0)# exit
Device C(cfg)# ip pim c-bsr loopback0
```

Configure device D.

```
Device D> privilege
Device D# config-system terminal
Device D(cfg)# ip mrouting-enable
Device D(cfg)# interface gigabitethernet 0/1
Device D(cfg-if-GigabitEthernet 0/1)# ip pim sparse-mode
Device D(cfg-if-GigabitEthernet 0/1)# exit
Device D(cfg)# interface gigabitethernet 0/2
Device D(cfg-if-GigabitEthernet 0/2)# ip pim sparse-mode
Device D(cfg-if-GigabitEthernet 0/2)# exit
Device D(cfg)# interface loopback 0
Device D(cfg-if-Loopback 0)# ip pim sparse-mode
Device D(cfg-if-Loopback 0)# exit
Device D(cfg)# interface loopback 1
Device D(cfg-if-Loopback 1)# ip pim sparse-mode
Device D(cfg-if-Loopback 1)# exit
Device D(cfg)# ip pim c-rp loopback 0
```

Configure device E.

```
Device E> privilege
Device E#config-system terminal
Device E(cfg)#ip mrouting-enable
Device E(cfg)#interface gigabitethernet 0/1
```

```
Device E(cfg-if-GigabitEthernet 0/1)#ip pim sparse-mode
Device E(cfg-if-GigabitEthernet 0/1)# exit
Device E(cfg)#interface gigabitethernet 0/2
Device E(cfg-if-GigabitEthernet 0/2)#ip pim sparse-mode
Device E(cfg-if-GigabitEthernet 0/2)# exit
```

(4) Establish the MSDP peer relationship between RPs, and modify the RP address in the SA message.

Configure device D.

```
Device D(cfg)# ip msdp add-peer 30.30.30.30 source-interface loopback 1
Device D(cfg)# ip msdp originator-inf loopback 1
```

Configure device B.

```
Device B(cfg)# ip msdp add-peer 20.20.20.20 source-interface loopback 1
Device B(cfg)# ip msdp originator-inf loopback 1
```

(5) Configure a mesh group.

Configure device D.

```
Device D(cfg)# ip msdp peer-to-mesh-group mesh-name 30.30.30.30
```

Configure device B.

```
Device B(cfg)# ip msdp peer-to-mesh-group mesh-name 20.20.20.20
```

5. Verification

(1) Use the multicast source to send a packet (6.6.6.6, 225.1.1.1). After a host joins the group 225.1.1.1, this host can receive this packet.

(2) Display the status and SA message of the MSDP peer on device B.

Run the **display ip msdp-info summary** command to display status of the peers.

```
Device B# display ip msdp-info peer-summary
Msdp Peer Status Summary
Peer Address     As          State   Uptime/Downtime   Reset-Count   SA-Count   Peer-Description
20.20.20.20      Unknown     Up        00:01:420                     1          No description
```

Run the **display ip msdp-info sa-cache** command to display SA messages of the peers.

```
Device B# display ip msdp-info sa-cache
MSDP Source-Active Cache: 1 entries
  (6.6.6.6,225.1.1.1),RP:10.10.10.10,(M)BGP/AS unknown, 00:00:18/00:01:57, Peer 20.20.20.20
```

6. Configuration Files

- Device A configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
!
interface GigabitEthernet 0/2
```

```
 unset switch-port
 ip pim sparse-mode
!
```

- Device B configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ip pim sparse-mode
!
interface Loopback 0
 ip pim sparse-mode
!
interface Loopback 1
 ip pim sparse-mode
!
ip msdp originator-inf Loopback 1
ip msdp add-peer 20.20.20.20 source-interface Loopback 1
ip msdp peer-to-mesh-group mesh-name 20.20.20.20
!
ip pim c-rp Loopback 0
!
```

- Device C configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ip pim sparse-mode
!
interface Loopback 0
 ip pim sparse-mode
!
ip pim c-bsr Loopback 0
!
```

- Device D configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ip pim sparse-mode
!
interface Loopback 0
 ip pim sparse-mode
!
interface Loopback 1
 ip pim sparse-mode
!
ip msdp originator-inf Loopback 1
ip msdp add-peer 30.30.30.30 source-interface Loopback 1
ip msdp peer-to-mesh-group mesh-name 30.30.30.30
!
ip pim c-rp Loopback 0
!
```

- Device E configuration file

```
!
ip mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ip pim sparse-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ip pim sparse-mode
!
```

7. Common Errors

- The C-BSRs and C-RPs are configured on the same interface.
- The RP address in the SA message is not modified.
- SA messages cannot pass the peer RPF check.

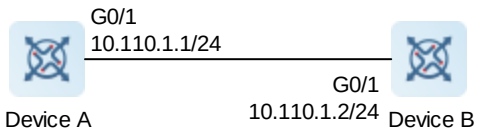
1.9.3 Enabling MD5 Authentication

1. Requirements

MD5 authentication is enabled on TCP connections between MSDP peers, and the number of SA cache entries from a specified MSDP peer is limited to prevent invalid TCP connections and improve MSDP security.

2. Topology

Figure 1-9 Topology of MD5 Authentication



3. Notes

- Establish the MSDP peer relationship between device A and device B.
- Enable MD5 authentication for the peer on device A.
- After MSDP timeout, configure the MD5 cipher of the peer on device B, which is the same as the cipher on device A. Then, the session is reconnected.
- Set the number of SA cache entries from the peer 10.110.1.2 to 10.

4. Procedure

(1) Configure IP addresses on interfaces of device A and device B, and enable basic functions of PIM-SM on the devices (omitted).

(2) Establish the MSDP peer relationship between device A and device B.

Add device B as an MSDP peer on device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ip msdp add-peer 10.110.1.2 source-interface gigabitEthernet 0/1
```

Add device A as an MSDP peer on device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ip msdp add-peer 10.110.1.1 source-interface gigabitEthernet 0/1
```

(3) Enable MD5 authentication for the peer on device A, and set the encryption level to 0 and the cypher to password123.

```
Device A(cfg)# ip msdp peer-connection password 10.110.1.2 0 password123
```

(4) After MSDP timeout, enable MD5 authentication for the peer on device B, and set the encryption level and cypher to be consistent with those on device A.

```
Device B(cfg)# ip msdp peer-connection password 10.110.1.1 0 password123
```

(5) Set the number of SA cache entries from the peer 10.110.1.2 to 10.

```
Device A(cfg)# ip msdp peer-sa-cahce-limit 10.110.1.2 10
```

5. Verification

(1) After the MSDP peer relationship is established between device A and device B, the peer status is up.

Check whether the status of the peer 10.110.1.2 is up on device A.

```
Device A# display ip msdp-info peer-detail
```

```
MSDP PEER 10.110.1.2 (No description), AS unknown
  Connection status:
    State: Up, Resets: 0, Connection source: GigabitEthernet 0/7 (10.110.1.1)
    Uptime(Downtime): 00:01:14, Message sent/received: 2/2
    Input messages discarded: 0
    Connection and counters cleared 00:01:44 ago
    Local Address of connection: 10.110.1.1
  SA Filtering:
    Input (S,G) Access-control-list filter: None
    Input (S,G) route-map filter: None
    Input RP Access-control-list filter: None
    Input RP Route-map filter: None
    Output (S,G) Access-control-list filter: None
    Output (S,G) Route-map filter: None
    Output RP Access-control-list filter: None
    Output RP Route-map filter: None
  SA-Requests:
    Input filter: None
  Peer ttl threshold: 0
  SAs learned from this peer: 0, SAs limit: No-limit
  Message counters:
    SA messages discarded: 0
    SA messages in/out: 0/0
    SA Requests discarded/in: 0/0
    SA Responses out: 0
    Data Packets in/out: 0/0
```

Check whether the status of the peer 10.110.1.1 is up on device B.

```
Device B# display ip msdp-info peer-detail
MSDP PEER 10.110.1.1 (No description), AS unknown
  Connection status:
    State: Up, Resets: 0, Connection source: GigabitEthernet 0/1 (10.110.1.2)
    Uptime(Downtime): 00:00:05, Message sent/received: 1/1
    Input messages discarded: 0
    Connection and counters cleared 00:00:14 ago
    Local Address of connection: 10.110.1.2
  SA Filtering:
    Input (S,G) Access-control-list filter: None
    Input (S,G) route-map filter: None
    Input RP Access-control-list filter: None
    Input RP Route-map filter: None
    Output (S,G) Access-control-list filter: None
    Output (S,G) Route-map filter: None
    Output RP Access-control-list filter: None
    Output RP Route-map filter: None
  SA-Requests:
    Input filter: None
```

```
Peer ttl threshold: 0
SAs learned from this peer: 0, SAs limit: No-limit
Message counters:
  SA messages discarded: 0
  SA messages in/out: 0/0
  SA Requests discarded/in: 0/0
```

(2) After MD5 authentication is enabled for the peer on device A, run the **display ip msdp-info peer-detail** command to check whether the peer relationship between device A and device B is disconnected and the peer is down.

(3) After MD5 authentication is enabled for the peer on device B, run the **display ip msdp-info peer-detail** command to check whether the peer relationship between device A and device B is established and the peer is up again.

(4) Set the number of SA cache entries from the peer 10.110.1.2 to 10 on device A, and send 20 (S, G) entries to device B. A message is displayed on device A, indicating that the number of entries exceeds the limit.

6. Configuration Files

- Device A configuration file

```
!
ip msdp add-peer 10.110.1.2 source-interface gigabitEthernet 0/1
ip msdp peer-connection password 10.110.1.2 0 password123
ip msdp peer-sa-cache-limit 10.110.1.2 10
!
```

- Device B configuration file

```
!
ip msdp add-peer 10.110.1.1 source-interface gigabitEthernet 0/1
ip msdp peer-connection password 10.110.1.1 0 password123
!
```

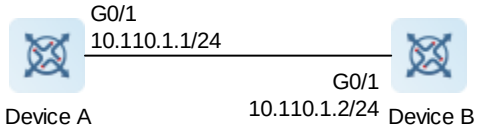
1.9.4 Managing MSDP Peers

1. Requirements

MSDP peer information is managed and configured, including configuring peer description information and disabling peer connections.

2. Topology

Figure 1-10 Topology of MSDP Peer Management



3. Notes

- Establish MSDP peer relationship between device A and device B.

- On device A, configure the peer description of device B as "peer-router-B".
- After 60 seconds, on device A, disable the peer device B.

4. Procedure

(1) Configure IP addresses on interfaces of device A and device B, and enable basic functions of PIM-SM on the devices (omitted).

(2) Establish the MSDP peer relationship between device A and device B.

Add device B as an MSDP peer on device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# ip msdp add-peer 10.110.1.2 source-interface gigabitEthernet 0/1
```

Add device A as an MSDP peer on device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ip msdp add-peer 10.110.1.1 source-interface gigabitEthernet 0/1
```

(3) On device A, configure the peer description of device B as "peer-router-B".

```
Device A(cfg)# ip msdp peer-description 10.110.1.2 peer-router-B
```

(4) After 60 seconds, on device A, disable the peer device B.

```
Device A(cfg)# ip msdp peer-closing 10.110.1.2
```

5. Verification

Run the **display ip msdp-info peer-detail** command to display the summary information of a specified peer, including the description and connection status of this MSDP peer.

```
Device A# display ip msdp-info peer-detail 10.110.1.2
MSDP PEER 10.110.1.2 (peer-router-B), AS unknown
  Connection status:
    State: Up, Resets: 3, Connection source: GigabitEthernet 0/1 (10.110.1.1)
    Uptime(Downtime): 00:30:39, Message sent/received: 50/50
    Input messages discarded: 0
    Connection and counters cleared 00:51:06 ago
    Local Address of connection: 10.110.1.1
  SA Filtering:
    Input (S,G) Access-control-list filter: None
    Input (S,G) route-map filter: None
    Input RP Access-control-list filter: None
    Input RP Route-map filter: None
    Output (S,G) Access-control-list filter: None
    Output (S,G) Route-map filter: None
    Output RP Access-control-list filter: None
    Output RP Route-map filter: None
  SA-Requests:
    Input filter: None
  Peer ttl threshold: 0
```

```
SAs learned from this peer: 0, SAs limit: No-limit
Message counters:
  SA messages discarded: 0
  SA messages in/out: 0/0
  SA Requests discarded/in: 0/0
  SA Responses out: 0
  Data Packets in/out: 0/0
```

6. Configuration Files

- Device A configuration file

```
!
ip msdp add-peer 10.110.1.2 source-interface gigabitEthernet 0/1
ip msdp peer-closing 10.110.1.2
ip msdp peer-description 10.110.1.2 peer-router-B
!
```

- Device B configuration file

```
!
ip msdp add-peer 10.110.1.1 source-interface gigabitEthernet 0/1
!
```

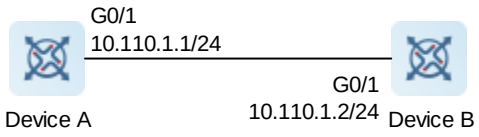
1.9.5 Modifying Protocol Parameters

1. Requirements

In some application scenarios, for example, MSDP peer creation, restart, or failback, you can shorten the TCP reconnection interval to accelerate convergence of the MSDP peer relationship.

2. Topology

Figure 1-11 Topology for Protocol Parameter Modification



3. Notes

- Establish the MSDP peer relationship between device A and device B.
- Set the MSDP peer reconnection interval to 20 seconds on device A.

4. Procedure

(1) Configure IP addresses on interfaces of device A and device B, and enable basic functions of PIM-SM on the devices (omitted).

(2) Establish the MSDP peer relationship between device A and device B.

Add device B as an MSDP peer on device A.

```
Device A> privilege
```

```
Device A# config-system terminal
Device A(cfg)# ip msdp add-peer 10.110.1.2 source-interface gigabitEthernet 0/1
```

Add device A as an MSDP peer on device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# ip msdp add-peer 10.110.1.1 source-interface gigabitEthernet 0/1
```

(3) Set the MSDP peer reconnection interval to 20 seconds on device A.

```
Device(cfg)# ip msdp retry-period 20
```

5. Verification

(1) Disable the MSDP peer on device B, and enable the MSDP peer again immediately.

```
Device B(cfg)# ip msdp peer-closing 10.110.1.1
Device B(cfg)# unset ip msdp peer-closing 10.110.1.1
```

(2) Check whether the peer on device B can be reconnected in 20 seconds.

```
Device B# display ip msdp-info peer-detail 10.110.1.1
MSDP PEER 10.110.1.1 (peer-a), AS unknown
  Connection status:
    State: Up, Resets: 3, Connection source: GigabitEthernet 0/1 (10.110.1.2)
    Uptime(Downtime): 00:00:41, Message sent/received: 20/20
    Input messages discarded: 0
    Connection and counters cleared 01:37:02 ago
    Local Address of connection: 10.110.1.2
  SA Filtering:
    Input (S,G) Access-control-list filter: None
    Input (S,G) route-map filter: None
    Input RP Access-control-list filter: None
    Input RP Route-map filter: None
    Output (S,G) Access-control-list filter: None
    Output (S,G) Route-map filter: None
    Output RP Access-control-list filter: None
    Output RP Route-map filter: None
  SA-Requests:
    Input filter: None
  Peer ttl threshold: 0
  SAs learned from this peer: 0, SAs limit: No-limit
  Message counters:
    SA messages discarded: 0
    SA messages in/out: 0/0
    SA Requests discarded/in: 0/0
    SA Responses out: 0
    Data Packets in/out: 0/0
```

6. Configuration Files

- Device A configuration file

```
!
ip msdp retry-period 20
ip msdp add-peer 10.110.1.2 source-interface gigabitEthernet 0/1
!
```

- Device B configuration file

```
!
ip msdp add-peer 10.110.1.1 source-interface gigabitEthernet 0/1
!
```

Contents

1 Configuring IPv6 Multicast Route Management 1
1.1 Introduction 1
1.1.1 IPv6 Multicast Routing and Forwarding 1
1.1.2 RPF Check 1
1.1.3 Protocols and Standards 4
1.2 Configuration Task Summary 4
1.3 Prerequisites 4
1.4 Configuring Basic Functions of IPv6 Multicast 4
1.4.1 Overview 4
1.4.2 Restrictions and Guidelines 4
1.4.3 Prerequisites 5
1.4.4 Procedure 5
1.5 Configuring the Maximum Number of Entries in an IPv6 Multicast Routing Table 5
1.5.1 Overview 5
1.5.2 Restrictions and Guidelines 6
1.5.3 Prerequisites 6
1.5.4 Procedure 6
1.6 Configuring an IPv6 Multicast Border 6
1.6.1 Overview 6
1.6.2 Restrictions and Guidelines 6
1.6.3 Prerequisites 6
1.6.4 Procedure 6

1.7 Configuring an IPv6 Static Multicast Route 7
1.7.1 Overview 7
1.7.2 Restrictions and Guidelines 7
1.7.3 Prerequisites 7
1.7.4 Procedure 7
1.8 Enabling Forced Forwarding of Multicast Packets by Software 8
1.8.1 Overview 8
1.8.2 Restrictions and Guidelines 8
1.8.3 Prerequisites 8
1.8.4 Procedure 8
1.9 Enabling L2 Direction Control for Multicast Streams 8
1.9.1 Overview 8
1.9.2 Restrictions and Guidelines 8
1.9.3 Prerequisites 8
1.9.4 Procedure 8
1.10 Configuring RPF Route Selection Based on the Longest Match Rule 9
1.10.1 Overview 9
1.10.2 Restrictions and Guidelines 9
1.10.3 Prerequisites 9
1.10.4 Procedure 9
1.11 Monitoring 9
1.12 Configuration Examples 11
1.12.1 Configuring an IPv6 Static Multicast Route to Modify the RPF Route 11
1.12.2 Configuring a Static Multicast Route to Connect the RPF Route 18

1 Configuring IPv6 Multicast Route Management

1.1 Introduction

Multicast route management covers how to create and modify multicast routes to control multicast packet forwarding and how to detect and maintain correctness of multicast forwarding paths. Multicast route management mainly includes the static multicast routing and reverse path forwarding (RPF) check features. Static multicast routing is used to modify or connect RPF routes, and RPF check is used to ensure that multicast data is forwarded on correct paths.

1.1.1 IPv6 Multicast Routing and Forwarding

An IPv6 multicast routing table includes MLD group entries, multicast protocol routing entries, multicast routing entries, and multicast hardware routing entries. These entries are maintained by different multicast routing protocols in the multicast network. The optimal IPv6 multicast routes are delivered to the hardware routing table and used to guide IPv6 multicast data forwarding. The RPF mechanism ensures that multicast data is forwarded on correct paths.

- MLD group entries: They are created when a host sends MLD join packets and used to maintain group member information and ask the multicast routing protocol to create (S, G) or (*, G) entries.
- Multicast protocol routing entries: They function as basis of multicast routing and forwarding and are maintained by different multicast routing protocols (the most widely applied protocol is PIM).
- Multicast routing entries: They are generated by the multicast routing management module and delivered to a multicast hardware routing table. Different multicast routing protocols can be summarized based on the multicast routing entries to select an optimal route.
- Hardware routing entries: They are generated based on routing tables and used to guide multicast data forwarding.

1.1.2 RPF Check

1. Basic Principles

In unicast routing and forwarding, a device gets to know the outbound interface of packets based on the destination address. In multicast routing and forwarding, the destination address in multicast packets is a multicast address, based on which, a receiver address cannot be determined. However, the source address in multicast packets is determined and based on this source address, multicast packets can be forwarded on correct paths.

When receiving a multicast packet, a multicast device queries the unicast route of the multicast source based on the source address in the multicast packet and checks whether the outbound interface of this unicast route is consistent with the inbound interface of the multicast packet. If yes, it is considered that this multicast packet is forwarded on a correct path. The inbound interface of the multicast packet is referred to as an RPF interface.

2. RPF Check Procedure

Unicast route, MBGP route, and static multicast route are the basis of RPF check. If these three types of routing entries exist on a multicast device, the multicast device performs the following checks in order after receiving a multicast packet:

(1) Select an optimal route from each of the three types of routing entries, respectively, based on the source address in the packet.

- The outbound interface of a unicast route is an RPF interface and the next hop is an RPF neighbor.
- The outbound interface of an MBGP route is an RPF interface and the next hop is an RPF neighbor.
- The static multicast route is manually configured and an RPF interface and neighbor are specified.

(2) Select one route out of the three optimal routes as the RPF route. The selection rule is as follows:

- If the longest match rule is not configured for selecting an RPF route, the one with the highest priority is selected as the RPF route. If the three routes have the same priority, the RPF route is selected in the order of static multicast route, MBGP route, and unicast route.
- If the longest match rule is configured for selecting an RPF route, the one with the longest match is selected as the RPF route. If the three routes have the same mask, the one with the highest priority is selected as the RPF route. If they have the same priority, the RPF route is selected in the order of static multicast route, MBGP route, and unicast route.

(3) Compare the RPF interface of the RPF route with the inbound interface of the multicast packet.

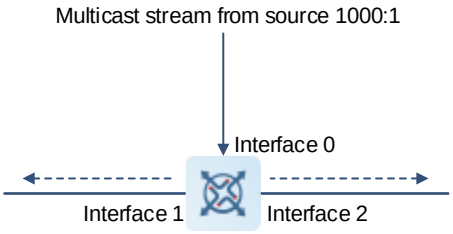
- If they are consistent, the route passes the RPF check and the multicast packet is forwarded to the downlink devices.
- If they are inconsistent, the route fails the RPF check and the received multicast packet is discarded.

Table 1-1 Routing Table on a Multicast Device

Network	Interface
1000::/64	Interface 0
2000::/64	Interface 1
3000::/64	Interface 2

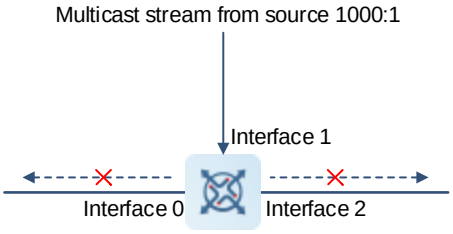
As shown in Figure 1-1, the multicast device receives a multicast packet from the source 1000::1 through interface 0 and as listed in Table 1-1, the outbound interface of the route to the source 1000::1 is also interface 0. The outbound interface of the route is consistent with the inbound interface of the multicast packet. The route passes the RPF check and the multicast packet is forwarded on the correct path.

Figure 1-1 Topology of RPF Check Success



As shown in Figure 1-2, the multicast device receives a multicast packet from the source 1000::1 through interface 1 and the outbound interface of the route to the source 1000::1 is Interface 0. The outbound interface of the route is inconsistent with the inbound interface of the multicast packet. The route fails the RPF check and the multicast packet is discarded.

Figure 1-2 Topology of RPF Check Failure



3. Application Scenarios of RPF Check in Multicast Forwarding

A multicast device ensures that multicast packets are forwarded on correct paths and avoids loops on the forwarding paths based on the RPF check mechanism. In actual applications, if the multicast device performs RPF check each time it receives a multicast packet, the multicast device is overloaded. Therefore, upon receiving a multicast packet, the multicast device checks for corresponding (S, G) hardware routing entry, rather than performs RPF check. The checking process is as follows:

- If no (S, G) hardware routing entry exists, the multicast device performs RPF check, uses the RPF interface as an inbound interface, creates a multicast routing entry, and delivers the entry to the multicast forwarding table.
  - If the route passes the RPF check, the inbound interface of the multicast packet is consistent with the RPF interface. The multicast packet is forwarded to all the outbound interfaces in the forwarding table.
  - If the route fails the RPF check, the inbound interface of the multicast packet is inconsistent with the RPF interface. The multicast packet is discarded.
- If (S, G) hardware routing entries exist, the multicast device checks whether the inbound interface in the forwarding entry is consistent with the inbound interface of the multicast packet.
  - If they are consistent, the multicast packet is forwarded to all the outbound interfaces.
  - If they are inconsistent, the multicast device performs RPF check. If the RPF interface is consistent with the inbound interface in the hardware routing entry, the source path of the packet is incorrect but the (S, G) entry is correct. The multicast packet is discarded. If the RPF interface is inconsistent with the inbound interface in the hardware routing entry, the (S, G) entry expired. The inbound interface in the

entry is updated as an RPF interface. Then, the multicast device checks whether the RPF interface is consistent with the inbound interface of the packet. If they are consistent, this multicast packet is forwarded to all outbound interfaces in the forwarding table. Otherwise, this multicast packet is discarded.

1.1.3 Protocols and Standards

- RFC 2710: Multicast Listener Discovery (MLDv1) for IPv6
- RFC 3810: Multicast Listener Discovery Version 2 (MLDv2) for IPv6
- RFC 4601: Protocol Independent Multicast-Sparse Mode (PIM-SM)
- RFC 5059: Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)
- RFC 4607: Source-Specific Multicast for IP

1.2 Configuration Task Summary

The IPv6 multicast route management configuration includes the following tasks:

(1) Configuring Basic Functions of IPv6 Multicast

(2) (Optional) Configuring the Maximum Number of Entries in an IPv6 Multicast Routing Table

(3) (Optional) Configuring an IPv6 Multicast Border

(4) (Optional) Configuring an IPv6 Static Multicast Route

(5) (Optional) Enabling Forced Forwarding of Multicast Packets by Software

(6) (Optional) Enabling L2 Direction Control for Multicast Streams

(7) (Optional) Configuring RPF Route Selection Based on the Longest Match Rule

1.3 Prerequisites

- In the multicast network, all nodes are reachable via IPv6 unicast routes.
- Configuring Basic Functions of IPv6 Multicast.

1.4 Configuring Basic Functions of IPv6 Multicast

1.4.1 Overview

The multicast routing function is enabled in the local network and a PIM network is created to provide data sources and user hosts in the network with the IPv6 multicast service.

A device maintains the routing table used for multicast packet forwarding over an IPv6 multicast routing protocol (for example, PIM-SMv6), and learns information about the status of group members in the directly-connected network segments over the MLDv1/v2 protocol. A host joins a specific IPv6 multicast group by sending an MLD REPORT message.

1.4.2 Restrictions and Guidelines

- The IPv6 multicast routing function should be enabled on each multicast device.
- The IPv6 multicast routing function is enabled prior to different IPv6 multicast protocols.
- The IPv6 multicast routing function and the MLD Snooping function are mutually exclusive.

- The IPv6 multicast protocol function must be enabled on each interface in the multicast network.
- After L3 multicast is enabled in the private VLAN and super VLAN and a multicast source exists in a sub VLAN, an extra entry whose ingress is the sub VLAN into which the multicast stream enters must be copied due to the validity check during multicast forwarding. This increases a multicast hardware entry and decreases the multicast capacity by one.

1.4.3 Prerequisites

In the multicast network, all nodes are reachable via IPv6 unicast routes.

1.4.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IPv6 multicast routing function.

ipv6 mrouting-enable

The IPv6 multicast routing function is disabled by default.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(5) Configure an IPv6 multicast protocol.

ipv6 pim dense-mode

No IPv6 multicast protocol is enabled by default.

1.5 Configuring the Maximum Number of Entries in an IPv6 Multicast Routing Table

1.5.1 Overview

Every multicast data packet received by a device is used to maintain relevant IPv6 multicast routing entries. However, excessive multicast routing entries may exhaust device memory and cause device performance to decrease. Users can restrict the number of entries in the IPv6 multicast routing table based on the actual networking conditions and service performance requirements.

1.5.2 Restrictions and Guidelines

Due to limitations by hardware resources, routing entries that exceed the range permitted by hardware can be forwarded only by software, which causes performance to decrease.

1.5.3 Prerequisites

Configuring Basic Functions of IPv6 Multicast.

1.5.4 Procedure

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of entries in an IPv6 multicast routing table.

ipv6 multicast max-route *limit* [*threshold*]

By default, a maximum of 4000 entries can be added to an IPv6 multicast routing table.

The configured value of *threshold* must be smaller than or equal to the configured value of *limit*.

1.6 Configuring an IPv6 Multicast Border

1.6.1 Overview

An IPv6 multicast border is configured to define the transmission range of multicast packets. When a multicast border is configured on an interface, this interface cannot forward or receive multicast packets, including those sent from the local host.

1.6.2 Restrictions and Guidelines

- The ACL referenced in this command can be a standard ACL or an extended ACL. If an extended ACL is used, only destination addresses need to be matched.
- This command can be used to filter MLD and PIM-SMv6 protocol packets relevant to the IPv6 multicast group range. Multicast data streams are not sent or received by this multicast border interface.

1.6.3 Prerequisites

- Configuring Basic Functions of IPv6 Multicast.
- Before an IPv6 multicast border is configured, ensure that the corresponding ACL is created and a multicast group range is specified. For details about ACL configuration, see *Configuring ACL*.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  interface *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Enable the IPv6 multicast routing function.

ipv6 multicast scope-boundary *acl-name* [**in** | **out**]

No IPv6 multicast border is configured by default.

1.7 Configuring an IPv6 Static Multicast Route

1.7.1 Overview

An IPv6 static multicast route is configured to specify an RPF interface or RPF neighbor for multicast packets from a specific multicast source. This function is used to modify and connect RPF routes.

- Modify an RPF route. A device expects to receive multicast packets from a source through a specified interface. However, the RPF check result indicates that the RPF interface is not this specified interface. In this case, you need to configure an IPv6 static multicast route to specify this interface as an RPF interface. Upon receiving multicast packets from the source, the device performs RPF check. If the RPF check result indicates that the RPF interface is not this specified interface, the route fails the RPF check.
- Connect an RPF route. If the unicast route is interrupted, the RPF routing fails. In this case, you can configure a static multicast route to specify the RPF interface to complete the RPF check and implement multicast packet forwarding.

1.7.2 Restrictions and Guidelines

To specify an outbound interface rather than a next-hop IP address for the IPv6 static multicast route, ensure that the outbound interface is of the point-to-point type.

1.7.3 Prerequisites

Configuring Basic Functions of IPv6 Multicast.

1.7.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an IPv6 static multicast route.

ipv6 rpf-mroute *ipv6-address/prefix-length* [*protocol*] { *ipv6-rpf-addrress* | *interface-type interface-number* } [*distance*]

No IPv6 static multicast route is configured by default.

1.8 Enabling Forced Forwarding of Multicast Packets by Software

1.8.1 Overview

After this function is enabled, all IPv6 multicast data packets destined for the CPU are forcedly forwarded by software.

1.8.2 Restrictions and Guidelines

The cache mechanism and forwarding performance problem may cause packet disorder if the multicast streams are heavy. You are not advised to run this command generally.

1.8.3 Prerequisites

Configuring Basic Functions of IPv6 Multicast.

1.8.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of forced forwarding of multicast packets by software.

msf6 force-forwarding

The function of forced forwarding of IPv6 multicast packets (destined for the CPU) by software is disabled by default.

1.9 Enabling L2 Direction Control for Multicast Streams

1.9.1 Overview

L2 direction control is enabled for multicast streams to control the forwarding of multicast streams on an interface. Multiple commands can be configured for a specified multicast stream so that the stream can be forwarded by multiple interfaces. After direction control is enabled for a multicast stream, the stream can be forwarded only by these configured interfaces. Other interfaces are not permitted to forward the stream.

1.9.2 Restrictions and Guidelines

This function controls only the forwarding of multicast streams on interfaces, but does not directly affect the processing of multicast protocols on the protocol packets. Some features (for example, establishment of L3 hardware routing entry) of multicast protocols (such as PIM-SMv6) are driven by multicast data streams, and therefore, the behavior of the multicast routing protocols may still be affected.

1.9.3 Prerequisites

Configuring Basic Functions of IPv6 Multicast.

1.9.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable L2 direction control for multicast streams.

ipv6 multicast static *ipv6-source-address ipv6-group-address interface-type interface-number*

L2 direction control is disabled for multicast streams by default.

1.10 Configuring RPF Route Selection Based on the Longest Match Rule

1.10.1 Overview

In multicast routing, various routing entries, such as static multicast routing entry, MBGP routing entry and unicast routing entry, may exist. If RPF route selection based on the longest match rule is configured, the route with the longest match mask is selected from the optimal routes of multiple routing entries as the RPF route. If the routes have the same mask length, the route with the highest priority is selected from the optimal routes. If they have the same priority, the RPF route is selected in the order of static multicast route, MBGP route, and unicast route.

1.10.2 Restrictions and Guidelines

It is recommended that RPF route selection based on the longest match rule be configured on each device unless otherwise specified.

1.10.3 Prerequisites

Configuring Basic Functions of IPv6 Multicast.

1.10.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure RPF route selection based on the longest match rule.

ipv6 multicast rpf-routing longest-match

By default, the route with the highest priority is selected as the RPF route. If the routes have the same priority, the RPF route is selected in the order of IPv6 static multicast route, IPv6 MBGP route and IPv6 unicast route.

1.11 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-2 Monitoring

Command	Purpose
cleanup ipv6 mrouting-table { *ipv6-group-address* [*ipv6-source–address*] | ***** }	Clears the IPv6 hardware routing entry.
cleanup ipv6 mrouting-table statistics { *ipv6-group-address* [*ipv6-source-address*] | ***** }	Clears the statistics about the IPv6 hardware routing entry.
display ipv6 mrouting-table [*ipv6-group-or-source-address* [*ipv6-group-or-source-address*]] [**sparse**] [**summary** | **count**]	Displays the IPv6 hardware routing entry.
display ipv6 rpf-routing *ipv6-source-address*	Displays the RPF information of a specified IPv6 source address.
display ipv6 mrouting-table static-route	Displays the IPv6 static multicast route information.
display ipv6 mvif [*interface-type interface-number*]	Displays information about the configured IPv6 multicast interface that takes effect.
display ipv6 mrf mfc	Displays the IPv6 L3 hardware routing entry.
display msf6 msc	Displays the IPv6 L2 and L3 hardware routing entry.
debugging nsm mcast6 all	Displays the running procedure of the multicast core.
debugging nsm mcast6 fib-msg	Displays the communication procedure between the IPv6 multicast core and protocol module.
debugging nsm mcast6 mif	Displays the interface running procedure of the IPv6 multicast core.
debugging nsm mcast6 stats	Displays the procedure for processing interfaces and behavior statistics of the IPv6 multicast core.
debugging ipv6 mrf6 forwarding	Displays the procedure for IPv6 L3 multicast forwarding.

Command	Purpose
debugging ipv6 mrf6 mfc	Displays the procedure for operating IPv6 L3 hardware routing entries.
debugging ipv6 mrf6 event	Displays the procedure for processing IPv6 L3 multicast forwarding events.
debugging msf6 forwarding	Displays the procedure for IPv6 L2 and L3 multicast packet forwarding.
debugging msf6 msc	Displays the procedure for operating IPv6 L2 and L3 hardware routing entries.
debugging msf6 ssp	Displays the bottom-layer hardware processing procedure of IPv6 L2 and L3 multicast forwarding.
debugging msf6 api	Displays the procedure for invoking API interfaces for IPv6 L2 and L3 multicast forwarding.
debugging msf6 event	Displays the procedure for processing IPv6 L2 and L3 multicast forwarding events.

1.12 Configuration Examples

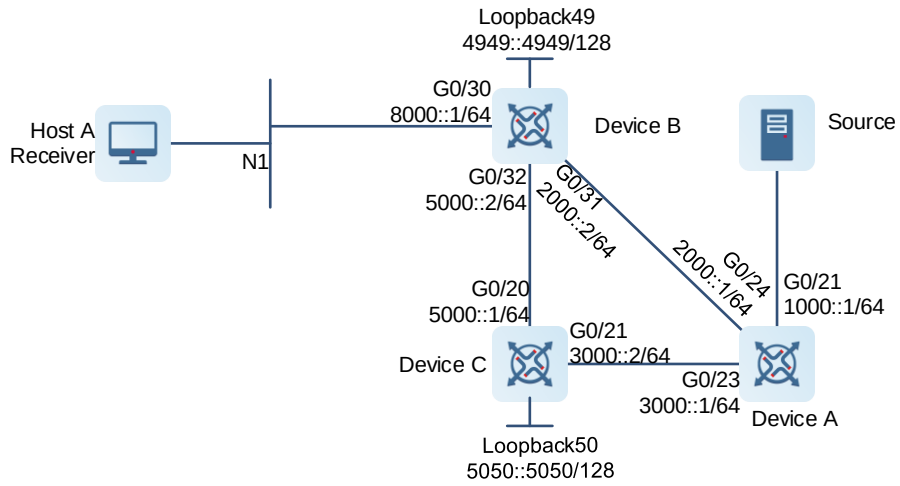
1.12.1 Configuring an IPv6 Static Multicast Route to Modify the RPF Route

1. Requirements

As shown in Figure 1-3, Device A, Device B, and Device C run the OSPF protocol and the PIM-SMv6 function is enabled on each interface. Originally, multicast data sent by the multicast source to Host A is forwarded through the path Device A > Device B. To reduce the load of the link from Device A to Device B, multicast data stream is separated from the unicast data stream. Therefore, the multicast data from the multicast source is sent through the path Device A > Device C > Device B.

2. Topology

Figure 1-3 Topology of IPv6 Static Multicast Route Configuration for Modifying the RPF Route



3. Notes

As shown in Figure 1-3, separate multicast data from unicast data by configuring a static multicast route on Device B to modify the RPF interface of the multicast data. Unicast data is forwarded through the path Device A > Device B, and multicast data is forwarded through the path Device A > Device C > Device B. The specific configuration method is as follows:

- Configure an IPv6 address and a unicast routing protocol on the multicast devices. A reachable unicast route is a basis of multicast protocol running.
- Enable the IPv6 multicast routing function on all multicast devices. Enable the PIM-SMv6 function on the interfaces connecting multicast devices, the interfaces connecting user hosts, and the multicast source interfaces. Configure C-BSR and C-RP on Device C.
- Configure an RPF static multicast route on Device B, specify Device C as an RPF neighbor, and modify the forwarding path of the multicast data packets.

4. Procedure

(1) Configure an IPv6 address and a unicast routing protocol on the multicast devices.

Configure an IPv6 address and a unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 address 1000::1/64
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/23)# ipv6 address 3000::1/64
```

```
DeviceA(cfg-if-GigabitEthernet 0/23)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# interface GigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 address 2000::1/64
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
DeviceA(cfg)# ipv6 router-protocol ospf 4949
```

Configure an IPv6 address and a unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 address 8000::1/64
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/31)# ipv6 address 2000::2/64
DeviceB(cfg-if-GigabitEthernet 0/31)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 address 5000::2/64
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ipv6 address 4949::4949/128
DeviceB(cfg-if-Loopback 49)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-Loopback 49)# exit
DeviceB(cfg)# ipv6 router-protocol ospf 4949
```

Configure an IPv6 address and a unicast routing protocol on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 address 5000::1/64
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 address 3000::2/64
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface Loopback 50
```

```
DeviceC(cfg-if-Loopback 50)# ipv6 address 5050::5050/128
DeviceC(cfg-if-Loopback 50)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-Loopback 50)# exit
DeviceC(cfg)# ipv6 router-protocol ospf 4949
```

(2) Enable the IPv6 multicast routing function on all multicast devices. Enable the PIM-SMv6 function on the interfaces connecting multicast devices, the interfaces connecting user hosts, and the multicast source interfaces. Configure C-BSR and C-RP on Device C.

Enable the multicast routing and PIM-SMv6 related functions on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# interface GigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
```

Enable the multicast routing and PIM-SMv6 related functions on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 mrouting-enable
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ipv6 pim sparse-mode
DeviceB(cfg-if-Loopback 49)# exit
```

Enable the multicast routing and PIM-SMv6 related functions on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ipv6 mrouting-enable
DeviceC(cfg)# ipv6 mrouting-enable
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
```

```
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ipv6 pim sparse-mode
DeviceC(cfg-if-Loopback 50)# exit
```

Configure C-BSR and C-RP on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6 pim c-bsr GigabitEthernet 0/21
DeviceA(cfg)# ipv6 pim c-rp GigabitEthernet 0/21
```

(3) Configure an RPF static multicast route on Device B, specify Device C as an RPF neighbor, and modify the forwarding path of the multicast data packets.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6    mrouting-table 1000::/64 5000::1
```

5. Verification

(1) Run the **display ipv6 mld groups** command on Device B to check whether Host A joins the group FF15::11 and the multicast source 1000::2 sends packets to the group FF15::11.

```
DeviceB# display ipv6 mld groups
MLD Connected Group Membership
Group Address                          Interface                          Uptime      Expires
Last Reporter
ff15::11                               GigabitEthernet 0/30               00:12:17    00:03:00
fe80::8205:88ff:fe22:b468
```

(2) Run the **display ipv6 mrouting-table** command on Device A to display the hardware routing entry and check whether the next hop is Device C and the outbound interface is GigabitEthernet 0/23.

```
DeviceA# display ipv6 mrouting-table

IPv6 Multicast Routing Table
Flags: I - Immediate Stat, T - Timed Stat, F - Forwarder installed,
           R - RPT, S - SPT, s - SSM Group
Timers: Uptime/Stat Expiry
Interface State: Interface

(1000::2, ff15::11), uptime 00:08:53, stat expires 00:03:12
Owner PIMSMV6, Flags: TFS
  Incoming interface: GigabitEthernet 0/21
  Outgoing interface list:
    GigabitEthernet 0/23
```

(3) Run the **display ipv6 rpf-routing** command on Device B to check whether the next hop of the multicast source is Device C and the RPF interface is GigabitEthernet 0/32.

```
DeviceB# display ipv6 rpf-routing 1000::
RPF information for 1000::
RPF interface: GigabitEthernet 0/32
   RPF neighbor: 5000::1
   RPF route: 1000::/64
   RPF type: static-route
   RPF recursion count: 0
   Doing distance-preferred lookups across tables
Distance: 0
   Metric: 0
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/21
 unset switch-port
 ipv6 address 1000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/23
 unset switch-port
 ipv6 address 3000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/24
 unset switch-port
 ipv6 address 2000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim c-bsr GigabitEthernet 0/21
ipv6 pim c-rp GigabitEthernet 0/21
!
```

- Device B configuration file

```
hostname DeviceB
!
ipv6 mrouting-enable
```

```
!
ipv6 mroute 1000::/64 5000::1
!
interface GigabitEthernet 0/30
 unset switch-port
 ipv6 address 8000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/31
 unset switch-port
 ipv6 address 2000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/32
 unset switch-port
 ipv6 address 5000::2/64
 ipv6 pim sparse-mode
!
interface Loopback 49
 ipv6 address 4949::4949/128
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
```

- Device C configuration file

```
hostname DeviceC
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/20
 unset switch-port
 ipv6 address 5000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/21
 unset switch-port
 ipv6 address 3000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 50
```

```
 ipv6 address 5050::5050/128
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
```

7. Common Errors

- IPv6 unicast routing is incorrectly configured.
- The IPv6 multicast routing function is not enabled on a multicast device.
- No IPv6 multicast protocol is enabled on an interface.

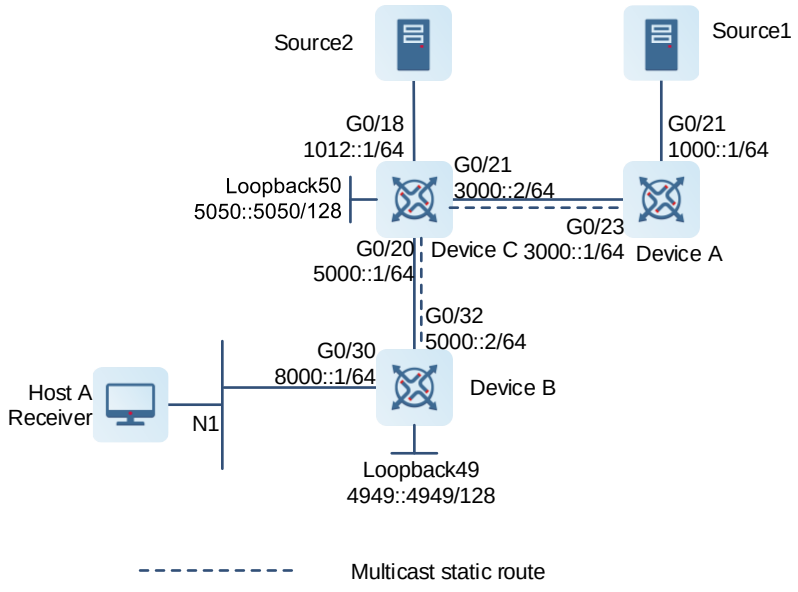
1.12.2 Configuring a Static Multicast Route to Connect the RPF Route

1. Requirements

As shown in Figure 1-4, Device B and Device C run the OSPF protocol, and the unicast route to Device A is isolated. The PIM-SMv6 function is enabled on the interfaces of Device A, Device B and Device C. Host A can normally receive multicast data packets from Source 2, and it is required to receive multicast data packets from Source 1 as well.

2. Topology

Figure 1-4 Topology of IPv6 Static Multicast Route Configuration for Connecting the RPF Route



3. Notes

Configure a static multicast route to establish an RPF route that is destined to source 1 based on the path Device B > Device C > Device A. The specific configuration method is as follows:

- Configure an IPv6 address for each interface of the devices, and configure the OSPF protocol on Device B and Device C to separate multicast routes from unicast routes on Device A.

- Enable the IPv6 multicast routing function, configure the PIM-SMv6 protocol on each L3 interface of the devices and configure the C-BSR and C-RP in the global configuration, and enable the MLD function on the interface connected to hosts.
- Configure static multicast routes to source 1 on Device B and Device C.

4. Procedure

(1) Configure an IPv6 address for each interface of the devices, and configure the OSPF protocol on Device B and Device C to separate multicast routes from unicast routes on Device A.

Configure an IPv6 address for the interface on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 address 1000::1/64
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/23)# ipv6 address 3000::1/64
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# ipv6 route-config 5050::5050/128 3000::2
```

Configure an IPv6 address and the OSPF protocol for the interface on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 address 8000::1/64
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 ospf 4949 area 0
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 address 5000::2/64
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ipv6 address 4949::4949/128
DeviceB(cfg-if-Loopback 49)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-Loopback 49)# exit
DeviceB(cfg)# ipv6 router-protocol ospf 4949
DeviceB(cfg-router)# exit
```

Configure an IPv6 address and the OSPF protocol for the interface on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/18
DeviceC(cfg-if-GigabitEthernet 0/18)# unset switch-port
```

```
DeviceC(cfg-if-GigabitEthernet 0/18)# ipv6 address 1012::1/64
DeviceC(cfg-if-GigabitEthernet 0/18)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/18)# exit
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 address 5000::1/64
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 ospf 4949 area 0
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 address 3000::2/64
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ipv6 address 5050::5050/128
DeviceC(cfg-if-Loopback 50)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-Loopback 50)# exit
DeviceC(cfg)# ipv6 router-protocol ospf 4949
DeviceC(cfg-router)# exit
```

(2) Enable the IPv6 multicast routing function, configure the PIM-SMv6 protocol on each L3 interface of the devices and configure the C-BSR and C-RP in the global configuration, and enable the MLD function on the interface connected to hosts.

Enable the IPv6 multicast routing function and configure the PIM-SMv6 protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
```

Enable the IPv6 multicast routing function and configure the PIM-SMv6 protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 mrouting-enable
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ipv6 pim sparse-mode
DeviceB(cfg-if-Loopback 49)# exit
```

Enable the IPv6 multicast routing function and functions related to the PIM-SMv6 protocol on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ipv6 mrouting-enable
DeviceC(cfg)# interface GigabitEthernet 0/18
DeviceC(cfg-if-GigabitEthernet 0/18)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/18)# exit
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ipv6 pim sparse-mode
DeviceC(cfg-if-Loopback 50)# exit
DeviceC(cfg)# ipv6 pim c-bsr Loopback 50
DeviceC(cfg)# ipv6 pim c-rp Loopback 50
```

Enable Source 1 (1000::2/64) and Source 2 (1012::2/64) to send multicast data packets to the multicast group (FF14::10). Add Host A to the multicast group (FF14::10) so that Host A can receive multicast data packets from Source 2 but cannot receive multicast data packets from Source 1.

Run the **display ipv6 rpf -routing**1000::2 command on Device C. No route exists. No RPF route to Source 1 exists on Device C.

(3) Configure static multicast routes to source 1 on Device B and Device C.

On Device B, configure an RPF static multicast route and configure Device C as an RPF neighbor to Source 1.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 rpf-mroute 1000::/64 5000::1
```

On Device C, configure an RPF static multicast route and configure Device A as an RPF neighbor to Source 1.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ipv6 rpf-mroute 1000::/64 3000::1
```

5. Verification

Run the **display ipv6 rpf-routing** 1000::2 command on Device B to display the RPF information of Source 1. The RPF information is displayed as follows:

```
DeviceB# display ipv6 rpf-routing 1000::2
RPF information for 1000::2
RPF interface: GigabitEthernet 0/32
  RPF neighbor: 5000::1
  RPF route: 1000::/64
  RPF type: static-route
```

```
  RPF recursion count: 0
  Doing distance-preferred lookups across tables
Distance: 0
  Metric: 0
```

Run the **display ipv6 rpf** 1000::2 command on Device C to display the RPF information of Source 1. The RPF information is displayed as follows:

```
DeviceC# display ipv6 rpf-routing 1000::2
RPF information for 1000::2
RPF interface: GigabitEthernet 0/21
  RPF neighbor: 3000::1
  RPF route: 1000::/64
  RPF type: static-route
  RPF recursion count: 0
  Doing distance-preferred lookups across tables
Distance: 0
  Metric: 0
```

Enable Source 1 and Source 2 to send multicast data packets to the multicast group (FF14::10), and add Host A to the multicast group.

Run the **display ipv6 pim-info sparse-mode mroute** command on Device C to display routing entry information. The multicast entries of Source 1 exist on Device C. Host A normally receives multicast data packets from Source 1.

```
DeviceC# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 1
(S,G) Entries: 2
(S,G,rpt) Entries: 2
FCR Entries: 0
REG Entries: 0

(*, ff14::10)
RP: 5050::5050
RPF nbr: ::
RPF idx: None
Upstream State: JOINED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
Joined
0 . . . j . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Asserted
0 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
FCR:

(1000::2, ff14::10)
RPF nbr: fe80::2d0:f8ff:fe22:358d
RPF idx: GigabitEthernet 0/21
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 11 seconds
kat expires in 209 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Joined
0 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
```

```
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(1000::2, ff14::10, rpt)
RP: 5050::5050
RPF nbr: ::
RPF idx: None
Upstream State: PRUNED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Pruned
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . o . . . . . .
1 . . . . . .
```

```
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(1012::2, ff14::10)
RPF nbr: fe80::2d0:f8ff:fe22:363d
RPF idx: GigabitEthernet 0/18
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 19 seconds
kat expires in 147 seconds
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . o . . . . . .
1 . . . . . .
2 . . . . . .
```

```
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(1012::2, ff14::10, rpt)
RP: 5050::5050
RPF nbr: ::
RPF idx: None
Upstream State: PRUNED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Pruned
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

DeviceC#

DeviceB> privilege
DeviceB# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
```

```
(*,G) Entries: 1
(S,G) Entries: 0
(S,G,rpt) Entries: 0
FCR Entries: 2
REG Entries: 0

(*, ff14::10)
RP: 5050::5050
RPF nbr: fe80::274:9cff:feee:f49f
RPF idx: GigabitEthernet 0/32
Upstream State: JOINED
jt_timer expires in 58 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  i  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  j  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
FCR:
Source: 1000::2
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Outgoing
0  .  0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
  KAT timer running, 120 seconds remaining
  Packet count 1
Source: 1012::2
    00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Outgoing
0  .  0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
  KAT timer running, 99 seconds remaining
  Packet count 1
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/21
  unset switch-port
  ipv6 address 1000::1/64
  ipv6 pim sparse-mode
!
interface GigabitEthernet 0/23
  unset switch-port
  ipv6 address 3000::1/64
  ipv6 pim sparse-mode
!
ipv6 route-config 5050::5050/128 3000::2
!
```

- Device B configuration file

```
hostname DeviceB
!
ipv6 mrouting-enable
```

```
!
ipv6 rpf-mroute 1000::/64 5000::1
!
interface GigabitEthernet 0/30
 unset switch-port
 ipv6 address 8000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/32
 unset switch-port
 ipv6 address 5000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
```

- Device C configuration file

```
hostname DeviceC
!
ipv6 mrouting-enable
!
ipv6 rpf-mroute 1000::/64 3000::1
!
interface GigabitEthernet 0/18
 unset switch-port
 ipv6 address 1012::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/20
 unset switch-port
 ipv6 address 5000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/21
 unset switch-port
 ipv6 address 3000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 50
 ipv6 address 5050::5050/128
 ipv6 ospf 4949 set-area 0
```

```
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim c-bsr Loopback 50
ipv6 pim c-rp Loopback 50
!
```

7. Common Errors

- IPv6 unicast routing is incorrectly configured.

Contents

1 Configuring MLD 1
1.1 Introduction 1
1.1.2 Working Principles of MLDv1 2
1.1.3 Working Principles of MLDv2 6
1.1.4 MLD SSM Mapping 7
1.1.5 MLD Proxy 7
1.1.6 Protocols and Standards 8
1.2 Restrictions and Guidelines 8
1.3 Configuration Task Summary 8
1.4 Prerequisites 9
1.5 Configuring Basic Functions of MLD 9
1.5.1 Overview 9
1.5.2 Restrictions and Guidelines 9
1.5.3 Prerequisites 9
1.5.4 Procedure 9
1.6 Configuring an MLD Version 10
1.6.1 Overview 10
1.6.2 Restrictions and Guidelines 10
1.6.3 Prerequisites 10
1.6.4 Procedure 10
1.7 Enabling Security Control for MLD Multicast 11
1.7.1 Overview 11

1.7.2 Configuration Tasks 11
1.7.3 Prerequisites 11
1.7.4 Enabling Access Control for Multicast Groups 11
1.7.5 Configuring the Maximum Number of Group Members 12
1.8 Configuring MLD Querier Parameters 13
1.8.1 Overview 13
1.8.2 Restrictions and Guidelines 13
1.8.3 Prerequisites 14
1.8.4 Configuration Tasks 14
1.8.5 Configuring the Last Member Query Interval 14
1.8.6 Configuring the Last Member Query Times 15
1.8.7 Configuring the General Group Query Interval 15
1.8.8 Configuring the Maximum Response Time of Query Packets 16
1.8.9 Configuring the Other Querier Keepalive Time 17
1.8.10 Configuring the Robustness Variable of the MLD Querier 17
1.9 Configuring Static MLD Groups 18
1.9.1 Overview 18
1.9.2 Restrictions and Guidelines 18
1.9.3 Prerequisites 18
1.9.4 Procedure 19
1.10 Configuring a Device Interface as a Simulated Host to Join a Group 19
1.10.1 Overview 19
1.10.2 Restrictions and Guidelines 19
1.10.3 Prerequisites 19

1.10.4 Procedure....19
1.11 Enabling the MLD Proxy Function....20
1.11.1 Overview....20
1.11.2 Restrictions and Guidelines....20
1.11.3 Prerequisites....20
1.11.4 Procedure....20
1.12 Enabling the SSM Mapping Function....21
1.12.1 Overview....21
1.12.2 Restrictions and Guidelines....21
1.12.3 Prerequisites....21
1.12.4 Procedure....21
1.13 Monitoring....21
1.14 Examples....23
1.14.1 Configuring Basic Functions of MLD....23
1.14.2 Configuring the MLD Proxy Function....31
1.14.3 Configuring SSM Mapping....35
1.15 Common Misconfigurations....38
1.15.1 What Can I Do when On-demand Multicast Service Is Unavailable on Client....39

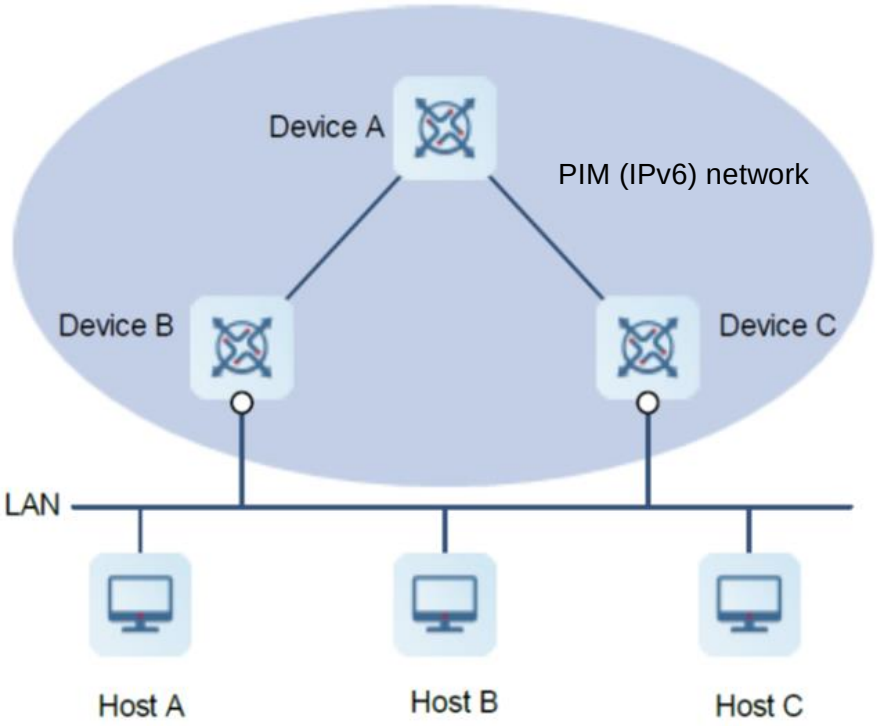
1 Configuring MLD

1.1 Introduction

The Multicast Listener Discovery (MLD) protocol manages IPv6 multicast members.

MLD defines how to establish and maintain membership between multicast devices and member hosts in an IPv6 network, to decide a forwarding path of multicast streams. Figure 1-1 shows the topology of MLD in an IPv6 multicast network.

Figure 1-1 Topology of MLD in a Multicast Network



O Interface with the MLD function enabled

There are two versions of MLD: MLDv1 is based on version 2 of the Internet Group Management Protocol (IGMP) and MLDv2 is based on version 3 of IGMP. A later MLD version is backward-compatible with an earlier MLD version. MLDv1 completes IPv6 member management based on the query and report mechanism. MLDv2 supports member hosts to decide whether to receive packets from a specified multicast source on the basis of MLDv1.

- MLD of both versions supports the Any-Source Multicast (ASM) model.
- MLDv2 can be directly applied to the Source-Specific Multicast (SSM) model.
- MLDv1 can be applied to the SSM model only when MLD SSM mapping is configured.

1.1.2 Working Principles of MLDv1

The querier election mechanism, general group query and report mechanism, join mechanism, and leave mechanism are defined in MLDv1.

1. Types of MLDv1 Packets

MLDv1 supports the following four types of packets:

- General group query packet: A querier sends general group query packets to all multicast devices and hosts in the same network segment to understand groups with members.
- Group-specific query packet: A querier sends query packets to specified multicast groups in the same network segment to understand whether these groups contain members.
- Report packet: A host sends report packets to the querier to respond to the query packets or join a group.
- Leave packet: A host sends leave packets to the querier when the host leaves a group, notifying the querier that it leaves the specified group.

2. Working Mechanisms of MLDv1

- Querier election

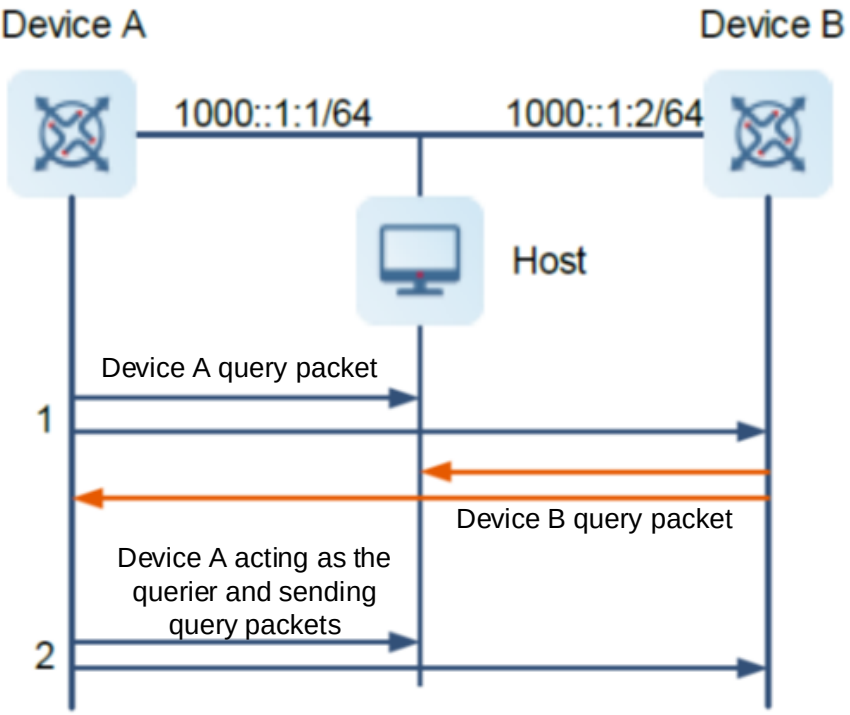
  If multiple multicast devices exist in a network segment, only one multicast device needs to send query packets as a querier. This avoids redundancy. When the querier in the network segment fails, other devices in the same network segment need to take the role and process MLD protocol packets. During querier election, interface addresses of the devices are compared. The device with a smallest IP address is elected as the querier, as shown in Figure 1-2.

  a Device A and Device B consider themselves as the querier and send query packets to the local network segment.

  b They compare the source IPv6 address in the received packets with the address of their interfaces. The address of Device A is smaller. Therefore, Device A acts as the querier to periodically send query packets.

  c Device B fails in the election and starts a query keepalive timer. When the timer expires, Device B considers that the querier fails and starts to send query packets to initiate a new round of querier election.

Figure 1-2 Querier Election Diagram



- General group query and report

  The MLD querier periodically sends general group query packets to all hosts and multicast devices in the local network segment, and understands and maintains group membership based on the report packets replied by the member hosts. Figure 1-3 shows the general group query and report procedure.

  a The MLD querier Device A periodically sends general group query packets to the destination address FF02::1.

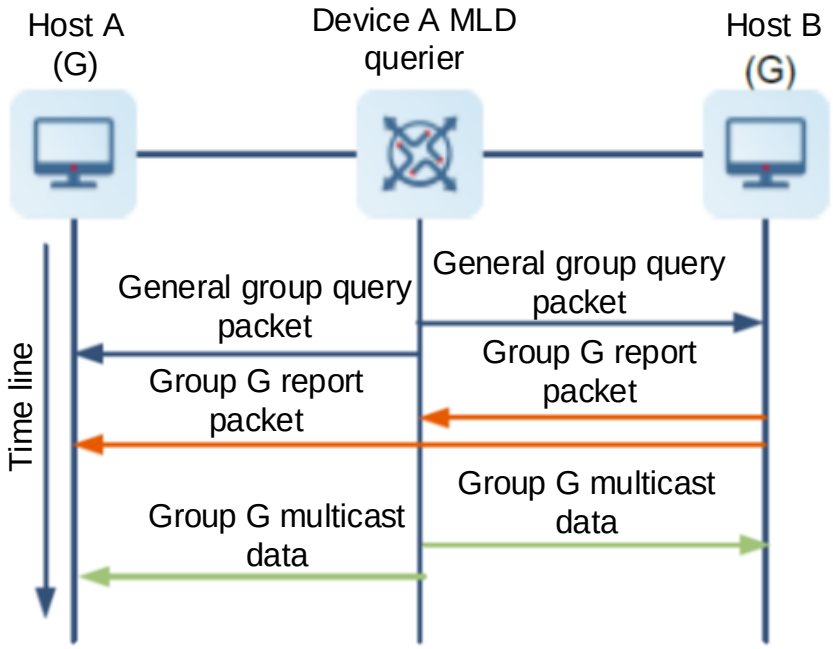
  b Upon receiving the general group query packets, Host A and Host B in the multicast group G start respective timers Timer-G-A and Timer-G-B.

  c The host whose timer expires first sends report packets destined for the multicast group G.

  If the timer Timer-G-A of Host A expires first, Host A sends report packets to the multicast group G. Upon receiving the report packets, Host B stops its timer and sends report packets in the network segment.

  d Upon receiving the report packets destined for the multicast group G from Host A, the MLD querier knows that there is member in the multicast group G. The multicast routing protocol generates an (*, G) entry. Upon receiving multicast data destined for the multicast group G, the MLD querier forwards the multicast data in the network segment.

Figure 1-3 General Group Query and Report Diagram



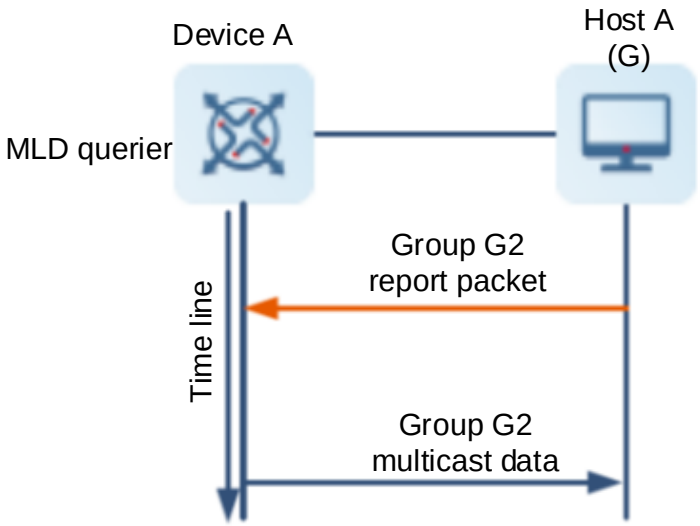
- Member join

  When a host tries to join a multicast group, it takes some time for the host to join the group if the host waits for a general group query packet from the querier and replies a report packet. If the host can proactively send a report packet to the MLD querier, the querier can immediately generate a multicast forwarding entry upon receiving the packet. If multicast devices in the group receive multicast data, the new host can quickly get multicast service. Figure 1-4 shows the new member join procedure.

  a Before receiving a general group query packet from Device A, Host A proactively sends a report packet destined for the multicast group G, to notify that Host A joins the multicast group G.

  b Upon receiving the report packet, the MLD querier Device A knows that there is a member in the multicast group G in the current network segment. The multicast routing protocol generates an (*, G) entry. Upon receiving multicast data destined for the multicast group G, the MLD querier forwards the multicast data in the network segment.

Figure 1-4 Member Join Diagram



- Member leave

  Upon receiving a leave packet from a member to plan to leave the multicast group G, the MLD querier sends a group-specific query packet to understand whether there is any other member existing in the group in the network segment. If not, the MLD querier deletes the corresponding membership to reduce redundant multicast traffic in the network and save bandwidth. Figure 1-5 shows the member leave procedure.

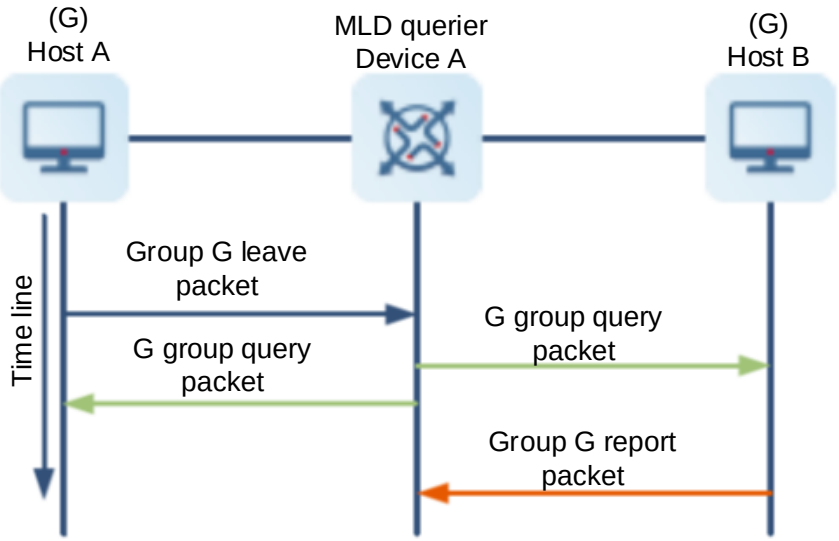
  a Host A sends a leave packet destined for all multicast devices in the local network segment to leave the multicast group G.

  b Upon receiving the leave packet, the MLD querier Device A sends a group-specific query packet destined for the multicast group G.

  c If there is any other member (for example, Host B) in the multicast group G, Host B sends a report packet for the multicast group G upon receiving the query packet. The MLD querier continues to maintain the membership of the multicast group G upon receiving the report packet. The (*, G) entry remains existent. Upon receiving multicast data destined for the multicast group G, the MLD querier forwards the multicast data in the network segment.

  d If there is no other member in the multicast group G, the querier receives no report packet for the multicast group G. After the timer expires, the (*, G) entry is deleted. Upon receiving multicast data destined for the multicast group G, the MLD querier does not forward the multicast data in the network segment.

Figure 1-5 Member Leave Diagram



1.1.3 Working Principles of MLDv2

1. Changes of MLDv2 from MLDv1

- In MLDv2, when you add a host to a group, you can specify the host to receive multicast data from specified sources. In MLDv1, you cannot specify sources for a host.
- In MLDv1, an MLDv1 packet can carry information of only one multicast group. In MLDv2, an MLDv2 packet can carry information of multiple multicast groups. This relieves the interaction problem of a great number of packets between hosts and the querier.

2. Types of MLDV2 Packets

- Query packets include general group query packet, group-specific query packet, and source-specific query packet (check whether members in a group receive multicast data from the specific source).
- Member report packets include multicast group and multicast source information. A source can be filtered in INCLUDE or EXCLUDE mode. (G, INCLUDE, (S1, S2...)) means to receive multicast data sent only from specified sources S1, S2, ... to the group G. (G, EXCLUDE, (S1, S2...)) means to reject multicast data sent from specified sources S1, S2, ... to the group G.

3. Working Mechanisms of MLDv2

Compared with MLDv1, MLDv2 supports you to select sources for a host to receive multicast data when you add the host.

- Specify sources to receive multicast data: You can specify sources for the host to receive multicast data when you add the host to a multicast group.
- Query a specified source: Upon receiving a report packet to modify the correspondence between multicast sources and groups from members, the MLD querier sends a specified source query packet. If you expect a member to receive multicast data from a specified source, configure a report packet. The MLD querier updates the source list of this group based on the report.

1.1.4 MLD SSM Mapping

In the source-specific multicast (SSM) model, a host must run MLDv2 if you want to specify source addresses for the host. If a host supports only MLDv1, a multicast source cannot be specified for the host. To enable a host that supports only MLDv1 to use the SSM service, you can enable the MLD SSM mapping function on the multicast device.

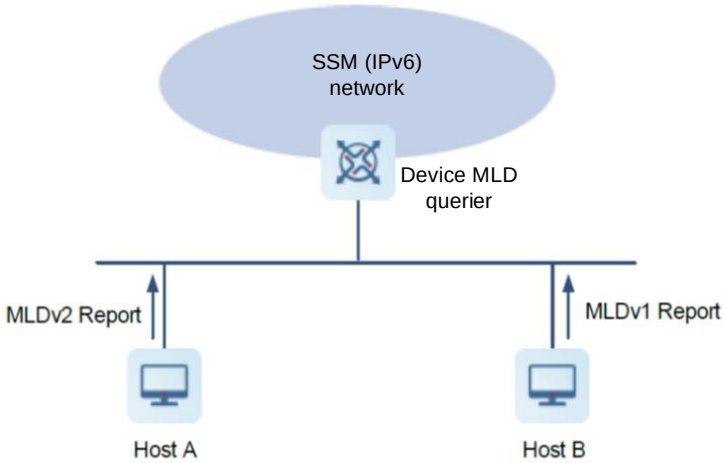
Based on the MLD SSM mapping mechanism, static SSM address mapping rules can be configured on the multicast device to convert the (*, G) information in MLDv1 report packets to (S, G) information so that the SSM service can be used by the host that supports only MLDv1.

In the SSM network shown in Figure 1-6, the MLD multicast device runs MLDv2, Host A supports only MLDv1, and Host B supports MLDv2. The MLD SSM mapping function can be enabled on the MLD querier to provide SSM service for Host A.

In this network, when the device receives an MLDv1 report packet from Host A, the device checks the address of the IPv6 multicast group G in the packet.

- If the multicast group G is not in the IPv6 SSM group address range, the ASM service is provided.
- If the multicast group G is in the IPv6 SSM group address range, whether the ASM service can be provided depends on the following situation:
  - If no MLD SSM mapping rule is configured on the device for the multicast group G, the SSM service cannot be provided and the packet is discarded.
  - If an MLD SSM mapping rule is configured on the device for the multicast group G, the (*, G) information in the MLDv1 report packet is mapped to (G, INCLUDE, (S1, S2)) and the SSM multicast service is provided.

Figure 1-6 Topology of MLD SSM Mapping



1.1.5 MLD Proxy

In some user scenarios, if you do not want devices connected to hosts to run complete multicast routing protocols, you can enable the MLD proxy function.

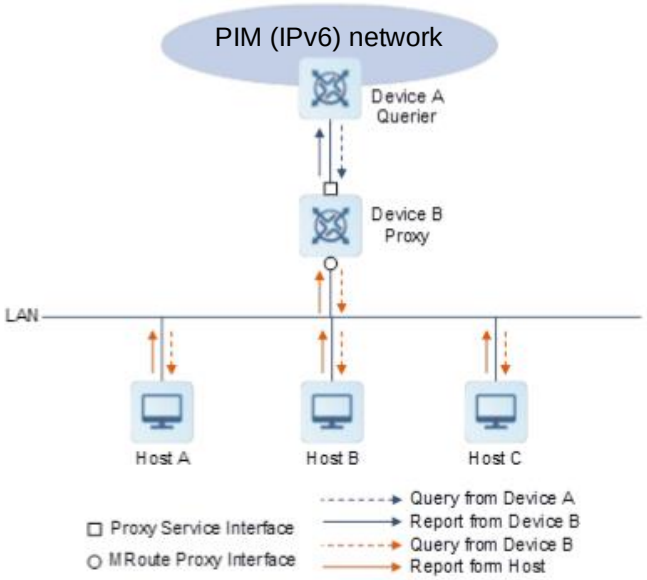
On the user side, the MLD proxy acts as a querier and periodically queries user information. On the route side, the MLD proxy acts a host and sends user information to a network when necessary. In this way, L2 multicast

flooding is suppressed, user information is more effectively obtained and controlled, and the protocol packets and network load are reduced to some extent.

Two types of MLD proxy interfaces are defined to facilitate understanding:

- MLD proxy service interface: This interface is responsible for behaviors of hosts. It receives query packets from uplink devices, replies report packets based on current group membership, and sends report or leave packets when the membership changes.
- MLD MRoute proxy interface: This interface is responsible for behaviors of the querier. It sends query packets, collects host information, and maintains membership.

Figure 1-7 Topology of MLD Proxy



1.1.6 Protocols and Standards

- RFC 2710: Multicast Listener Discovery (MLDv1) for IPv6
- RFC 3810: Multicast Listener Discovery Version 2 (MLDv2) for IPv6

1.2 Restrictions and Guidelines

- In an IPv6 network, all L3 devices must run the IPv6 PIM protocol before multicast forwarding entries can be established normally.
- Device interfaces must run MLDv2 before hosts supporting either MLD version can use the SSM service.
- A member host can be a PC or a set top box (STB), and it can normally receive multicast data by the multicast client software.

1.3 Configuration Task Summary

MLD configuration includes the following tasks:

(1) Configuring Basic Functions of MLD

(2) (Optional) Configuring an MLD Version

(3) Enabling Security Control for MLD Multicast. All the configuration tasks below are optional. Select the configuration tasks as required.

- Enabling Access Control for Multicast Groups
- Configuring the Maximum Number of Group Members

(4) Configuring MLD Querier Parameters. All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the Last Member Query Interval
- Configuring the Last Member Query Times
- Configuring the General Group Query Interval
- Configuring the Maximum Response Time of Query Packets
- Configuring the Other Querier Keepalive Time
- Configuring the Robustness Variable of the MLD Querier

(5) (Optional) Configuring Static MLD Groups

(6) (Optional) Configuring a Device Interface as a Simulated Host to Join a Group

(7) (Optional) Enabling the MLD Proxy Function

(8) (Optional) Enabling the SSM Mapping Function

1.4 Prerequisites

An IPv6 unicast routing protocol is configured to ensure that device nodes are reachable through IPv6 unicast routes.

1.5 Configuring Basic Functions of MLD

1.5.1 Overview

The multicast routing function is enabled on the local network and group information on the local network is collected so that member hosts can join multicast groups to receive multicast data packets.

1.5.2 Restrictions and Guidelines

- The IPv6 multicast routing function should be enabled on each multicast device on the local network unless otherwise specified.
- The PIM-SMv6 function should be enabled on interfaces connected to the local network unless otherwise specified.

1.5.3 Prerequisites

All PIM interfaces should be reachable through IPv6 unicast routes.

1.5.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IPv6 multicast routing function.

ipv6 mrouting-enable

The IPv6 multicast routing function is disabled by default.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(5) Enable the PIM-SMv6 function.

ipv6 pim sparse-mode

The PIM-SMv6 function on an interface is disabled by default.

1.6 Configuring an MLD Version

1.6.1 Overview

At present, there are two versions of MLD. Multicast devices running MLDv2 can identify MLDv1 packets. Multicast devices running MLDv1 cannot identify MLDv2 packets.

1.6.2 Restrictions and Guidelines

- After an MLD version is configured, MLD automatically restarts.
- The MLD version on a multicast device must be later than or equal to the MLD version on a host.
- It is recommended that an MLD version be configured on all multicast device interfaces directly connected to the local network unless otherwise specified.

1.6.3 Prerequisites

Configuring Basic Functions of MLD.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure an MLD version.

ipv6 mld specify-version { **1** | **2** }

MLDv2 is running on an interface by default.

1.7 Enabling Security Control for MLD Multicast

1.7.1 Overview

- To prevent a host from joining an IPv6 multicast group, you can enable the function of filtering the MLD group on the interface directly connected to a network segment.
- To prevent a great number of hosts from joining an IPv6 multicast group and affecting interface performance of devices, you can configure the maximum number of members on the interface directly connected to a network segment.
- To prevent a great number of hosts from joining an IPv6 multicast group and affecting the overall performance of devices, you can configure the maximum number of members on the devices.

1.7.2 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Enabling Access Control for Multicast Groups
- Configuring the Maximum Number of Group Members

1.7.3 Prerequisites

Configuring Basic Functions of MLD.

1.7.4 Enabling Access Control for Multicast Groups

1. Overview

After enabling access control for multicast groups on an interface, you can control hosts from joining groups in the directly connected network segment. ACLs are used to limit the group address range. Report packets denied by the ACLs are discarded.

When MLDv2 is enabled, this command supports extended ACLs to precisely filter source record information in MLDv2 packets. If the received MLD report packets follow the format of (S1, S2, S3, ..., Sn, G), the corresponding ACL must be configured for filtering. S1, S2, S3, ..., Sn correspond to source addresses in ACL entries, and G corresponds to the destination address in the ACL entries.

2. Restrictions and Guidelines

It is recommended that access control be enabled on all multicast device interfaces directly connected to the local network unless otherwise specified.

3. Prerequisites

Before a multicast group address range is configured, ensure that the corresponding ACL is created and a multicast group range is specified. For details about ACL configuration, see *Configuring ACL*.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Enable access control for multicast groups.

ipv6 mld restricted-group *acl-name*

An interface allows hosts to join any group by default.

1.7.5 Configuring the Maximum Number of Group Members

1. Overview

To prevent a great number of hosts from joining an IPv6 multicast group, you can limit the number of members.

- In the global configuration mode, the number of members in an MLD group on a device is limited.
- In the interface configuration mode, the number of members in an MLD group on an interface is limited.

If the number of group members exceeds the interface limit or global limit, subsequent report packets are ignored.

If an exceptional list is configured, report packets in a specified range can be normally processed, and the records about these group members are not counted.

Interface and global limits can be configured separately. If the global limit is smaller than the interface limit, the global limit prevails.

2. Restrictions and Guidelines

It is recommended that the maximum number of group members be configured on all multicast device interfaces directly connected to the local network unless otherwise specified.

It is recommended that the maximum number of group members be configured on multicast devices directly connected to the local network unless otherwise specified.

3. Prerequisites

- Before an exceptional list is configured, ensure that the corresponding ACL is created and an exceptional multicast group range is specified. For details about ACL configuration, see *Configuring ACL*.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum number of MLD group members. Select one of the following to configure.

- Configure the maximum number of MLD group members on a device.

  ipv6 mld max-member-record *max-member-record* [**except** *acl-name*]

- Configure the maximum number of MLD group members on an interface.

  interface *interface-type interface-number*

  ipv6 mld max-member-record *max-member-record* [**except** *acl-name*]

By default, the maximum number of MLD group members on a device is **64000**, and that on an interface is **4000**.

1.8 Configuring MLD Querier Parameters

1.8.1 Overview

MLD maintains IPv6 membership by query packets and report packets. If multiple multicast devices exist in a network segment, only one multicast device needs to send query packets as a querier. IP addresses of multicast devices are compared. The device with the smallest address is elected as the querier. The querier periodically sends query packets to obtain local member information and ensure that a group has at least one host. If no host is available in a group, the group is deleted by the device.

The MLD querier work can be adjusted using multiple parameters. When the default parameter values are used, the querier can normally work. Network optimization can be implemented by adjusting querier parameter values to avoid excessive protocol packets and network congestion. Also, membership can be updated and maintained in time.

1.8.2 Restrictions and Guidelines

- It is recommended that querier parameters be configured on all multicast device interfaces directly connected to the local network unless otherwise specified.
- In actual network deployment, to accurately, timely, and stably maintain group membership and avoid

non-queriers from initiating a new round of election, ensure that the following durations are arranged in ascending order: Maximum response time of general group query packet < Sending interval of general group query packet < Other querier keepalive time.

1.8.3 Prerequisites

Configuring Basic Functions of MLD.

1.8.4 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the Last Member Query Interval
- Configuring the Last Member Query Times
- Configuring the General Group Query Interval
- Configuring the Maximum Response Time of Query Packets
- Configuring the Robustness Variable of the MLD Querier
- Configuring the Other Querier Keepalive Time

1.8.5 Configuring the Last Member Query Interval

1. Overview

Before a host leaves a multicast group, the host sends a leave packet. Upon receiving the leave packet, the querier continuously sends group-specific/source-specific query packets and waits for responses from hosts. This parameter specifies the query packet sending interval.

2. Restrictions and Guidelines

- It is recommended that this parameter be configured on all multicast device interfaces directly connected to the local network unless otherwise specified.
- The unit of the last member query interval is 0.1 seconds. If the parameter value is 10, the actual sending interval is 1 second.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.

interface loopback *interface-number*

(4) Configure the last member query interval.

ipv6 mld last-member-query-period *interval*

The group-specific query packet sending interval is **1** second by default.

1.8.6 Configuring the Last Member Query Times

1. Overview

Before a host leaves a multicast group, the host sends a leave packet. Upon receiving the leave packet, the querier continuously sends specific group/source query packets and waits for responses from the host. This parameter specifies the query packet sending times.

2. Restrictions and Guidelines

It is recommended that this parameter be configured on all multicast device interfaces directly connected to the local network unless otherwise specified.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the last member query times.

ipv6 mld last-member-query-number *count*

The group-specific query packet sending times is **2** by default.

1.8.7 Configuring the General Group Query Interval

1. Overview

The MLD querier periodically sends query packets to all hosts and multicast devices on the shared network. This parameter specifies the interval for sending query packets.

2. Restrictions and Guidelines

It is recommended that this parameter be configured on all multicast device interfaces directly connected to the local network unless otherwise specified.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the general group query interval.

ipv6 mld query-period *interval*

The default interval for sending general group query packets is **125** seconds.

1.8.8 Configuring the Maximum Response Time of Query Packets

1. Overview

After the MLD querier sends query packets, member hosts reply upon receiving the query packets before the maximum response time expires. If no response is received after timeout, it is considered that no member exists in the group of the directly connected network segment and the group information is deleted.

2. Restrictions and Guidelines

It is recommended that this parameter be configured on all multicast device interfaces directly connected to the local network unless otherwise specified.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*

- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the maximum response time of query packets.

ipv6 mld max-response-time *max-response-time*

The maximum response time of query packets is **10** seconds by default.

1.8.9 Configuring the Other Querier Keepalive Time

1. Overview

When a non-querier does not receive a query packet in the same network segment within a specified period, the non-querier considers there is only querier (itself) in the directly connected network segment and initiates a new round of querier election. This parameter specifies the timeout time for the non-queriers to wait for query packets.

2. Restrictions and Guidelines

It is recommended that this parameter be configured on all interfaces where MLD is enabled on the local network unless otherwise specified.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the other querier keepalive time.

ipv6 mld other-querier-timeout *timeout*

The keepalive time of other querier is **255** seconds by default.

1.8.10 Configuring the Robustness Variable of the MLD Querier

1. Overview

When the query packets sent by the querier are lost due to network instability, the query packets are retransmitted. The retransmission times are referred to as the robustness variable of the MLD querier.

In the event of packet loss, query packet retransmission occurs in either of the following situations:

- When the querier is started, it sends general group query packets.

- Upon receiving leave packets, the querier sends group-specific/source-specific query packets to check whether there are members existing in the specific multicast group.

2. Restrictions and Guidelines

It is recommended that this parameter be configured on all interfaces where MLD is enabled on the local network unless otherwise specified.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure the robustness variable of the MLD querier.

ipv6 mld robustness-variable *robustness*

The robustness variable of the MLD querier is **2** by default.

1.9 Configuring Static MLD Groups

1.9.1 Overview

In some application scenarios, a static multicast group may need to be configured on the interfaces connected to group members in the following cases:

1. The group members are stable and multicast data needs to be quickly and stably forwarded to member interfaces.

2. The group members cannot send report packets but need to receive certain multicast data. After the static multicast group is configured, the multicast devices consider that group members exist on these interfaces and forward multicast data to these interfaces.

1.9.2 Restrictions and Guidelines

A static multicast group must be manually configured on an interface connected to group members.

1.9.3 Prerequisites

Configuring Basic Functions of MLD.

1.9.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure a static MLD group.

ipv6 mld static-group-entries *ipv6-group-address*

No interface is added to any static multicast group by default.

1.10 Configuring a Device Interface as a Simulated Host to Join a Group

1.10.1 Overview

In some test scenarios, no host is available or can be added to a multicast group. In this case, you need to configure device interfaces to simulate hosts and add the hosts to a specific multicast group.

1.10.2 Restrictions and Guidelines

- An interface can be added to multiple groups at the same time.

1.10.3 Prerequisites

The PIM protocol must be configured and the IPv6 multicast routing function must be enabled on the interfaces to be configured as hosts.

1.10.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Configure a device interface as a host to join a group.

ipv6 mld add-to-join-group *ipv6-group-address*

No interface is added to any multicast group by default.

1.11 Enabling the MLD Proxy Function

1.11.1 Overview

The MLD router proxy function collects local member information. For a host, the MLD proxy functions as a router and sends query packets.

For the querier, the MLD proxy functions as a host and sends report packets.

1.11.2 Restrictions and Guidelines

- The MLD proxy service interface is configured on the interface of routers directly connected to the upstream devices unless otherwise specified.
- The MRoute proxy interface is configured on the interface of hosts directly connected to the downstream devices unless otherwise specified.
- A maximum of 32 proxy service interfaces can be configured on a device. After receiving MLD query packets, the proxy service interface sends report packets based on the MLD group member records.

1.11.3 Prerequisites

Configuring Basic Functions of MLD.

1.11.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet*-type interface-number
- Enter the Layer 3 link aggregation configuration mode.
  interface aggregateport *interface-number*
- Enter the SVI configuration mode.
  interface vlan *interface-number*
- Enter the loopback interface configuration mode.
  interface loopback *interface-number*

(4) Enable the MLD proxy service function.

ipv6 mld enable-proxy-service

The MLD proxy service function is disabled by default.

(5) Enable the MLD MRoute proxy function.

ipv6 mld enable-mrouting-proxy *interface-type interface-number*

The MRoute proxy function is disabled by default.

1.12 Enabling the SSM Mapping Function

1.12.1 Overview

MLDv2 supports source filtering while MLDv1 does not. However, the multicast devices provide the SSM mapping function to implement source filtering and enable MLDv1 hosts to use the SSM service. Multicast device interfaces run MLDv2. When receiving MLDv1 report packets, a multicast device adds the static mapping source address.

1.12.2 Restrictions and Guidelines

The SSM service must be enabled and the SSM group address range must be consistent on all multicast devices.

1.12.3 Prerequisites

- Configuring Basic Functions of MLD.
- Before a static mapping entry is configured, ensure that the corresponding ACL is created and a multicast group range is specified. For details about ACL configuration, see *Configuring ACL*.

1.12.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the SSM mapping function.

ipv6 mld ssm-policy enable

The SSM mapping function is disabled by default.

(4) (Optional) Configure a static mapping entry.

ipv6 mld ssm-policy static *acl-name ipv6-source-address*

No static mapping entry is configured by default.

1.13 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup ipv6 mld-cache dynamicgroup [*ipv6-group-address*] [*interface-type interface-number*]	Clears dynamic group member records in the MLD cache.
cleanup ipv6 mld-cache interface *interface-type interface-number*	Clears all MLD statistics and group member records on the interface.
display ipv6 mld-info groups [*ipv6-group-address* | *interface-type interface-number*] [**detail**]	Displays groups directly connected to the device and group information learned from MLD.
display ipv6 mld-info interface [*interface-type interface-number*]	Displays configurations of the interface.
display ipv6 mld-info ssm-mapping [*ipv6-group-address*]	Displays the SSM mapping information.
display debugging	Displays the MLD debugging switch status.
debugging ipv6 mld-info all	Debugs all MLD information.
debugging ipv6 mld-info decode	Debugs MLD packet resolution.
debugging ipv6 mld-info encode	Debugs MLD packet encoding.
debugging ipv6 mld-info events	Debugs MLD event information.
debugging ipv6 mld-info fsm	Debugs MLD event information.
debugging ipv6 mld-info tib	Debugs MLD state machine information.
debugging ipv6 mld-info warning	Debugs MLD alarms.

1.14 Examples

1.14.1 Configuring Basic Functions of MLD

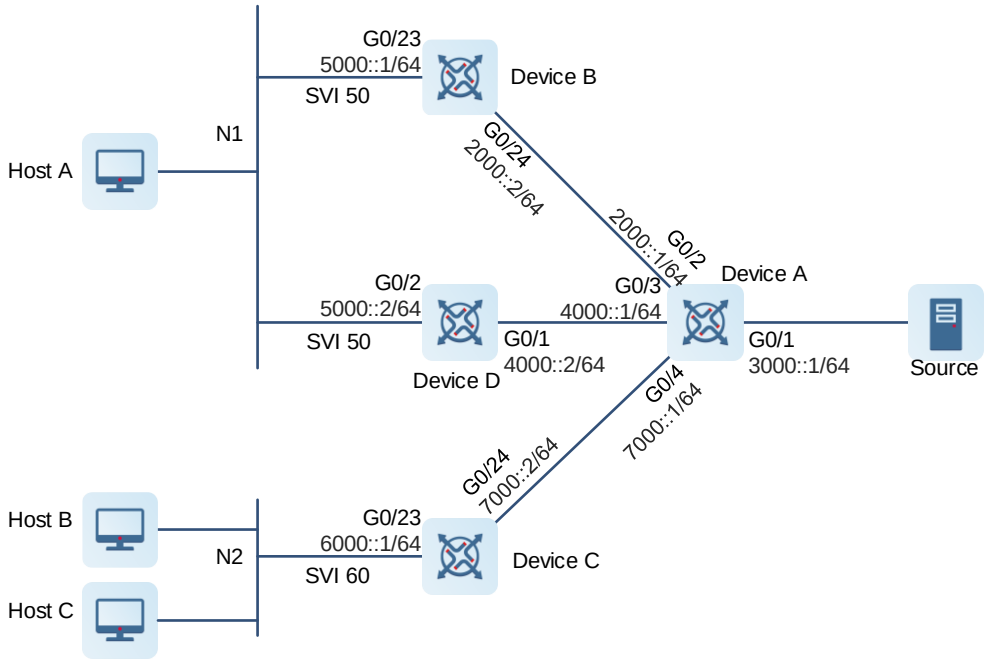
1. Requirements

As shown in Figure 1-8, Host A, Host B, and Host C residing in shared networks N1 and N2 of IPv6 want to receive video information from the source in multicast mode. The static multicast group FF13::1 is configured on N1 shared network, and hosts on N2 shared network can join the multicast group FF14::1.

- OSPFv3 runs in the network to make unicast routes between any two nodes reachable.
- Device B and Device D are connected to the shared network N1 through SVI 50. The IPv6 address of Device B is smaller than the IPv6 address of Device D. Therefore, Device B acts as the querier, and Host A is the multicast information receiver in the shared network N1.
- Device C is connected to the shared network N2 through SVI 60, and Host B is the multicast information receiver in the shared network N2.
- Device B, Device C, and Device D are connected to other devices in the IPv6 PIM network through L3 interfaces.

2. Topology

Figure 1-8 Topology for Basic Functions of MLD



3. Notes

- Configure an IPv6 unicast routing protocol OSPFv3 on a multicast device and ensure that unicast routes are reachable to the Loopback interface.
- Enable the IPv6 multicast routing function on all multicast devices.
- Configure the PIM-SMv6 protocol on the interconnected interfaces of devices, interfaces connected to hosts,

and interfaces connected to sources.

- Configure a static multicast group on the interfaces of the devices on shared network N1.
- Enable the multicast group access control function on the interfaces of the devices on shared network N2.

4. Procedure

(1) Configure an IPv6 address for each interface and configure the unicast routing protocol.

Configure an IPv6 address for the interface of Device A and configure the OSPFv3 protocol to implement internetworking. Perform the same configuration on Device B, Device C, and Device D.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6 router-protocol ospf 5050
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 address 3000::1/64
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface GigabitEthernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/2)# ipv6 address 2000::1/64
DeviceA(cfg-if-GigabitEthernet 0/2)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
DeviceA(cfg)# interface GigabitEthernet 0/3
DeviceA(cfg-if-GigabitEthernet 0/3)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/3)# ipv6 address 8000::1/64
DeviceA(cfg-if-GigabitEthernet 0/3)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/3)# exit
DeviceA(cfg)# interface GigabitEthernet 0/4
DeviceA(cfg-if-GigabitEthernet 0/4)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/4)# ipv6 address 7000::1/64
DeviceA(cfg-if-GigabitEthernet 0/4)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/4)# exit
DeviceA(cfg)# interface loopback 0
DeviceA(cfg-if-Loopback 0)# ipv6 address 1111::1/128
DeviceA(cfg-if-Loopback 0)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-Loopback 0)# exit
```

(2) Enable the IPv6 multicast routing function on all devices, and configure the PIM-SMv6 protocol on the interconnected interfaces of devices, interfaces connected to hosts, and interfaces connected to sources.

Enable the IPv6 multicast routing function and configure the PIM-SMv6 protocol on Device A. Perform the same configuration on Device B, Device C, and Device D.

```
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface GigabitEthernet 0/2
```

```
DeviceA(cfg-if-GigabitEthernet 0/2)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
DeviceA(cfg)# interface GigabitEthernet 0/3
DeviceA(cfg-if-GigabitEthernet 0/3)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/3)# exit
DeviceA(cfg)# interface GigabitEthernet 0/4
DeviceA(cfg-if-GigabitEthernet 0/4)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/4)# exit
DeviceA(cfg)# interface loopback 0
DeviceA(cfg-if-Loopback 0)# ipv6 pim sparse-mode
DeviceA(cfg-if-Loopback 0)# exit
```

Configure an RP. Configure a C-BSR and a C-RP on Device A.

```
DeviceA(cfg)# ipv6 pim c-bsr Loopback 0
DeviceA(cfg)# ipv6 pim c-rp Loopback 0
```

(3) Configure a static multicast group on the interfaces of the devices on shared network N1.

Configure a static multicast group on SVI 50 of Device B, and set the static multicast group address to FF13::1. Perform the same configuration on Device D.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface VLAN 50
DeviceB (cfg-if-VLAN 50)# ipv6 mld static-group-entries ff13::1
DeviceB(cfg-if-VLAN 50)# exit
DeviceB(cfg)# interface GigabitEthernet 0/23
DeviceB(cfg-if-GigabitEthernet 0/23)# switch-port access vlan 50
DeviceB(cfg-if-GigabitEthernet 0/23)# exit
DeviceB(cfg)# interface TenGigabitEthernet 0/27
DeviceB(cfg-if-TenGigabitEthernet 0/27)# switch-port mode trunk
DeviceB(cfg-if-TenGigabitEthernet 0/27)# switch-port trunk native vlan 50
DeviceB(cfg-if-TenGigabitEthernet 0/27)# exit
```

(4) Enable the multicast group access control function on the interfaces of the devices on shared network N2.

Enable the multicast group access control function on SVI 60 of Device C, and add the hosts in network N2 to the multicast group FF14::1.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ipv6 access-control-list ipv6-group
DeviceC(cfg-ipv6acl)# permit ipv6 any host FF14::1
DeviceC(cfg-ipv6acl)# exit
DeviceC(cfg)# interface VLAN 60
DeviceC(cfg-if-VLAN 60)# ipv6 mld restricted-group ipv6-group
DeviceC(cfg-if-VLAN 60)# exit
```

5. Verification

Run the **display ipv6 mld-info interface** command to display MLD configuration of each device. Device B is taken as an example. Device B is the querier, because the IPv6 address of SVI 50 on Device B in the shared network N1 is smaller.

```
DeviceB# display ipv6 mld-info interface VLAN 50
Interface VLAN 50 (Index 4146)
 MLD Enabled, Active, Querier, Version 2 (default)
 Internet address is fe80::274:9cff:feee:53cb
 MLD interface max-member-record is 4000
 MLD interface has 0 group-record states
 MLD interface has 0 add-to-group records
 MLD interface has 1 static-group-entries records
 MLD activity: 10 joins, 1 leaves
 MLD query interval is 125 seconds
 MLD querier timeout is 255 seconds
 MLD max query response time is 10 seconds
 Last member query response interval is 10 (1/10s)
 Last member query count is 2
 Group Membership interval is 260
 Robustness Variable is 2
```

Run the **display ipv6 pim-info sparse-mode mroute** command to display the multicast routing entries on each device. The multicast routing entries on Device B are as follows:

```
DeviceB# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 1
(S,G) Entries: 0
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(*, ff13::1)
RP: 1111::1
RPF nbr: fe80::274:9cff:fe11:c9e7
RPF idx: GigabitEthernet 0/24
Upstream State: NOT JOINED
 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .i  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
7 . . . . . . . . . . .. . .. . .. . . . . . . . . . . . . . .
Joined
0 . . . . . . . ... .. . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . .. . . . .. . . . .. . . . . . . . .
2 . . . .. . . .. . . . .. . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . .. . . .. . . . . .. . . . . . . . . . . .
4 . . . . . . . .. . . . .. . . . . .. . . . . . . . . . . . .
5 . . . . .. . . . . . . . .. . . . .. . . . . . . . . . . . .
6 . . . .. . . . . . . .. . . . . . . . .. . . . . . . . . . .
7 . . . .. . . . . . .. . . . . . . .. . . . . . . . . . . . .
Asserted
0 . . . . . . . .. . . . .. . . . . .. . . . . . . . . . . . .
1 . . . . . . .. . . . .. . . . . . .. . . . . . . . . . . . .
2 . . . . . . .. . . . . .. . . . . .. . . . . . . . . . . . .
3 . . . . . . .. . . . . .. . . . . .. . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . .. . .. . . . . . .. . .
5 . . . .. . . . . . . . . .. . . . .. . . . . . . . . . . . .
6 . . . .. . . .. . . .. . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . .. .. . . . .. . . . . . . . . . . . . . .
FCR:
```

Run the **display ipv6 mld-info groups** command to display the static multicast group FF13::1 that SVI50 joins on Device B and Device D.

```
DeviceB# display ipv6 mld-info groups
MLD Connected Group Membership
Group Address                 Interface        Uptime     Expires    Last Reporter
ff13::1                      VLAN 50           00:11:07   stopped    ::
```

Try to add a host to the multicast groups FF13::1 and FF14::1 through SVI 60, and run the **display ipv6 mld-info groups** command to display that the host joins in only the multicast group FF14::1 through SVI 60.

```
DeviceC# display ipv6 mld-info groups
MLD Connected Group Membership
Group Address                 Interface        Uptime     Expires    Last Reporter
ff14::1                      VLAN 60           00:01:03   00:03:51
fe80::274:9cff:feee:f49f
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ipv6 address 3000::1/64
 ipv6 ospf 5050 set-area 0
```

```
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ipv6 address 2000::1/64
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/3
 unset switch-port
 ipv6 address 4000::1/64
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/4
 unset switch-port
 ipv6 address 7000::1/64
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 0
 ipv6 address 1111::1/128
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 5050
 graceful-restart
!
ipv6 pim c-bsr Loopback 0
ipv6 pim c-rp Loopback 0
!
end
```

- Device B configuration file

```
hostname DeviceB
!
ipv6 mrouting-enable
!
vlan range 1,50
!
interface GigabitEthernet 0/23
 switch-port access vlan 50
!
interface GigabitEthernet 0/24
 unset switch-port
 ipv6 address 2000::2/64
 ipv6 ospf 5050 set-area 0
```

```
 ipv6 pim sparse-mode
!
interface TenGigabitEthernet 0/27
 switch-port mode trunk
 switch-port trunk native vlan 50
!
interface Loopback 0
 ipv6 address 2222::2/128
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface VLAN 50
 ipv6 address 5000::1/64
 ipv6 ospf 5050 set-area 0
 ipv6 mld static-group-entries ff13::1
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 5050
 graceful-restart
!
End
```

- Device C configuration file

```
hostname DeviceC
!
ipv6 access-control-list ipv6-group
 10 permit ipv6 any host FF14::1
!
ipv6 mrouting-enable
!
vlan range 1,60
!
interface GigabitEthernet 0/23
 switch-port access vlan 60
!
interface GigabitEthernet 0/24
 unset switch-port
 ipv6 address 7000::2/64
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 0
 ipv6 address 3333::3/128
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface VLAN 60
```

```
 ipv6 address 6000::1/64
 ipv6 mld restricted-group ipv6-group
 ipv6 mld add-to-group ff13::1
 ipv6 mld add-to-group ff14::1
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 5050
 graceful-restart
!
end
```

- Device D configuration file

```
hostname DeviceD
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ipv6 address 4000::2/64
 ipv6 ospf 5050 set-area 0
!
interface TenGigabitEthernet 0/51
 switch-port mode trunk
 switch-port trunk native vlan 50
!
interface Loopback 0
 ipv6 address 4444::4/128
!
interface VLAN 50
 ipv6 address 5000::2/64
 ipv6 ospf 5050 set-area 0
 ipv6 mld static-group-entries ff13::1
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 5050
 graceful-restart
!
end
```

7. Common Errors

- The multicast routing function is not enabled on a multicast device.
- No multicast interface is available on the network.

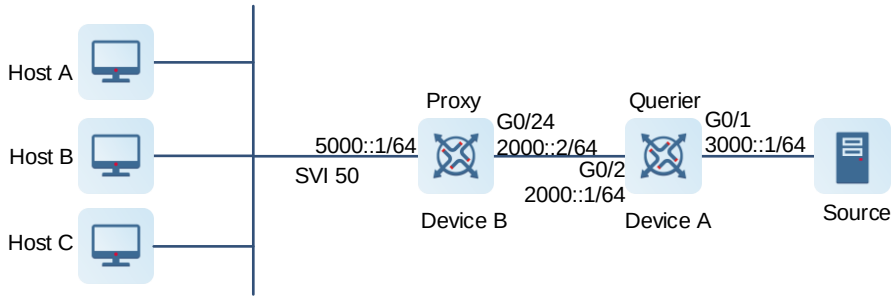
1.14.2 Configuring the MLD Proxy Function

1. Requirements

As shown in Figure 1-9, the PIM-SMv6 protocol is configured in the network. Host A and Host B in the stub network receive multicast data from the source. The multicast group addresses are FF13::1 and FF14::1. The MLD proxy function is enabled on Device B to maintain membership and normal forwarding of multicast data.

2. Topology

Figure 1-9 Topology of the MLD Proxy



3. Notes

- Configure an IPv6 unicast routing protocol OSPFv3 on a multicast device and ensure that unicast routes are reachable to the Loopback interface.
- Enable the IPv6 multicast routing function on all multicast devices, and configure the PIM-SMv6 protocol on the interconnected interfaces of devices, interfaces connected to hosts, and interfaces connected to sources.
- Configure a proxy service interface.
- Configure an MRoute proxy interface.

4. Procedure

(1) Configure an IPv6 unicast routing protocol OSPFv3 on a multicast device and ensure that unicast routes are reachable to the Loopback interface.

Configure an IPv6 address for the interface of Device A and configure the OSPFv3 protocol to implement internetworking. Perform the same configuration on Device B.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6 router-protocol ospf 5050
DeviceA(cfg)# interface GigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 address 3000::1/64
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface GigabitEthernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/2)# ipv6 address 2000::1/64
```

```
DeviceA(cfg-if-GigabitEthernet 0/2)# ipv6 ospf 5050 area 0
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
DeviceA(cfg)# interface loopback 0
DeviceA(cfg-if-Loopback 0)# ipv6 address 1111::1/128
DeviceA(cfg-if-Loopback 0)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-Loopback 0)# exit
```

(2) Enable the IPv6 multicast routing function on all multicast devices, and configure the PIM-SMv6 protocol on the interconnected interfaces of devices, interfaces connected to hosts, and interfaces connected to sources.

Enable the IPv6 multicast routing function and configure the PIM-SMv6 protocol on Device A. Perform the same configuration on Device B.

```
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface GigabitEthernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
DeviceA(cfg)# interface loopback 0
DeviceA(cfg-if-Loopback 0)# ipv6 pim sparse-mode
DeviceA(cfg-if-Loopback 0)# exit
```

Configure an RP. Configure a C-BSR and a C-RP on Device A.

```
DeviceA(cfg)# ipv6 pim c-bsr Loopback 0
DeviceA(cfg)# ipv6 pim c-rp Loopback 0
```

(3) Configure a proxy service interface.

Configure GigabitEthernet 0/24 on Device B as a proxy service interface.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/24
DeviceB(cfg-if-GigabitEthernet 0/24)# ipv6 mld enable-proxy-service
DeviceB(cfg-if-GigabitEthernet 0/24)# exit
```

(4) Configure an MRoute proxy interface.

Configure SVI 50 of Device B as an MRoute proxy interface.

```
DeviceB(cfg)# interface VLAN 50
DeviceB(cfg-if-VLAN 50)# ipv6 mld enable-mrouting-proxy GigabitEthernet 0/24
```

5. Verification

Run the **display ipv6 mld-info groups** command on Device B to display the proxy entries.

```
DeviceB# display ipv6 mld-info groups
MLD Connected Group Membership
Group Address                    Interface          Uptime     Expires    Last Reporter
ff13::1                          VLAN 50            00:03:24   00:04:14
fe80::274:9cff:feee:f49f
```

```
ff14::1                                  VLAN 50                 00:00:06   00:04:14
fe80::274:9cff:feee:54cb

MLD Proxy-server Connected Group Membership
Group Address                              Interface                Uptime
ff13::1                                  GigabitEthernet 0/24    00:03:24
ff14::1                                  GigabitEthernet 0/24    00:00:06
```

Run the **display ipv6 mld-info groups** command on Device A to display that addresses of members joining the multicast groups FF13::1 and FF14::1 are IPv6 addresses of the Device B proxies.

```
DeviceA# display ipv6 mld-info groups
MLD Connected Group Membership
Group Address                              Interface                Uptime      Expires     Last Reporter
ff13::1                                  GigabitEthernet 0/2     00:09:56   00:03:44
fe80::274:9cff:feee:53cb
ff14::1                                  GigabitEthernet 0/2     00:00:23   00:03:56
fe80::274:9cff:feee:53cb
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ipv6 address 3000::1/64
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/2
 unset switch-port
 ipv6 address 2000::1/64
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 0
 ipv6 address 1111::1/128
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 5050
 graceful-restart
!
ipv6 pim c-bsr Loopback 0
ipv6 pim c-rp Loopback 0
```

```
!
End
```

- Device B configuration file

```
hostname DeviceB
!
ipv6 mrouting-enable
!
vlan range 1,50
!
interface GigabitEthernet 0/23
 switch-port access vlan 50
!
interface GigabitEthernet 0/24
 unset switch-port
 ipv6 address 2000::2/64
 ipv6 ospf 5050 set-area 0
ipv6 mld enable-proxy-service
 ipv6 pim sparse-mode
!
interface TenGigabitEthernet 0/27
 switch-port mode trunk
 switch-port trunk native vlan 50
!
interface Loopback 0
 ipv6 address 2222::2/128
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface VLAN 50
 ipv6 address 5000::1/64
 ipv6 ospf 5050 set-area 0
ipv6 mld enable-mrouting-proxy GigabitEthernet 0/24
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 5050
 graceful-restart
!
End
```

7. Common Errors

The basic functions of MLD are not enabled.

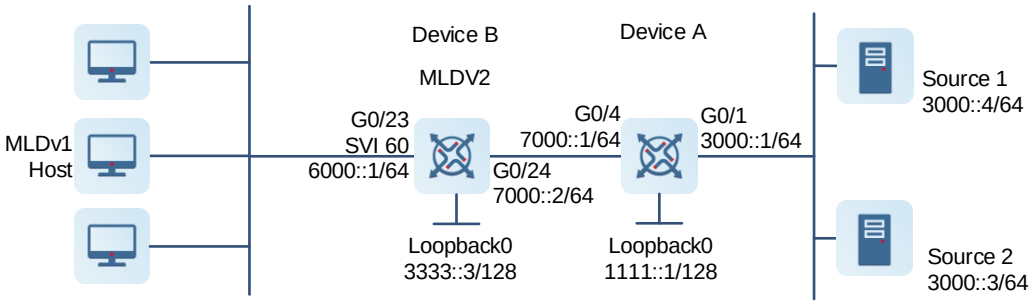
1.14.3 Configuring SSM Mapping

1. Requirements

As shown in Figure 1-10, the SSM model is used to provide multicast service in an IPv6 multicast network. The default SSM group address range in the network is FF3X::/32. Hosts run MLDv1 and do not support MLDv2. Hosts are expected to receive multicast packets from Source 2 and ignore multicast packets from Source 1.

2. Topology

Figure 1-10 Topology of SSM Mapping



3. Notes

- Configure an IPv6 unicast routing protocol OSPFv3 on a multicast device and ensure that unicast routes are reachable to the Loopback interface.
- Enable the IPv6 multicast routing function on all multicast devices, and configure the PIM-SMv6 protocol on the interconnected interfaces of devices, interfaces connected to hosts, and interfaces connected to sources.
- Configure an SSM group address range.
- Enable the SSM mapping function.
- Configure an SSM mapping static linked list.

4. Procedure

(1) Configure an IPv6 unicast routing protocol OSPFv3 on a multicast device and ensure that unicast routes are reachable to the Loopback interface.

Configure an IPv6 address for the interface of Device A and configure the OSPFv3 protocol to implement internetworking. Perform the same configuration on Device B.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6 router-protocol ospf 5050
DeviceA(cfg)# interface GigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 address 3000::1/64
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface GigabitEthernet 0/4
```

```
DeviceA(cfg-if-GigabitEthernet 0/4)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/4)# ipv6 address 7000::1/64
DeviceA(cfg-if-GigabitEthernet 0/4)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/4)# exit
DeviceA(cfg)# interface loopback 0
DeviceA(cfg-if-Loopback 0)# ipv6 address 1111::1/128
DeviceA(cfg-if-Loopback 0)# ipv6 ospf 5050 set-area 0
DeviceA(cfg-if-Loopback 0)# exit
```

(2) Enable the IPv6 multicast routing function on all multicast devices, and configure the PIM-SMv6 protocol on the interconnected interfaces of devices, interfaces connected to hosts, and interfaces connected to sources.

Enable the IPv6 multicast routing function and configure the PIM-SMv6 protocol on Device A. Perform the same configuration on Device B.

```
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface GigabitEthernet 0/4
DeviceA(cfg-if-GigabitEthernet 0/4)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/4)# exit
DeviceA(cfg)# interface loopback 0
DeviceA(cfg-if-Loopback 0)# ipv6 pim sparse-mode
DeviceA(cfg-if-Loopback 0)# exit
```

Set the MLD version on SVI 60 connected to hosts to 2 on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface VLAN 60
DeviceB(cfg-if-VLAN 60)# ipv6 mld specify-version 2
DeviceB(cfg-if-VLAN 60)# exit
```

(3) Configure an SSM group address range.

Configure an SSM multicast address range on all devices. Take Device A as an example and perform the same configuration on Device B.

```
DeviceB(cfg)# ipv6 pim enable-ssm default
```

(4) Enable the SSM mapping function.

Enable the SSM mapping function on Device B.

```
DeviceB(cfg)# ipv6 mld ssm-policy enable
```

(5) Configure an SSM mapping static linked list.

Map the multicast addresses in FF34::/64 to Source 1 on Device B.

```
DeviceB(cfg)# ipv6 access-control-list map-acl
DeviceB(cfg-ipv6acl)# permit ipv6 any FF34::/64
DeviceB(cfg)# ipv6 mld ssm-policy static map-acl 3000::3
```

5. Verification

Display the SSM mapping information.

```
DeviceB# display ipv6 mld-info ssm-mapping
SSM Mapping : Enabled
Database      : Static mappings configured
```

Display multicast source information corresponding to the multicast group FF34::1.

```
DeviceB# display ipv6 mld-info ssm-mapping FF34::1
Group address: ff34::1
Database        : Static
Source list   : 3000::3
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ipv6 address 3000::1/64
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/4
 unset switch-port
 ipv6 address 7000::1/64
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 0
 ipv6 address 1111::1/128
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 5050
 graceful-restart
!
ipv6 pim enable-ssm default
!
end
```

- Device B configuration file

```
hostname DeviceB
!
ipv6 access-control-list map-acl
```

```
 30 permit ipv6 any FF34::/64
!
ipv6 mrouting-enable
!
vlan range 1,60
!
interface GigabitEthernet 0/23
 switch-port access vlan 60
!
interface GigabitEthernet 0/24
 unset switch-port
 ipv6 address 7000::2/64
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 0
 ipv6 address 3333::3/128
 ipv6 ospf 5050 set-area 0
 ipv6 pim sparse-mode
!
interface VLAN 60
 ipv6 address 6000::1/64
 ipv6 ospf 5050 area 0
 ipv6 pim sparse-mode
 ipv6 mld specify-version 2
!
ipv6 router-protocol ospf 5050
 graceful-restart
!
ipv6 mld ssm-policy enable
ipv6 mld ssm-policy static map-acl 3000::3
ipv6 pim enable-ssm default
!
end
```

7. Common Errors

- The basic functions of MLD are not enabled.
- The SSM service is not enabled.
- No SSM address range is configured.

1.15 Common Misconfigurations

- Basic functions of MLD are not configured on the proxy service interface.
- No SSM service address is configured when the SSM service is used.
- The SSM service address ranges on different multicast devices are inconsistent.

1.15.1 What Can I Do when On-demand Multicast Service Is Unavailable on Client

- Fault Symptom

  The multicast function is enabled. But I failed to use the on-demand multicast service through a client.

- Troubleshooting

  a Check whether the connection between the on-demand client and multicast server is normal.

  Run the **display ipv6 route** command to check for any entry of unicast route to the multicast source on the server. If no, configure a unicast routing protocol.

  b Check configuration on the devices.

  Run the **display ipv6 pim-info** { **dense-mode | sparse-mode** } **interface** command to check for PIM information on an interface. If no information about the interface is displayed, the PIM-SMv6 function is not enabled. Run the **ipv6 pim sparse-mode** command to enable the PIM-SMv6 function on the interface.

  c Check whether the multicast receiver normally joins the group.

  Run the **display ipv6 mld-info groups** command to check whether the host joins the corresponding multicast group on the multicast device residing in the same network segment as the host.

  d Check whether the multicast entries on the multicast source and receiver are normal.

  Run the **display ipv6 pim-info** { **dense-mode | sparse-mode** } command to check whether the multicast entries on the devices are normal.

Contents

1 Configuring PIM-SMv6 1
1.1 Introduction 1
1.2 Principles 1
1.2.1 ASM Model 1
1.2.2 SSM Model 6
1.2.3 Protocols and Standards 7
1.3 Restrictions and Guidelines 8
1.4 Configuration Task Summary 8
1.5 Enabling the PIM-SMv6 Function 8
1.5.1 Overview 8
1.5.2 Restrictions and Guidelines 8
1.5.3 Prerequisites 9
1.5.4 Procedure 9
1.6 Configuring PIM-SMv6 in the ASM Model 9
1.6.1 Overview 9
1.6.2 Configuration Tasks 10
1.6.3 Prerequisites 10
1.6.4 Configuring a Static RP 10
1.6.5 Configuring C-BSRs and Related Parameters 11
1.6.6 Configuring C-RPs and Related Parameters 12
1.6.7 Configuring Registration Parameters 15
1.6.8 Configuring the Last Hop Device to Switch over from the RPT to SPT 18

1.7 Configuring PIM-SMv6 in the SSM Model 18
1.7.1 Overview 18
1.7.2 Restrictions and Guidelines 18
1.7.3 Prerequisites 18
1.7.4 Procedure 19
1.8 Configuring PIM-SMv6 Neighbor Parameters 19
1.8.1 Overview 19
1.8.2 Configuration Tasks 19
1.8.3 Prerequisites 19
1.8.4 Configuring the Neighbor Message Time Parameters 19
1.8.5 Configuring Neighbor Tracking 20
1.8.6 Configuring Neighbor Filtering 21
1.9 Configuring DR Priority 22
1.9.1 Overview 22
1.9.2 Prerequisites 22
1.9.3 Procedure 22
1.10 Configuring Parameters of the Join/Prune Packets 22
1.10.1 Overview 22
1.10.2 Restrictions and Guidelines 23
1.10.3 Configuration Tasks 23
1.10.4 Prerequisites 23
1.10.5 Configuring Join/Prune Packet Sending Interval 23
1.10.6 Configuring the Prune Delay 24
1.11 Enabling PIM-SMv6 BFD Correlation 25

1.12 Monitoring 25

1.13 Configuration Examples 26

1.13.1 Configuring Basic Functions of PIM-SMv6 in the ASM Model 26

1.13.2 Configuring Basic Functions of PIM-SMv6 in the SSM Model 49

1 Configuring PIM-SMv6

1.1 Introduction

The Protocol Independent Multicast (PIM) is an intra-domain multicast routing protocol that is independent from unicast routing protocols. Any unicast routing protocol, for example, static route, Routing Information Protocol (RIP), and Open Shortest Path First (OSPF), can provide routing information for IP multicast. Multicast routing is independent of unicast routing, and simply uses unicast routing tables to generate multicast routing information. The Protocol Independent Multicast-Sparse Mode version 6 (PIM-SMv6) is used most widely, and applies to networks of any form, especially large networks where group members are scattered in a wide area. Compared with other intra-domain multicast routing protocols, PIM-SMv6 has irreplaceable advantages.

1.2 Principles

Suppose that not all hosts need to receive multicast data, PIM-SMv6 sends multicast data to hosts that require the multicast data.

A PIM multicast device provides multicast service in any-source multicast (ASM) and source-specific multicast (SSM) models. For groups in the SSM address range, the SSM model is used. For other groups, the ASM model is used. In the ASM model, a host joins a specified multicast group G without the multicast source S specified. In the SSM model, a host can join a specified multicast group G to receive multicast data from a specified source S.

1.2.1 ASM Model

1. Basic Principles

PIM-SMv6 in the ASM model is implemented as follows:

- A rendezvous point (RP) is used as a root of a rendezvous point tree (RPT) for transition. The designated router (DR) connected to a host sends a join message to the RP of a multicast group, and the multicast devices on the route create (*, G) entries hop by hop. Based on the entries, an RPT is created with the RP as the root.
- To send data to a multicast group, the source DR sends a register message to the RP to create an (S, G) entry on the RP. After the registration succeeds, the multicast source sends multicast data to the RP. After the multicast data arrives at the RP, it is copied and sent to the host via the RPT.

In the ASM model, the key mechanisms of PIM-SMv6 include PIM neighbor creation, DR election, RP, bootstrap router (BSR), multicast source registration, shortest path tree (SPT) switchover, and assertion.

2. PIMv6 Neighbor Creation

Neighbor relationship must be created between PIM multicast devices to form a PIM network, so that PIM control messages and multicast packets can be transmitted on the network. A Hello message is sent from a PIM interface. In the multicast packet with the Hello message encapsulated, the destination address is FF02::D (indicating all PIM multicast devices in the same network segment), the source address is the IP address of the PIM interface, and the time to live (TTL) value is 1. The Hello message is used to:

- Discover PIM neighbors.

  PIM multicast devices in the same network segment receive the multicast packet with the destination address as FF02::D, obtain neighbor information, and create neighbor relationship. After the neighbor relationship is created, the PIM interfaces can receive and process other PIM packets and create multicast routing entries.

  When a PIM interface is enabled or detects a new neighbor, a Triggered-Hello-Delay message is used to generate a random time. Within the time, the interface sends Hello messages.

- Negotiate protocol parameters. The Hello message contains the following key parameters:
  - **DR Priority**: The priority of a multicast device for DR election. The device with a higher priority is selected.
  - **Holdtime**: The neighbor is considered reachable before timeout.
  - **Propagation Delay**: Duration between the time when a PIM device receives a prune packet and the time when the PIM device sends the prune packet to the uplink devices.
  - **Override-Interval**: The interval to override the prune.

- Maintain neighbor relationship.

  A Hello message is sent periodically between PIM multicast devices. If no new Hello message is received from a PIM neighbor within the hold time, this neighbor is considered unreachable and then removed from the neighbor list. Any change of PIM neighbors will cause change of the multicast topology in the network. If an uplink or downlink neighbor in a multicast distribution tree (MDT) is unreachable, multicast routes converge again and the MDT is reshaped.

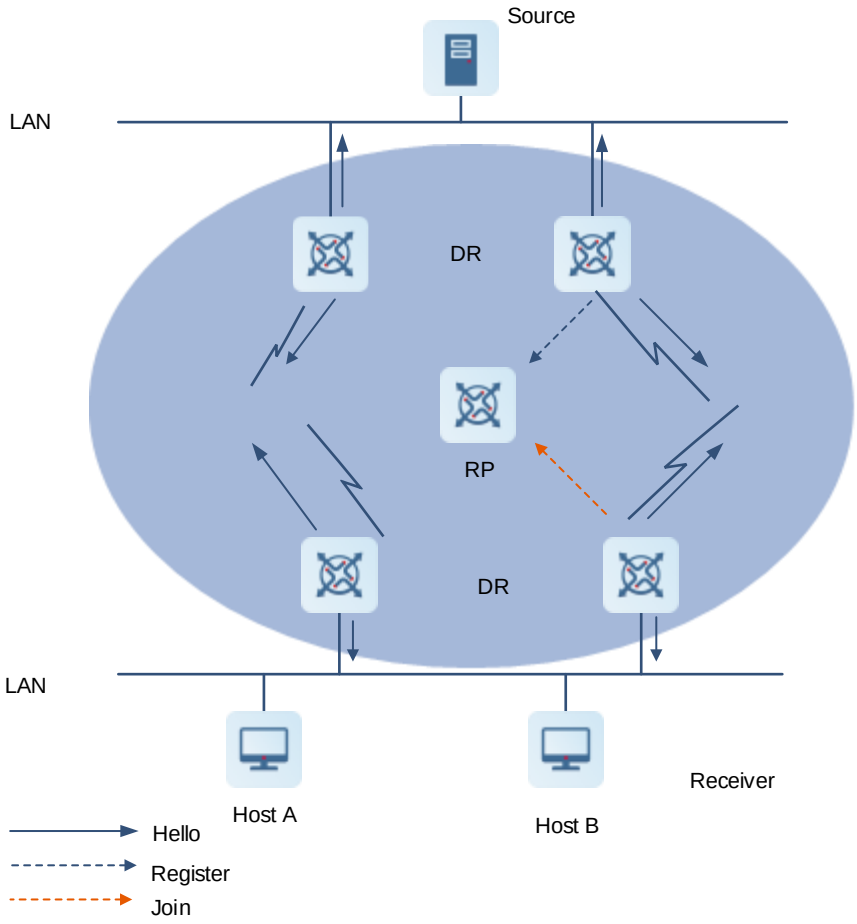
3. DR Election

As shown in Figure 1-1, the source DR sends a register message to the RP. Member DRs send join messages to the RP and forward multicast data to members in a multicast group. In the network segment to which members and multicast source are connected, multiple PIM devices obtain IP addresses and DR priorities of neighbors based on the Hello messages sent between the PIM devices during PIM neighbor creation. The PIM device that is elected becomes a member DR or source DR. The election rules are as follows:

(1) The DR priorities are compared. The device with a higher priority is selected.

(2) If the DR priorities are the same or there is any neighbor in the shared network segment that does not support DR priority, the IP addresses of the devices are compared. The device with a larger IP address is selected.

Figure 1-1 Schematic Diagram of DR Election



4. BSR Election

In the PIM network, the BSR periodically generates bootstrap messages (BSMs) that include a series of candidate-RPs (C-RPs) and corresponding group addresses. The BSMs are sent in the entire domain hop by hop. All PIM multicast devices in the network receive the BSMs and record the C-RPs and corresponding group addresses.

In the PIM-SMv6 domain, one or multiple candidate-BSRs (C-BSRs) are configured and the BSR is elected based on rules.

5. RP Election

In the PIMv6 network, the RP processes register messages of the source DR and join packets of group members. The process from the multicast source to members of the multicast group is divided into two sections: The SPT that uses the source DR as the root and the RP as the leaf, and the RPT that uses the RP as the root and member DRs as leaves. In the PIM-SMv6 network with the ASM model, the RP must act as a transition to create or forward data.

In the PIM network, an RP can serve one or more groups, and a group can be served by only one RP. In the PIM network, the RP can be configured as static or dynamic.

- Static RP

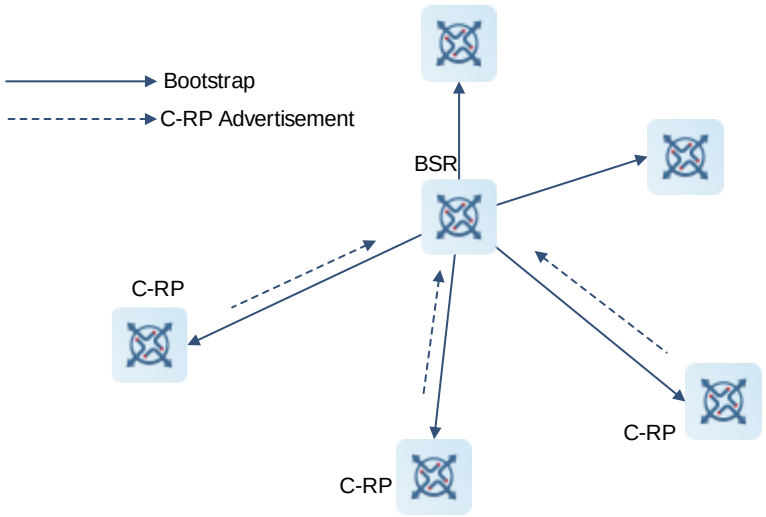
A static RP sets the RP address on each PIM multicast device to enable the entire PIMv6 network to know the RP address.

- Dynamic RP

  Multiple C-RPs can be configured in a PIM-SMv6 domain to elect an RP dynamically. These C-RPs add their own addresses and served multicast groups to advertisement packets and sent the packets to the BSR periodically in unicast mode. The BSR is elected from the configured C-BSRs based on rules. The BSR periodically generates BSMs that include a series of C-RPs and corresponding group addresses. The BSMs are sent in the entire domain. Each PIM device in the domain receives and saves these BSMs, and applies the same hash algorithms to compute and elect an RP for each group.

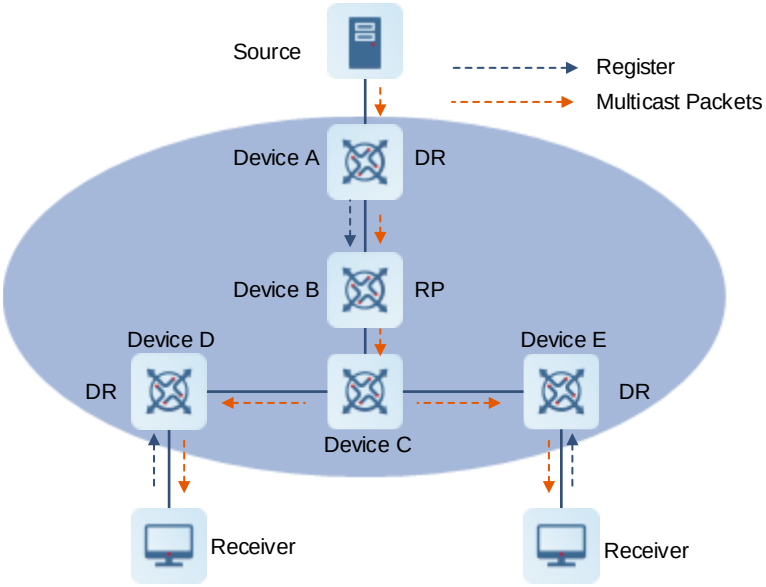
  - BSR election process: Initially, each C-BSR considers itself a BSR and sends a BSM throughout the network. The BSMs carry the C-BSR addresses and priorities. Upon receiving the BSMs, each PIM device compares the C-BSR addresses and priorities to elect the BSR.
  - RP election process: The C-RPs periodically send advertisement packets to the BSR in unicast mode to advertise the C-RP addresses, priorities, and addresses of the served groups. The BSR collects the information as RP-Set, encapsulates the information in a BSM, and sends the messages to each PIM device in the network. Upon receiving the RP-Set, the PIM devices in the network use the same algorithms for computation to obtain the RP that serves each group. Each PIM device obtains the RP position of each group. Figure 1-2 shows the schematic diagram of dynamic RP election.

Figure 1-2 Schematic Diagram of Dynamic RP Election



6. Multicast Source Registration

Figure 1-3 Schematic Diagram of Multicast Source Registration



As shown in Figure 1-3, when a multicast source sends data to a group, the source DR sends a register message to the RP so that the RP obtains the source information and a multicast packet and forwards the multicast packet to the members in the group. The forwarding process is described as follows:

(1) The multicast source sends a multicast packet to the source DR.

(2) The source DR encapsulates the multicast packet in a register message and sends the message to the RP in unicast mode.

(3) Upon receiving the register message, the RP decapsulates the message and forwards the multicast packet to the members in a group via the RPT. Meanwhile, the RP sends an (S, G) join packet to the source. The PIM devices on the route create (S, G) entries, and finally create an SPT from the source to the RP.

(4) The source DR S forwards multicast packets based on this SPT. The multicast data is sent to the RP in the form of multicast packets.

(5) Upon receiving multicast packets from the source, the RP discards the register message and sends a register-stop message to the source.

(6) Upon receiving the register-stop message, the source DR stops encapsulating the register message, forwards multicast packets via the SPT, and periodically sends a register-probe message to the RP.

(7) Upon receiving the register-probe message, the RP sends a register-stop message to the source DR and updates the (S, G) entries.

(8) Upon receiving the register-stop message, the source DR continues forwarding multicast packets via the SPT without encapsulating the register message.

7. RPT Construction

When a group member joins the network, the DR connecting to the group member sends a join packet to the RP. If a multicast source has sent a multicast packet to the RP, the multicast packet will be sent to the group member via the RPT.

The RPT construction procedure is as follows:

- When a member DR receives an MLD (*, G) INCLUDE Report packet, if the member DR is not an RP of the group G, the member DR sends a (*, G) join packet to the RP and the multicast devices on the route create the (*, G) entries hop by hop to generate an RPT that uses the RP as the root.
- During RPT construction, each PIM device checks the reverse path forwarding (RPF) when sending a join packet to the RP to find the unicast route to the RP. The outbound interface of the unicast route is an uplink interface, and the next hop of the unicast route is the RPF neighbor. The join packet is sent from the member DR to the RP.

8. SPT Switchover

In the PIM-SMv6 network of the ASM model, one or multiple multicast groups correspond to one RP. Before SPT switchover, all packets of multicast groups sent to the RP service must be transitioned by the RP. When the sending rate of multicast packets grows, the RP load is increased. To solve this problem, the PIM-SMv6 relieves the RP load by SPT switchover. SPT switchover can be triggered by the RP or a member DR.

- SPT switchover triggered by the RP

  After receiving a register message from the source DR, the RP decapsulates the register message, forwards the multicast packets in the message to the group members via the RPT, and sends an (S, G) join packet to the source DR. The PIM multicast devices on the route create (S, G) entries hop by hop to create an SPT from the RP to the source.

- SPT switchover triggered by a member DR

  a When receiving the first multicast packet from the RP, the member DR obtains the source address.

  b The member DR sends an (S, G) join packet to the source. The PIM devices on the route create (S, G) entries and create an SPT from the member DR to the source DR.

  c After the SPT is created, the member DR sends a prune packet to the RP hop by hop and removes the downlink interface from the (*, G) entries. After the pruning, the RP does not forward multicast packets to the member DR via the RPT.

  d The SPT from the member DR to the source DR does not pass the RP. The RP sends a prune packet to the source DR to remove the downlink interface from the (S, G) entries. After the pruning, the source DR does not forward packets to the RP via the original "RP-source DR" SPT.

1.2.2 SSM Model

1. Basic Principles

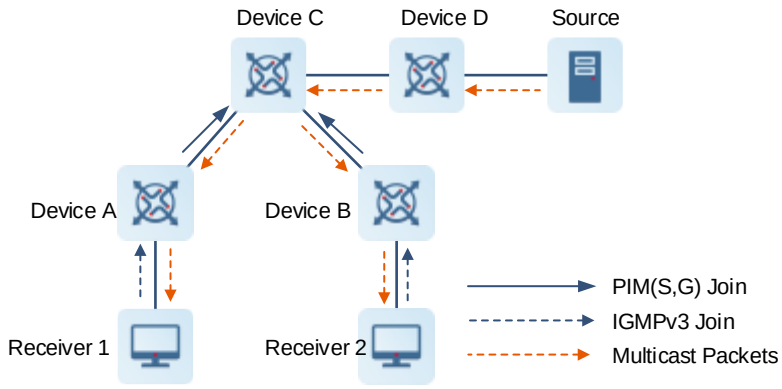
In the SSM model, members in a multicast group can specify to receive or reject data from a multicast source. The group members get to know the (S, G) entries of sources through access to the server or receiving advertisements. After understanding requirement of group members, the member DR sends a join packet to the source DR. This join packet is transmitted to the multicast source hop by hop to create an SPT from the source to the group members.

Note

- To implement the SSM model on IPv6, complete multicast interconnection through the PIM-SMv6 and configure MLDv2 to manage relationship between the hosts and devices.
- SSM model requirements: The hosts must know the addresses of multicast sources in advance using other network services so that the hosts can select multicast sources.

- On the devices that reside on the route from the hosts to the multicast source, the PIM-SMv6 and SSM functions must be enabled.

Figure 1-4 SSM Model



Caution

After the SSM function is enabled, the default group range of the SSM is FF3x::/32 (x cannot be 1 or 2). You can modify the group range of the SSM using commands.

In the SSM model, the key mechanisms of PIM-SMv6 include neighbor creation, DR election, and SPT creation.

2. PIMv6 Neighbor Creation

For more information about the PIMv6 neighbor creation mechanism in the SSM model, see the PIMv6 neighbor creation mechanism in the ASM model in 1.2.1 2. PIMv6 Neighbor Creation.

3. DR Election

For more information about the DR election mechanism in the SSM model, see the DR election mechanism in the ASM model in 1.2.1 3. DR Election.

4. SPT Construction

The SPT construction mechanism is as follows:

(1) The member DR learns through MLDv2 that a group member wants to receive data of a multicast source and sends an (S, G) join packet to the source hop by hop.

(2) The PIMv6 multicast devices on the route receive the (S, G) join packet and create the (S, G) entries, creating an SPT from the group member to the source.

(3) After the SPT is created, the multicast source distributes the multicast packets to the group member via the SPT.

1.2.3 Protocols and Standards

- RFC 4601: Protocol Independent Multicast-Sparse Mode (PIM-SM)
- RFC 5059: Bootstrap Router (BSR) Mechanism for Protocol Independent Multicast (PIM)
- RFC 3962: Protocol Independent Multicast-Dense Mode protocol

- RFC 4607: Source-Specific Multicast for IP

1.3 Restrictions and Guidelines

- Before PIM-SMv6 is enabled, enable the multicast routing and forwarding function in global configuration mode. Otherwise, multicast packets cannot be sent even if PIM-SMv6 is enabled.
- In the IPv6 multicast network, the PIM (IPv6) protocol must be run on all L3 devices. Otherwise, a forwarding route cannot be created.

1.4 Configuration Task Summary

The PIM-SMv6 configuration includes the following tasks:

(1) Enabling the PIM-SMv6 Function

(2) Configuring a model for PIM-SMv6. Select one of the following to configure:

- Configuring PIM-SMv6 in the ASM Model
- Configuring PIM-SMv6 in the SSM Model

(3) Configuring PIM-SMv6 Neighbor Parameters. All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the Neighbor Message Time Parameters
- Configuring Neighbor Tracking
- Configuring Neighbor Filtering

(4) (Optional) Configuring DR Priority

(5) Configuring Parameters of the Join/Prune Packets. All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring Join/Prune Packet Sending Interval
- Configuring the Prune Delay

(6) (Optional) Enabling PIM-SMv6 BFD Correlation

1.5 Enabling the PIM-SMv6 Function

1.5.1 Overview

Neighbor relationship must be created between PIM multicast devices to form a PIM-SMv6 network, so that PIM control messages and multicast packets can be transmitted on the network. The PIM-SMv6 function must be configured on all interfaces so that PIM neighbor creation, multicast entry creation, and multicast packet forwarding can be successful.

In a PIMv6 network, if you want an interface, for example, a stub network device interface connecting to hosts, to receive multicast packets without involving the PIMv6 network topology construction, you can set the interface to the PIM-SMv6 passive mode. After the PIM-SMv6 passive mode is configured, the interface does not process PIM packets. This improves device performance.

1.5.2 Restrictions and Guidelines

- The IPv6 multicast routing function must be enabled on each multicast device.

- The PIM-SMv6 function should be enabled on the following interfaces: interfaces interconnecting multicast devices, interfaces connecting the static RP, C-RP and C-BSRs, and interfaces connecting multicast sources or hosts.

1.5.3 Prerequisites

All PIM interfaces can be reached via IPv6 unicast routes.

1.5.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IPv6 multicast routing.

ipv6 mrouting-enable

The IPv6 multicast routing function is disabled by default.

(4) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(5) Enable the PIM-SMv6 function.

ipv6 pim sparse-mode

The PIM-SMv6 function is disabled by default.

(6) (Optional) Enable the PIM-SMv6 passive mode.

ipv6 pim sparse-mode passive

The PIM-SMv6 passive mode is disabled by default.

1.6 Configuring PIM-SMv6 in the ASM Model

1.6.1 Overview

The PIM-SMv6 network supports the multicast service in the ASM and SSM models. Either one model or two models can be configured.

PIM-SMv6 in the ASM model can provide hosts with multicast services of any source. A host joining a multicast group can receive multicast packets destined to this multicast group from any source.

If you require a PIMv6 network to support the multicast service in the ASM model, an RP must be configured. There are three methods to configure an RP:

- Configure a static RP.
- Configure a dynamic RP.
- Configure static and dynamic RPs. If static and dynamic RPs are available at the same time, dynamic RPs prevail.

1.6.2 Configuration Tasks

The PIM-SMv6 configuration in the ASM model includes the following tasks. Configure at least one RP, either dynamic or static. You can also configure both a static RP and a dynamic RP.

- Configuring a Static RP
- Configuring a Dynamic RP
  - Configuring C-BSRs and Related Parameters
  - Configuring C-RPs and Related Parameters
  - (Optional) Configuring Registration Parameters
- (Optional) Configuring the Last Hop Device to Switch over from the RPT to SPT

1.6.3 Prerequisites

In the PIM-SMv6 multicast network, all nodes are reachable via IPv6 unicast routes.

1.6.4 Configuring a Static RP

1. Overview

You can manually configure a static RP to prevent frequent packet interaction between the BSR and C-RPs or between the BSR and PIM multicast devices and avoid bandwidth occupation. If the multicast network is simple and small, you can configure a static RP.

2. Restrictions and Guidelines

- Static RPs must be configured on every multicast device.
- The multicast device selected as a static RP must boast high performance.
- You are advised to use the loopback interface as a static RP address.
- Static RPs configured on multicast devices should be consistent, including RP address and address range of the served groups.
- If multiple static RPs are configured to correspond to different addresses of groups, the RPs can implement load balancing.
- If multiple static RPs serve the same group, the static RP with a larger IPv6 address is selected first.
- You can use the control list to configure an RP to serve multiple (referencing an ACL) or all (not referencing an ACL) multicast groups. However, an RP can adopt only one configuration.

3. Prerequisites

Before an address range of a group is configured for an RP, ensure that the corresponding ACL is created and specify a multicast group range for the static RP to serve. For more information about ACL configuration, see *Configuring ACL*.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static RP.

ipv6 pim static-rp-addr *ipv6-rp-address* [*acl-name*]

No static RP is configured by default.

1.6.5 Configuring C-BSRs and Related Parameters

1. Overview

In the PIM-SMv6 domain of the ASM model, a unique BSR must be available. The BSR collects and releases RP information.

- The BSR is elected from C-BSRs based on rules. The election rules are as follows:
  a The priorities of the C-BSRs are compared. The device with a higher priority is selected.
  b If the priorities are the same, the IP addresses are compared. The device with a larger IP address is selected.

  The winner is elected as the BSR in the domain and periodically sends BSMs in the domain. A BSM contains the BSR IP address, RP-Set and hash mask information. The hash mask is used for hash calculation during C-RP election. In the hold time of a BSM, if other C-BSRs do not receive the BSM from the BSR, the BSR is considered unreachable and a new round of BSR election is initiated.
- To avoid fraud by an illegitimate BSR, a BSR range can be specified on all PIM devices in the domain so that the PIM devices ignore a BSM that is sent from an illegitimate BSR.
- To limit BSM flooding, you can configure a BSR border on an interface. Upon receiving BSMs, this interface discards the BSMs without forwarding them. This limits the BSM flooding.

2. Restrictions and Guidelines

- C-BSRs must be configured on one or multiple multicast devices.
- The multicast devices selected as C-BSRs must boast high performance. The C-RPs and C-BSRs can reside on the same or different multicast devices.
- You are advised to use the loopback interface as a C-BSR address.
- If multiple C-BSRs are configured, they can implement redundancy backup.
- You are advised to configure a valid BSR address range on all PIM devices in the domain.
- A BSR border must be configured on the interface between domains.

3. Prerequisites

- Enabling the PIM-SMv6 Function.
- Before a valid BSR address range is configured on PIM devices, ensure that the corresponding ACL is created and specify a valid BSR range. For more information about ACL configuration, see *Configuring ACL*.

4. Procedure

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Configure C-BSRs and related parameters.

ipv6 pim c-bsr *interface-type interface-number* [*hash-mask-length* [*priority-value*]]

No C-BSR is configured by default.

(2) (Optional) Configure a valid BSR address range on all PIM devices.

ipv6 pim accept-bsr list *acl-name*

All BSMs are received by default.

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) (Optional) Configure a BSR border on the interface between PIM-SMv6 domains.

ipv6 pim bsr-boundary

No BSR border is configured by default.

If multiple domains exist, the BSR border must be configured on the interface between two domains.

1.6.6 Configuring C-RPs and Related Parameters

1. Overview

- After a multicast device is configured as a C-RP, the C-RP periodically sends an advertisement packet to the BSR in unicast mode. The packet includes the RP priority, advertisement packet hold time (greater than the interval of sending the advertisement packet), RP address, address of the served group, and prefix of the multicast address. Upon receiving the advertisement packet, the BSR collects the information in the packet as RP-Set in the advertisement packet hold time, encapsulates the information in a BSM, and sends the message to all PIM devices. Upon receiving the BSM from the BSR, the PIM devices in the domain apply the same algorithms for hash calculation to calculate the C-RP serving each multicast group. After the

advertisement packet hold time expires, if the BSR does not receive a next advertisement packet from the C-RP, this C-RP is considered unreachable.

- To prevent fraud by an illegitimate C-RP, a C-RP range can be specified on the BSR. If an advertisement packet comes from an illegitimate C-RP, the packet is ignored.
- The BSR can determine whether to receive an advertisement packet with the prefix of the multicast address being 0 from a C-RP. When the prefix of the multicast address is 0, this C-RP serves all multicast groups.
- When receiving the BSM from the BSR, each PIM multicast device in the domain applies rules to select the RP. The RP election rules are as follows:
  a The address of the multicast group that a host joins is compared with the mask length of the group address range that the C-RP serves. The device with the longest mask is selected.
  b If the addresses are consistent with the mask length, the priorities of the C-RPs are compared. The device with the smallest priority value is selected.
  c If the priorities are consistent, hash values are calculated and compared. The device with the largest hash value is selected.
  d If the hash values are the same, the IP addresses of the C-RPs are compared. The device with the largest IP address is selected.

2. Restrictions and Guidelines

- C-RPs must be configured on one or multiple multicast devices.
- The multicast devices selected as C-RPs must boast high performance. The C-RPs and C-BSRs can reside on the same or different multicast devices.
- You are advised to use the loopback interface as a C-RP address.
- If multiple C-RPs are configured to serve the same group, they can implement redundancy backup.
- If multiple C-RPs are configured to serve different groups, they can implement load balancing.
- A valid C-RP range can be configured on the BSR.
- To specify a group range for C-RPs to serve, you can add the ACL option to the **ip pim rp-candidate** command. The group range is calculated based on the permit access control entry (ACE), rather than the deny ACE.

3. Prerequisites

- Enabling the PIM-SMv6 Function.
- If the **group-acl** parameter is used to limit the range of groups served by the RP, ensure that the corresponding ACL is created and specify a multicast group range for the RP to serve. For more information about ACL configuration, see *Configuring ACL*
- Before a valid C-RP address range is configured on PIM BSRs, ensure that the corresponding ACL is created and specify a multicast group range for the static RP to serve. For more information about ACL configuration, see *Configuring ACL*.

4. Procedure

- Configure a C-RP and related parameters.
  a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Configure a C-RP.

ipv6 pim c-rp *interface-type interface-number* [**priority** *priority-value*] [**interval** *interval-seconds*] [**group-acl** *acl-name*]

No C-RP is configured by default.

- (Optional) Configure the BSR to receive C-RP advertisement packets with the prefix of the multicast address being 0. This C-RP serves all the multicast groups.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Configure the BSR to receive advertisement packets from the C-RP that serves all multicast groups.

ipv6 pim accept-crp-with-null-group

The function for the BSR to receive advertisement packets from the C-RP that serves all multicast groups is disabled by default.

You are advised to enable this function on all C-BSRs.

- (Optional) Configure a valid C-RP range on the BSR.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Configure a valid C-RP range.

ipv6 pim accept-crp list *acl-name*

The C-BSRs receive all external C-RP advertisement packets by default.

You are advised to enable this function on all C-BSRs.

- (Optional) Configure all PIM multicast devices to ignore RP priority and reselect an RP.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Configure all PIM multicast devices to ignore the RP priority for RP election.

ipv6 pim ignore-rp-set-priority

If the address of the multicast group that a host joins is consistent with the mask length of the group address range that the C-RPs serve, the RP with the highest priority is selected by default.

You can configure each multicast device to ignore C-RP priority.

- (Optional) Configure static RP priority to be higher than dynamic RP priority on all PIM multicast devices.
  a Enter the privileged EXEC mode.

  privilege

  b Enter the global configuration mode.

  config-system terminal

  c Configure the static RP priority.

  ipv6 pim static-rp-preferred

  The dynamic RP priority is higher than the static RP priority by default.

  You can configure each multicast device to ignore dynamic RP priority.

1.6.7 Configuring Registration Parameters

1. Overview

When a multicast source sends multicast data, the data arrives at the source DR first. Upon receiving the multicast data, the source DR encapsulates the multicast data as a register message and sends the message to the RP. Related parameters of the register message are configured on the source DR and RP.

- You can configure the following parameters on the source DR:
  - RP reachability checking: Before sending a register message, the source DR determines whether the RP is reachable. If the RP is reachable, the message is sent. Otherwise, the message is not sent.
  - Limit on the rate of sending the register message: The rate of sending each register message in (S, G) status is limited. If multiple register messages in (S, G) status are sent from the same source, the rate of the register messages is limited to the product of *N* × Limit value.
  - Source address in the register message: The source address in the register message is the address of the DR interface connected to the multicast source by default. If a security policy is configured on the RP to filter out this address, the RP does not receive register messages from this source DR. In this case, you can specify a reachable source address as the source address of the register message.
  - Register suppression time: When the RP receives multicast packets sent by the multicast source via the SPT, the RP sends a register-stop message to the source DR. Within the register suppression time, the source DR stops encapsulating the register message and sends multicast packets to the RP via the SPT. After the register suppression time expires, if the source DR does not receive another register-stop message from the RP, the source DR continues to encapsulate the register message and send the message to the RP.
  - Null message detection time: Before the register suppression time expires, the RP receives a null message from the source DR, refreshes the local (S, G) entries, and sends a register-stop message to the source DR so that the source DR refreshes the register suppression timer. Otherwise, the source DR continues to encapsulate the register message and send multicast packets to the RP in the form of the register message.
- You can configure the following parameters on the RP:
  - (S, G) address range of source: Provide multicast services for hosts in the permitted multicast group in the ASM model. After this parameter is configured, the RP replies a register-stop message upon receiving data from an unauthorized source.

- Length of register message checksum: Adjust the calculation method of the register message checksum and calculate the checksum of the entire PIM protocol packets (including encapsulated multicast packets) or the checksum of PIM headers of separate register messages.
- Register message decapsulation and delivery to downlink interface: Configure this parameter on the RP to decapsulate the received register message and forward the multicast packets in the register message.
- Register message (S, G) entry timeout period: Configure this parameter to configure the timeout period of (S, G) entries. The (S, G) entries age if no packet is received within the register message timeout period.

2. Restrictions and Guidelines

- The source address in the specified register message must be reachable in unicast mode. You are advised to use the loopback interface of the source DR as the source address of the register message.
- If the **ip pimp-info rp-register-kat** command is not configured, modifying the register suppression time on the RP changes the (S, G) entry timeout period.
- Configuring a limit to the register message rate reduces the load of the source DR and RP. Only register messages under the limit are sent.
- The register-probe time cannot be greater than a half of the register suppression time. Otherwise, the configuration failed and an alarm is generated.
- The sum of the three times of register suppression time and the register-probe time does not exceed 65,535. Otherwise, the configuration failed and an alarm is generated.

3. Prerequisites

- Enabling the PIM-SMv6 Function.
- Before an (S, G) address range is configured for the source, if the **list** parameter is specified, ensure that the corresponding ACL is created and specify an (S, G) address range for the source. For more information about ACL configuration, see *Configuring ACL*
- Before an (S, G) address range is configured for the source, if the **route-map** parameter is specified, ensure that the corresponding route map is created and specify an (S, G) address range for the source. For more information about route map configuration, see *Configuring Routing Policies*.
- Before the register message checksum calculation is applied to entire packets, if the **group-acl** parameter is specified, ensure that the corresponding ACL is created and specify the register message checksum calculation to apply to the entire packets in a multicast group address range. For more information about ACL configuration, see *Configuring ACL*.

4. Procedure

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configure registration parameters on the source DR.

  a Enter the privileged EXEC mode.

  privilege

  b Enter the global configuration mode.

  config-system terminal

c Enable RP reachability checking on the source DR.

ipv6 pim register-rp-reachability

The RP reachability checking function is disabled by default.

d Limit the rate of sending register messages on the source DR.

ipv6 pim limit-register-rate *rate*

The sending rate of register messages is not limited by default.

e Configure a source address in the register messages on the source DR.

ipv6 pim register-source-addr { *ipv6-local-address* | *interface-type interface-number* }

The source IP address in the register messages is the address of the DR interface connected to the source by default.

f Configure the register suppression time on the source DR.

ipv6 pim register-suppression *suppression-time*

The default register suppression time is **60** seconds.

g Configure the null message detection time on the source DR.

ipv6 pim probe-interval *interval*

The default register-probe time is **5** seconds.

- Configure registration parameters on the RP.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Configure the (S, G) address range for the source on the RP.

ipv6 pim restricted-register { **list** *acl-name* | **route-map** *route-map-name* } *

The (S, G) address range of register messages is not limited by default. The RP receives register messages with any (S, G) address.

You can configure to filter the source register message information on all C-RP or static RP multicast devices.

d Enable register message checksum calculation to be applied to the entire packets on the RP.

ipv6 pim register-checksum-wholepkt [**group-list** *acl-name*]

By default, the register message checksum of headers of only PIM packets and register messages, rather than the entire packets, is calculated.

If the **group-acl** parameter is not specified, this configuration applies to all group addresses.

e Configure the timeout period for the (S, G) entries created upon receiving the register message on the RP.

ipv6 pim rp-register-kat *interval*

By default, the timeout period for the (S, G) entries is the sum of three times of the register suppression time and the register-probe time.

You can configure the (S, G) entry timeout period on all C-RP or static RP multicast devices.

1.6.8 Configuring the Last Hop Device to Switch over from the RPT to SPT

1. Overview

If the multicast traffic in a network is heavy, the RP easily gets overloaded and then faulty. To relieve the RP load, you can initiate SPT switchover through the member DR. You can specify multicast groups to perform SPT switchover.

2. Prerequisites

- Enabling the PIM-SMv6 Function.
- Before an address range of a group is configured for SPT switchover, ensure that the corresponding ACL is created and specify a multicast group range for SPT switchover. For more information about ACL configuration, see *Configuring ACL*.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the SPT switchover function.

ipv6 pim spt-switching [**group-acl** *acl-name*]

The SPT switchover function is disabled by default.

If the **group-acl** parameter is not specified, all groups support SPT switchover.

1.7 Configuring PIM-SMv6 in the SSM Model

1.7.1 Overview

The PIM-SMv6 function in the SSM model can provide specified source services for hosts so that the hosts can determine to receive or reject multicast data of specified sources based on needs.

1.7.2 Restrictions and Guidelines

- If you require a PIM network to support the multicast service in the SSM model, the SSM function must be enabled.
- The SSM function must be enabled on each multicast device.
- The address range of SSM groups configured on all multicast devices must be consistent.
- MLDv2 or SSM mapping must be enabled.

1.7.3 Prerequisites

- In the multicast network, all nodes are reachable via IPv6 unicast routes.
- Enabling the PIM-SMv6 Function.
- Before an SSM group range is configured using an ACL, ensure that the corresponding ACL is created and specify an SSM multicast group range. For more information about ACL configuration, see *Configuring ACL*.

1.7.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the SSM function.

ipv6 pim enable-ssm { **default** | **range** *acl-name* }

The SSM function is disabled by default.

1.8 Configuring PIM-SMv6 Neighbor Parameters

1.8.1 Overview

PIM multicast devices transmit Hello messages to discover neighbors, coordinate protocol parameters and maintain neighbor relationships.

1.8.2 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the Neighbor Message Time Parameters
- Configuring Neighbor Tracking
- Configuring Neighbor Filtering

1.8.3 Prerequisites

- In the multicast network, all nodes are reachable via IPv6 unicast routes.
- Enabling the PIM-SMv6 Function.

1.8.4 Configuring the Neighbor Message Time Parameters

1. Overview

The neighbor message time parameters include hello message sending interval and hello message sending delay.

- The hello message sending interval can be configured, and the hello message hold time is 3.5 times the hello message sending interval. If no new hello message is received within the hello message hold time, the neighbor is considered unreachable.
- When the PIM interface is enabled or detects a new neighbor, random time is generated and it is referred to as hello message sending delay. Within the delay time, the interface sends hello messages out.

2. Restrictions and Guidelines

If the hello message hold time is greater than 65,535 seconds, the hello message sending interval is automatically set to 18,725 seconds.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Configure the neighbor message time parameters. Configure at least one of the tasks.

- Configure the hello message sending interval.

  ipv6 pim query-period *interval*

  The hello message is sent at an interval of **30** seconds by default.
- Configure the hello message sending delay.

  ipv6 pim triggered-hello-delay *delay*

  The default hello message sending delay is **5** seconds.

1.8.5 Configuring Neighbor Tracking

1. Overview

After the suppression capability of an interface is enabled, when a multicast device plans to send a join packet to an uplink multicast device but it receives a join packet sent from the neighbor to the uplink multicast device, this local multicast device suppresses its own join packet. If the suppression capability of the interface is disabled, the join packet can be sent. When the suppression capability of downlink hosts is disabled, an uplink device can determine the number of the downlink hosts based on the quantity of received join packets. This is neighbor tracking.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.

- **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Enable the neighbor tracking function.

ipv6 pim neighbor-tracking

The neighbor tracking function is disabled by default.

1.8.6 Configuring Neighbor Filtering

1. Overview

The neighbor filtering function can protect neighborhood. Controlling neighbors can strengthen PIM network security and limit the valid address range of neighbors. If a neighbor is filtered out based on an access filtering list, PIM-SMv6 does not create peer relationship with the neighbor or stops the peer relationship with this neighbor.

2. Prerequisites

Before an access filtering list is configured, ensure that the corresponding ACL is created and specify a valid neighbor address range. For more information about ACL configuration, see *Configuring ACL*.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Enable the neighbor filtering function.

ip pim peering-filter *acl-name*

The neighbor filtering function is disabled by default.

1.9 Configuring DR Priority

1.9.1 Overview

In a network segment shared by multicast sources and group members, multiple PIM multicast devices exist. The PIM multicast devices elect a DR based on neighbor messages. A device with the highest priority is selected. If the priorities are the same or some PIM devices do not support the DR priority function, the IP addresses are compared and the PIM multicast device with the largest IP address is selected. The selected PIM device functions as a DR, transmits protocol packets between the sources and members, and forwards multicast packets.

1.9.2 Prerequisites

- In the multicast network, all nodes are reachable via IPv6 unicast routes.
- Enabling the PIM-SMv6 Function.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Configure a DR priority in hello messages.

ipv6 pim dr-elected--priority *priority-value*

The default DR priority in hello messages is **1**.

1.10 Configuring Parameters of the Join/Prune Packets

1.10.1 Overview

A device sends a join packet to request an uplink device to forward multicast packets and sends a prune packet to request the uplink device to stop forwarding multicast packets. The request route is referred to as an RPT or SPT.

1.10.2 Restrictions and Guidelines

- The network administrator must ensure that the prune override interval is smaller than the join-prune packet hold time. Otherwise, a short interrupt may occur.
- When the PIM multicast device receives a prune message from an uplink interface, there are other downlink devices existent in the shared network segment. If this PIM multicast device still needs to receive multicast data, the PIM multicast device must send a prune override message to the uplink device within the override interval.

1.10.3 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring Join/Prune Packet Sending Interval
- Configuring the Prune Delay

1.10.4 Prerequisites

Enabling the PIM-SMv6 Function.

1.10.5 Configuring Join/Prune Packet Sending Interval

1. Overview

A PIM device periodically sends the join/prune packet to an uplink device to update the forwarding status. The join packet is used to request data forwarding and the prune packet is used to stop data forwarding. Upon receiving the join/prune packet, the uplink device starts the join/prune packet hold time timer.

- If a join packet hold time timer expires, the PIM device is suppressed to forward multicast packets of the corresponding multicast group through the downlink interface.
- If a prune packet hold time timer expires, the PIM device is restored to forward the multicast packets of the corresponding multicast group through the downlink interface.

2. Restrictions and Guidelines

The join/prune packet sending interval must be smaller than the join/prune packet hold time. Generally, the join/prune packet hold time is 3.5 times the join/prune packet sending interval.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the join/prune packet sending interval.

ipv6 pim jp-timer *interval*

The join/prune packet is sent at an interval of **60** seconds by default.

1.10.6 Configuring the Prune Delay

1. Overview

Upon receiving a prune packet from a downlink device, the multicast device does not perform prune immediately and it waits for a period of time. This period of time is divided into two parts: The part from the time when the multicast device receives a prune message to the time when the multicast device sends the prune message to the uplink device is referred to as the propagation delay; the part from the time when the PIM device receives a prune packet to the time when the device continues to forward the multicast packets to a downlink device is referred to as the override interval. The period from the time when the multicast device receives a prune message to the time when the prune action is performed is the sum of the propagation delay and the override interval.

2. Restrictions and Guidelines

In the period from the time when the multicast device receives a prune message to the time when the prune action is performed, if the PIM device receives a prune override message from the downlink interface, the PIM device cancels pruning.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Configure the prune message propagation delay.

ipv6 pim propagation-delay *propagation-delay*

The default propagation delay of an interface is 500 ms.

(5) Configure the prune override interval.

ipv6 pim override-interval *override-interval*

The default prune override interval is 2500 ms.

1.11 Enabling PIM-SMv6 BFD Correlation

1. Overview

In some shared multicast network, PIM neighbor status change triggers re-initiation of some mechanisms, for example, DR election mechanism. Generally, a PIM neighbor triggers a new round of election after the neighborhood times out, causing multicast data interruption. PIM bidirectional forwarding detection (BFD) correlation can quickly detect link status and neighbor faults in a shared network segment and trigger a new round of DR election. This reduces multicast data transmission interrupt time and ensures reliability of the multicast network.

2. Restrictions and Guidelines

- If DR or assertion election is required in a multicast network, you are advised to enable PIM-SMv6 BFD correlation.
- The PIM BFD function must be enabled on the interconnection interface of the multicast devices.

3. Prerequisites

- In the network, all nodes are reachable via IPv6 unicast routes.
- Enabling the PIM-SMv6 Function.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 3 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*
- Enter the SVI configuration mode.
  - **interface vlan** *interface-number*
- Enter the loopback interface configuration mode.
  - **interface loopback** *interface-number*

(4) Enable PIM-SMv6 BFD correlation.

ipv6 pim bfd

PIM-SMv6 BFD correlation is disabled by default.

1.12 Monitoring

Run the display command to check the running status of a configured function to verify the configuration effect.

Run the cleanup commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup ipv6 pim-info sparse-mode bsr rp-set *	Clears dynamic RP information.
cleanup ipv6 pim-info sparse-mode track	Resets the statistic start time and clears the counter of the PIMv6 packets.
display ipv6 pim-info sparse-mode bsr-router	Displays details about the BSR.
display ipv6 pim-info sparse-mode interface [*interface-type interface-number*] [**detail**]	Displays PIM-SMv6 information of an interface.
display ipv6 pim-info sparse-mode local-members [*interface-type interface-number*]	Displays local MLD information of a PIM-SMv6 interface.
display ipv6 pim-info sparse-mode mroute [*ipv6-group-or-source-address* [*ipv6-group-or-source-address*]]	Displays PIM-SMv6 routing information.
display ipv6 pim-info sparse-mode neighbor [**detail**]	Displays PIM-SMv6 neighbor information.
display ipv6 pim-info sparse-mode nexthop	Displays next hop information, including interface number, address, and metric value of a next hop.
display ipv6 pim-info sparse-mode embedded-rp mapping	Displays all RPs and the groups served by the RPs on the local device.
display ipv6 pim-info sparse-mode rp-group-hash *ipv6-group-address*	Displays RP information corresponding to a group address.
display ipv6 pim-info sparse-mode track	Displays the number of PIM packets sent and received since the statistic start time.

1.13 Configuration Examples

1.13.1 Configuring Basic Functions of PIM-SMv6 in the ASM Model

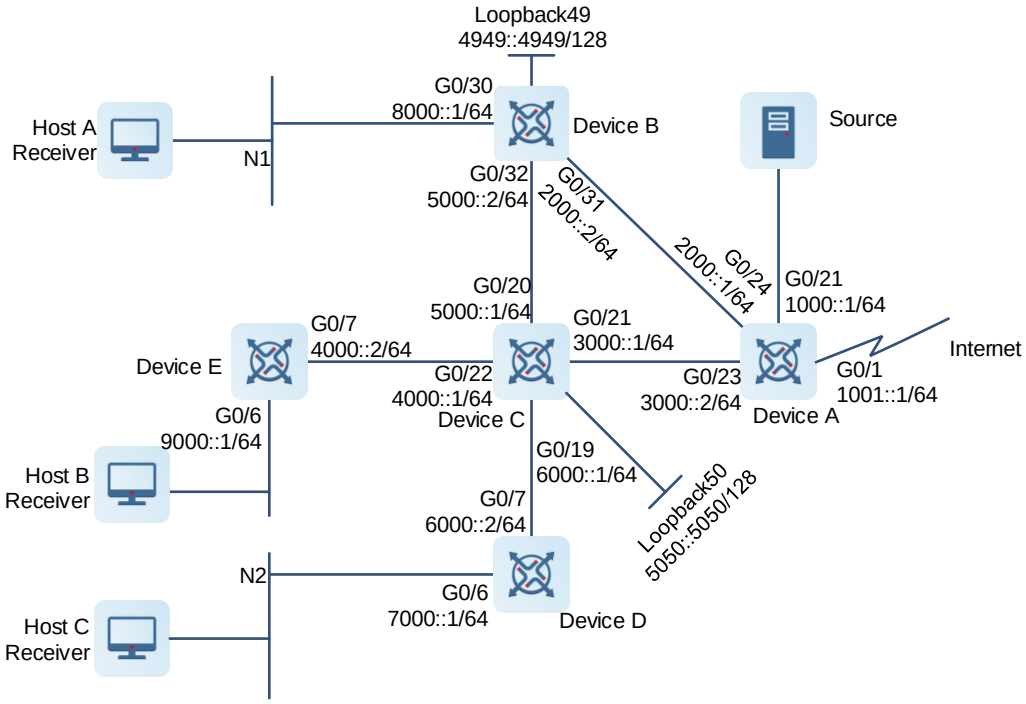
1. Requirements

As shown in Figure 1-5, the internal network is connected to the Internet. The PIM-SMv6 protocol is configured on each multicast device to provide ASM multicast services for hosts in networks N1 and N2 so that hosts

joining a multicast group can receive multicast data destined to this multicast group from any multicast source, including video-on-demand information.

2. Topology

Figure 1-5 PIM-SMv6 Multicast Network Topology in the ASM Model



3. Notes

- Configure IPv6 addresses and unicast routing protocols for multicast devices. The PIMv6 protocol in a multicast domain depends on unicast routing protocols. Normal communication of unicast routing lays a foundation of multicast routing. Ensure that each PIM node is reachable via unicast routes.
- Enable the IPv6 multicast function on all multicast devices. Enabling the multicast function is a prerequisite of PIM-SMv6 configuration.
- Enable the basic functions of PIM-SMv6 on all interfaces of the multicast devices. The other functions of PIM-SMv6 can be enabled only after the basic functions of PIM-SMv6 are enabled.
- Enable the MLD function on the interface connected to hosts so that the hosts can join or leave a multicast group by sending MLD messages. The multicast devices acting as leaf nodes maintain group membership based on the MLD protocol.
- Configure RPs. In the PIM-SMv6 network in the ASM model, the RP is a transitional station for forwarding multicast data. It is recommended that the multicast device with many traffic flows be configured as an RP, for example, device C in Figure 1-5.
- Configure a BSR border on the interface connected to the Internet. BSMs must be transmitted in the PIM-SMv6 domain within the border to improve the multicast controllability.

4. Procedure

(1) Configure IP addresses and enable unicast routing protocols on interfaces of different devices.

Configure an IPv6 address for the interface of Device A and configure the OSPFv3 protocol to implement internetworking.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6 router-protocol ospf 4949
DeviceA(cfg-router)# exit
DeviceA(cfg)# interface gigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 address 1001::1/64
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 address 1000::1/64
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)#interface gigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/23)# ipv6 address 3000::2/64
DeviceA(cfg-if-GigabitEthernet 0/23)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# interface gigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 address 2000::1/64
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
```

Configure IPv6 addresses on interfaces of device B, and configure the OSPFv3 protocol to implement internetworking.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 router-protocol ospf 4949
DeviceB(cfg-router)# exit
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 address 8000::1/64
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/31)# ipv6 address 2000::2/64
DeviceB(cfg-if-GigabitEthernet 0/31)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 address 5000::2/64
```

```
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface loopback 49
DeviceB(cfg-if-Loopback 49)# ipv6 address 4949::4949/128
DeviceB(cfg-if-Loopback 49)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-Loopback 49)# exit
```

Configure IPv6 addresses on interfaces of device C, and configure the OSPFv3 protocol to implement internetworking.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ipv6 router-protocol ospf 4949
DeviceC(cfg-router)# exit
DeviceC(cfg)# interface GigabitEthernet 0/19
DeviceC(cfg-if-GigabitEthernet 0/19)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/19)# ipv6    address 6000::1/64
DeviceC(cfg-if-GigabitEthernet 0/19)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/19)# exit
DeviceC(cfg)#interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 address 5000::1/64
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 address 3000::1/64
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface GigabitEthernet 0/22
DeviceC(cfg-if-GigabitEthernet 0/22)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/22)# ipv6 address 4000::1/64
DeviceC(cfg-if-GigabitEthernet 0/22)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/22)# exit
DeviceC(cfg)# interface loopback 50
DeviceC(cfg-if-Loopback 50)# ipv6 address 5050::5050/128
DeviceC(cfg-if-Loopback 50)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-Loopback 50)# exit
```

Configure IPv6 addresses on interfaces of device D, and configure the OSPFv3 protocol to implement internetworking.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# ipv6 router-protocol ospf 4949
DeviceD(cfg-router)# exit
DeviceD(cfg)# interface GigabitEthernet 0/6
DeviceD(cfg-if-GigabitEthernet 0/6)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/6)# ipv6 address 7000::1/64
```

```
DeviceD(cfg-if-GigabitEthernet 0/6)# ipv6 ospf 4949 set-area 0
DeviceD(cfg-if-GigabitEthernet 0/6)# exit
DeviceD(cfg)# interface GigabitEthernet 0/7
DeviceD(cfg-if-GigabitEthernet 0/7)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/7)# ipv6 address 6000::2/64
DeviceD(cfg-if-GigabitEthernet 0/7)# ipv6 ospf 4949 set-area 0
DeviceD(cfg-if-GigabitEthernet 0/7)# exit
```

Configure IPv6 addresses on interfaces of device E, and configure the OSPFv3 protocol to implement internetworking.

```
DeviceE> privilege
DeviceE# config-system terminal
DeviceE(cfg)# ipv6 router-protocol ospf 4949
DeviceE(cfg-router)# exit
DeviceE(cfg)# interface GigabitEthernet 0/6
DeviceE(cfg-if-GigabitEthernet 0/6)# unset switch-port
DeviceE(cfg-if-GigabitEthernet 0/6)# ipv6 address 9000::1/64
DeviceE(cfg-if-GigabitEthernet 0/6)# ipv6 ospf 4949 set-area 0
DeviceE(cfg-if-GigabitEthernet 0/6)# interface GigabitEthernet 0/7
DeviceE(cfg-if-GigabitEthernet 0/7)# unset switch-port
DeviceE(cfg-if-GigabitEthernet 0/7)# ipv6 address 4000::2/64
DeviceE(cfg-if-GigabitEthernet 0/7)# ipv6 ospf 4949 set-area 0
DeviceE(cfg-if-GigabitEthernet 0/7)# exit
```

(2) Enable the multicast and PIM-SMv6 functions on all multicast devices, and enable the MLD function on all multicast devices connected to hosts.

Enable the multicast and PIM-SMv6 functions on device A.

```
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface gigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface gigabitEthernet 0/23
DeviceA(cfg-if-GigabitEthernet 0/23)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/23)# exit
DeviceA(cfg)# interface gigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
```

Enable the IPv6 multicast, IPv6-SMv6, and MLD functions on device B.

```
DeviceB(cfg)# ipv6 mrouting-enable
DeviceB(cfg)# interface gigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface gigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface gigabitEthernet 0/32
```

```
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface loopback 49
DeviceB(cfg-if-Loopback 49)# ipv6 pim sparse-mode
DeviceB(cfg-if-Loopback 49)# exit
```

Enable the IPv6 multicast and PIM-SMv6 functions on device C.

```
DeviceC(cfg)# ipv6 mrouting-enable
DeviceC(cfg)# interface gigabitEthernet 0/19
DeviceC(cfg-if-GigabitEthernet 0/19)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/19)# exit
DeviceC(cfg)# interface gigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface gigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface gigabitEthernet 0/22
DeviceC(cfg-if-GigabitEthernet 0/22)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/22)# exit
DeviceC(cfg)# interface loopback 50
DeviceC(cfg-if-Loopback 50)# ipv6 pim sparse-mode
DeviceC(cfg-if-Loopback 50)# exit
```

Enable the IPv6 multicast, IPv6-SMv6, and MLD functions on device D.

```
DeviceD(cfg)# ipv6 mrouting-enable
DeviceD(cfg)# interface gigabitEthernet 0/6
DeviceD(cfg-if-GigabitEthernet 0/6)# ipv6 pim sparse-mode
DeviceD(cfg-if-GigabitEthernet 0/6)# exit
DeviceD(cfg)# interface gigabitEthernet 0/7
DeviceD(cfg-if-GigabitEthernet 0/7)# ipv6 pim sparse-mode
DeviceD(cfg-if-GigabitEthernet 0/7)# exit
```

Enable the IPv6 multicast, IPv6-SMv6, and MLD functions on device E.

```
DeviceE(cfg)# ipv6 mrouting-enable
DeviceE(cfg)# interface gigabitEthernet 0/6
DeviceE(cfg-if-GigabitEthernet 0/6)# ipv6 pim sparse-mode
DeviceE(cfg-if-GigabitEthernet 0/6)# exit
DeviceE(cfg)# interface gigabitEthernet 0/7
DeviceE(cfg-if-GigabitEthernet 0/7)# ipv6 pim sparse-mode
DeviceE(cfg-if-GigabitEthernet 0/7)# exit
```

(3) Configure RPs.

Note

There are two methods to configure RPs: dynamic RPs and static RPs. You can configure either or both types of RPs. If static and dynamic RPs are available at the same time, the dynamic RPs are preferred by default.

Configure dynamic RPs. One or multiple multicast devices in the PIM-SMv6 domain must be configured as C-RPs and C-BSRs. In this example, device B and device C are configured as C-RP and C-BSR. The address range of the served groups of the RP, as well as the interfaces of the C-RP and C-BSR are configured on device B and device C. The loopback interface is configured as the interface of the C-RP and C-BSR.

Configure C-RP and C-BSR and address range of the served groups of the RP on device B.

```
DeviceB(cfg)# ipv6 pim c-bsr loopback 49
DeviceB(cfg)# ipv6 access-control-list mcast-ipv6-list
DeviceB(cfg-ipv6acl)# permit ipv6 any ff13::/64
DeviceB(cfg)# ipv6 pim c-rp loopback 49 group-acl mcast-ipv6-list
```

Configure C-RP and C-BSR and address range of the served groups of the RP on device C.

```
DeviceC(cfg)# ipv6 pim c-bsr loopback 50
DeviceC(cfg)# ipv6 access-control-list mcast-ipv6-list
DeviceC(cfg-ipv6acl)# permit ipv6 any ff14::/64
DeviceC(cfg-stdacl)# exit
DeviceC(cfg)# ipv6 pim c-rp loopback 50 group-acl mcast-ipv6-list
```

Configure static RPs. The addresses of static RPs must be specified on all multicast devices.

Configure a static RP on device A.

```
DeviceA(cfg)# ipv6 pim static-rp-addr 5050::5050
```

Configure a static RP on device B.

```
DeviceB(cfg)# ipv6 pim static-rp-addr 5050::5050
```

Configure a static RP on device C.

```
DeviceC(cfg)# ipv6 pim static-rp-addr 5050::5050
```

Configure a static RP on device D.

```
DeviceD(cfg)# ipv6 pim static-rp-addr 5050::5050
```

Configure a static RP on device E.

```
DeviceE(cfg)# ipv6 pim static-rp-addr 5050::5050
```

(4) Configure a BSR border on the device A interface connected to the Internet.

```
DeviceA(cfg-if-GigabitEthernet 0/1)# ipv6 pim bsr-boundary
```

5. Verification

(1) Run the **display ipv6 pim-info sparse-mode interface** command to display the PIM-SMv6 running information on the interfaces.

The PIM-SMv6 running information on the interface of device A is as follows:

```
DeviceA# display ipv6 pim-info sparse-mode interface
Interface                                VIFindex Ver/     Nbr      DR
                                                  Mode     Count    Prior
GigabitEthernet 0/24                     3        v2/S     1        1
    Address          : fe80::2d0:f8ff:fe22:358d
    Global Address: 2000::1
    DR               : fe80::2d0:f8ff:feee:8c1f
```

```
GigabitEthernet 0/23                2          v2/S    1        1
      Address          : fe80::2d0:f8ff:fe22:358d
      Global Address: 3000::2
      DR               : this system
GigabitEthernet 0/21                1          v2/S    0        1
      Address          : fe80::2d0:f8ff:fe22:358d
      Global Address: 1000::1
      DR               : this system
```

The PIM-SMv6 running information on the interface of device B is as follows:

```
DeviceB# display ipv6 pim-info sparse-mode interface
Interface                           VIFindex Ver/     Nbr       DR
                                             Mode     Count     Prior
GigabitEthernet 0/31                2          v2/S    1        1
      Address          : fe80::2d0:f8ff:feee:8c1f
      Global Address: 2000::2
      DR               : this system
GigabitEthernet 0/32                3          v2/S    1        1
      Address          : fe80::2d0:f8ff:feee:8c1f
      Global Address: 5000::2
      DR               : this system
GigabitEthernet 0/30                1          v2/S    0        1
      Address          : fe80::2d0:f8ff:feee:8c1f
      Global Address: 8000::1
      DR               : this system
Loopback 49                           4          v2/S    0        1
      Address          : fe80::2d0:f8ff:feee:8c1f
      Global Address: 4949::4949
      DR               : this system
```

The PIM-SMv6 running information on the interface of device C is as follows:

```
DeviceC# display ipv6 pim-info sparse-mode interface
Interface                           VIFindex Ver/     Nbr       DR
                                             Mode     Count     Prior
GigabitEthernet 0/21                3          v2/S    1        1
      Address          : fe80::274:9cff:feee:f49f
      Global Address: 3000::1
      DR               : fe80::2d0:f8ff:fe22:358d
GigabitEthernet 0/22                4          v2/S    1        1
      Address          : fe80::274:9cff:feee:f49f
      Global Address: 4000::1
      DR               : this system
GigabitEthernet 0/20                2          v2/S    1        1
      Address          : fe80::274:9cff:feee:f49f
      Global Address: 5000::1
      DR               : fe80::2d0:f8ff:feee:8c1f
GigabitEthernet 0/19                1          v2/S    1        1
```

```
    Address          : fe80::274:9cff:feee:f49f
    Global Address: 6000::1
    DR               : this system
Loopback 50                          5          v2/S   0        1
    Address          : fe80::274:9cff:feee:f49f
    Global Address: 5050::5050
    DR               : this system
```

The PIM-SMv6 running information on the interface of device D is as follows:

```
DeviceD> privilege
DeviceD# display ipv6 pim-info sparse-mode interface
Interface                            VIFindex Ver/   Nbr     DR
                                              Mode   Count   Prior
GigabitEthernet 0/6                  1        v2/S   0       1
    Address          : fe80::274:9cff:fe11:c9e7
    Global Address: 7000::1
    DR               : this system
GigabitEthernet 0/7                  2        v2/S   1       1
    Address          : fe80::274:9cff:fe11:c9e7
    Global Address: 6000::2
    DR               : fe80::274:9cff:feee:f49f
```

The PIM-SMv6 running information on the interface of device E is as follows:

```
DeviceE# display ipv6 pim-info sparse-mode interface
Interface                            VIFindex Ver/   Nbr     DR
                                              Mode   Count   Prior
GigabitEthernet 0/6                  1        v2/S   0       1
    Address          : fe80::274:9cff:fec8:e27b
    Global Address: 9000::1
    DR               : this system
GigabitEthernet 0/7                  2        v2/S   1       1
    Address          : fe80::274:9cff:fec8:e27b
    Global Address: 4000::2
    DR               : fe80::274:9cff:feee:f49f
```

(2) Run the **display ipv6 pim-info sparse-mode bsr-router** command to display the election information of the BSRs.

The BSR election information on device A is as follows:

```
DeviceA# display ipv6 pim-info sparse-mode bsr-router
BSR address: 5050::5050
  Uptime:         00:05:28, BSR Priority: 64, Hash mask length: 126
  Expires:        00:01:54
  Role: Non-candidate BSR    Priority: 0, Hash mask length: 126
  State: Accept Preferred
```

The BSR election information on device B is as follows:

```
DeviceB# display ipv6 pim-info sparse-mode bsr-router
PIMv2 Bootstrap information
```

```
  BSR address: 5050::5050
  Uptime:         00:04:13, BSR Priority: 64, Hash mask length: 126
  Expires:        00:02:09
  Role: Candidate BSR    Priority: 64, Hash mask length: 126
  State: Candidate BSR

  Candidate RP: 4949::4949(Loopback 49)
    Advertisement interval 60 seconds
    Next Cand_RP_advertisement in 00:00:53
```

The BSR election information on device C is as follows:

```
DeviceC# display ipv6 pim-info sparse-mode bsr-router
PIMv2 Bootstrap information
This system is the Bootstrap Router (BSR)
  BSR address: 5050::5050
  Uptime:         00:04:25, BSR Priority: 64, Hash mask length: 126
  Next bootstrap message in 00:00:52
  Role: Candidate BSR    Priority: 64, Hash mask length: 126
  State: Elected BSR

  Candidate RP: 5050::5050(Loopback 50)
    Advertisement interval 60 seconds
    Next Cand_RP_advertisement in 00:00:44
```

The BSR election information on device D is as follows:

```
DeviceD# display ipv6 pim-info sparse-mode bsr-router
PIMv2 Bootstrap information
  BSR address: 5050::5050
  Uptime:         00:06:00, BSR Priority: 64, Hash mask length: 126
  Expires:        00:01:21
  Role: Non-candidate BSR    Priority: 0, Hash mask length: 126
  State: Accept Preferred
```

The BSR election information on device E is as follows:

```
DeviceE# display ipv6 pim-info sparse-mode bsr-router
PIMv2 Bootstrap information
  BSR address: 5050::5050
  Uptime:         00:06:23, BSR Priority: 64, Hash mask length: 126
  Expires:        00:01:58
  Role: Non-candidate BSR    Priority: 0, Hash mask length: 126
  State: Accept Preferred
```

(3) Run the **display ipv6 pim-info sparse-mode embedded-rp mapping** command to display the RP information on the multicast devices.

The RP information on device A is as follows:

```
DeviceA# display ipv6 pim-info sparse-mode embedded-rp mapping
PIM Group-to-RP Mappings
Group(s): ff13::/64
```

```
  RP: 4949::4949(Not self)
    Info source: 5050::5050, via bootstrap, priority 192
          Uptime: 03:10:58, expires: 00:01:32
Group(s): ff14::/64
  RP: 5050::5050(Not self)
    Info source: 5050::5050, via bootstrap, priority 192
          Uptime: 03:08:58, expires: 00:01:32
Group(s): ff00::/8, Static
    RP: 5050::5050(Not self), Static
          Uptime: 03:08:15
```

The RP information on device B is as follows:

```
DeviceB# display ipv6 pim-info sparse-mode embedded-rp mapping
PIM Group-to-RP Mappings
Group(s): ff13::/64
  RP: 4949::4949(Self)
    Info source: 5050::5050, via bootstrap, priority 192
          Uptime: 03:12:31, expires: 00:02:08
Group(s): ff14::/64
  RP: 5050::5050(Not self)
    Info source: 5050::5050, via bootstrap, priority 192
          Uptime: 03:10:26, expires: 00:02:08
Group(s): ff00::/8, Static
    RP: 5050::5050(Not self), Static
          Uptime: 03:09:35
```

The RP information on device C is as follows:

```
DeviceC# display ipv6 pim-info sparse-mode embedded-rp mapping
PIM Group-to-RP Mappings
Group(s): ff13::/64
  RP: 4949::4949(Not self)
    Info source: 4949::4949, via bootstrap, priority 192
          Uptime: 00:21:40, expires: 00:01:49
Group(s): ff14::/64
  RP: 5050::5050(Self)
    Info source: 4949::4949, via bootstrap, priority 192
          Uptime: 00:19:40, expires: 00:01:49
Group(s): ff00::/8, Static
    RP: 5050::5050(Self), Static
          Uptime: 00:18:52
```

The RP information on device D is as follows:

```
DeviceD# display ipv6 pim-info sparse-mode embedded-rp mapping
PIM Group-to-RP Mappings
Group(s): ff13::/64
  RP: 4949::4949(Not self)
    Info source: 4949::4949, via bootstrap, priority 192
          Uptime: 00:21:59, expires: 00:01:31
```

```
Group(s): ff14::/64
  RP: 5050::5050(Not self)
    Info source: 4949::4949, via bootstrap, priority 192
        Uptime: 00:19:59, expires: 00:01:31
Group(s): ff00::/8, Static
    RP: 5050::5050(Not self), Static
        Uptime: 00:19:09
```

The RP information on device E is as follows:

```
DeviceE# display ipv6 pim-info sparse-mode embedded-rp mapping
PIM Group-to-RP Mappings
Group(s): ff13::/64
  RP: 4949::4949(Not self)
    Info source: 4949::4949, via bootstrap, priority 192
        Uptime: 00:22:24, expires: 00:02:06
Group(s): ff14::/64
  RP: 5050::5050(Not self)
    Info source: 4949::4949, via bootstrap, priority 192
        Uptime: 00:20:24, expires: 00:02:06
Group(s): ff00::/8, Static
    RP: 5050::5050(Not self), Static
        Uptime: 00:19:31
```

(4) Run the **display ipv6 pim-info sparse-mode mroute** command to display the multicast routing tables of PIM-SMv6. The multicast source (1001:: 2/64) sends traffic to the multicast groups (FF14::1/64) and (FF13::1/64). Host A joins the multicast group (FF14::1/64), and host B joins the multicast group (FF13::1/64).

The routing table information on device A is as follows:

```
DeviceA# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 3
(S,G,rpt) Entries: 3
FCR Entries: 0
REG Entries: 3

(1000::2, ff13::1)
RPF nbr: ::
RPF idx: None
SPT bit: 1
Upstream State: JOINED
kat expires in 152 seconds
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . . j . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(1000::2, ff13::1, rpt)
RP: 4949::4949
RPF nbr: fe80::2d0:f8ff:feee:8c1f
RPF idx: GigabitEthernet 0/24
Upstream State: RPT NOT JOINED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
```

```
5 . . . . . .
6 . . . . . .
7 . . . . . .
Pruned
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(1000::2, ff14::1)
RPF nbr: ::
RPF idx: None
SPT bit: 1
Upstream State: JOINED
kat expires in 76 seconds
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . j . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
```

```
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(1000::2, ff14::1, rpt)
RP: 5050::5050
RPF nbr: fe80::274:9cff:feee:f49f
RPF idx: GigabitEthernet 0/23
Upstream State: RPT NOT JOINED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Pruned
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . . . .
```

```
1  . . . . . .
2  . . . . . .
3  . . . . . .
4  . . . . . .
5  . . . . . .
6  . . . . . .
7  . . . . . .

(1000::2, ff15::1)
RPF nbr: ::
RPF idx: None
SPT bit: 1
Upstream State: NOT JOINED
kat expires in 32 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  . . . . . .
1  . . . . . .
2  . . . . . .
3  . . . . . .
4  . . . . . .
5  . . . . . .
6  . . . . . .
7  . . . . . .
Joined
0  . . . . . .
1  . . . . . .
2  . . . . . .
3  . . . . . .
4  . . . . . .
5  . . . . . .
6  . . . . . .
7  . . . . . .
Asserted
0  . . . . . .
1  . . . . . .
2  . . . . . .
3  . . . . . .
4  . . . . . .
5  . . . . . .
6  . . . . . .
7  . . . . . .
Outgoing
0  . . . . . .
1  . . . . . .
2  . . . . . .
```

```
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(1000::2, ff15::1, rpt)
RP: 5050::5050
RPF nbr: fe80::274:9cff:feee:f49f
RPF idx: GigabitEthernet 0/23
Upstream State: RPT NOT JOINED
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Pruned
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
```

The routing table information on device B is as follows (only key information is displayed):

```
DeviceB# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 2
(S,G) Entries: 1
```

```
(S,G,rpt) Entries: 1
FCR Entries: 1
REG Entries: 0

(*, ff13::1)
RP: 4949::4949
RPF nbr: ::
RPF idx: None
Upstream State: JOINED
FCR:

(1000::2, ff13::1)
RPF nbr: fe80::2d0:f8ff:fe22:358d
RPF idx: GigabitEthernet 0/31
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 18 seconds
kat expires in 207 seconds

(1000::2, ff13::1, rpt)
RP: 4949::4949
RPF nbr: ::
RPF idx: None
Upstream State: PRUNED

(*, ff14::1)
RP: 5050::5050
RPF nbr: fe80::274:9cff:feee:f49f
RPF idx: GigabitEthernet 0/32
Upstream State: JOINED
jt_timer expires in 15 seconds
FCR:
Source: 1000::2
KAT timer running, 135 seconds remaining
  Packet count 1
```

The routing table information on device C is as follows (only key information is displayed):

```
DeviceC# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 2
(S,G) Entries: 2
(S,G,rpt) Entries: 2
FCR Entries: 1
REG Entries: 0
```

```
(*, ff13::1)
RP: 4949::4949
RPF nbr: fe80::2d0:f8ff:feee:8c1f
RPF idx: GigabitEthernet 0/20
Upstream State: JOINED
jt_timer expires in 40 seconds
FCR:
Source: 1000::2

 KAT timer running, 100 seconds remaining
 Packet count 1

(*, ff14::1)
RP: 5050::5050
RPF nbr: ::
RPF idx: None
Upstream State: JOINED
FCR:

(1000::2, ff14::1)
RPF nbr: fe80::2d0:f8ff:fe22:358d
RPF idx: GigabitEthernet 0/21
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 40 seconds
kat expires in 208 seconds

(1000::2, ff14::1, rpt)
RP: 5050::5050
RPF nbr: ::
RPF idx: None
Upstream State: PRUNED

(1000::2, ff15::1)
RPF nbr: fe80::2d0:f8ff:fe22:358d
RPF idx: GigabitEthernet 0/21
SPT bit: 0
Upstream State: NOT JOINED
kat expires in 210 seconds

(1000::2, ff15::1, rpt)
RP: 5050::5050
RPF nbr: ::
RPF idx: None
Upstream State: RPT NOT JOINED
```

The routing table information on device E is as follows (only key information is displayed):

```
DeviceE# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 1
(S,G) Entries: 0
(S,G,rpt) Entries: 0
FCR Entries: 1
REG Entries: 0

(*, ff13::1)
RP: 4949::4949
RPF nbr: fe80::274:9cff:feee:f49f
RPF idx: GigabitEthernet 0/7
Upstream State: JOINED
jt_timer expires in 22 seconds
FCR:
Source: 1000::2
KAT timer running, 172 seconds remaining
 Packet count 1
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/1
 unset switch-port
 ipv6 address 1001::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim bsr-boundary
!
interface GigabitEthernet 0/21
 unset switch-port
 ipv6 address 1000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/23
 unset switch-port
 ipv6 address 3000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
```

```
interface GigabitEthernet 0/24
 unset switch-port
 ipv6 address 2000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim static-rp-addr 5050::5050
!
```

- Device B configuration file

```
hostname DeviceB
!
ipv6 access-control-list mcast-ipv6-list
 10 permit ipv6 any FF13::/64
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/30
 unset switch-port
 ipv6 address 8000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/31
 unset switch-port
 ipv6 address 2000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/32
 unset switch-port
 ipv6 address 5000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 49
 ipv6 address 4949::4949/128
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
Ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim static-rp-addr 5050::5050
```

```
ipv6 pim c-bsr Loopback 49
ipv6 pim c-rp Loopback 49 group-acl mcast-ipv6-list
!
```

- Device C configuration file

```
hostname DeviceC
!
ipv6 access-control-list mcast-ipv6-list
 10 permit ipv6 any FF14::/64
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/19
 unset switch-port
 ipv6 address 6000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/20
 unset switch-port
 ipv6 address 5000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/21
 unset switch-port
 ipv6 address 3000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/22
 unset switch-port
 ip address 10.100.4.1 255.255.255.0
 ipv6 address 4000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 50
 ipv6 address 5050::5050/128
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim static-rp-addr 5050::5050
ipv6 pim c-bsr Loopback 50
```

```
ipv6 pim c-rp Loopback 50 group-acl mcast-ipv6-list
!
```

- Device D configuration file

```
hostname DeviceD
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/6
 unset switch-port
 ipv6 address 7000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/7
 unset switch-port
 ipv6 address 6000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim static-rp-addr 5050::5050
!
```

- Device E configuration file

```
hostname DeviceE
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/6
 unset switch-port
 ipv6 address 9000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/7
 unset switch-port
 ipv6 address 4000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim static-rp-addr 5050::5050
```

```
!
```

7. Common Errors

- The multicast routing function is not enabled on multicast devices.
- The multicast function is not enabled on any interface.
- No BSR border is configured between PIM-SMv6 domains.
- No RP is configured.
- No static RP is configured on a multicast device, or the configured static RP on this multicast device is not consistent with the configuration on other multicast devices.
- C-RPs are configured, but no C-BSR is configured.
- Static RP, C-RP or C-BSR is unreachable via unicast routes.

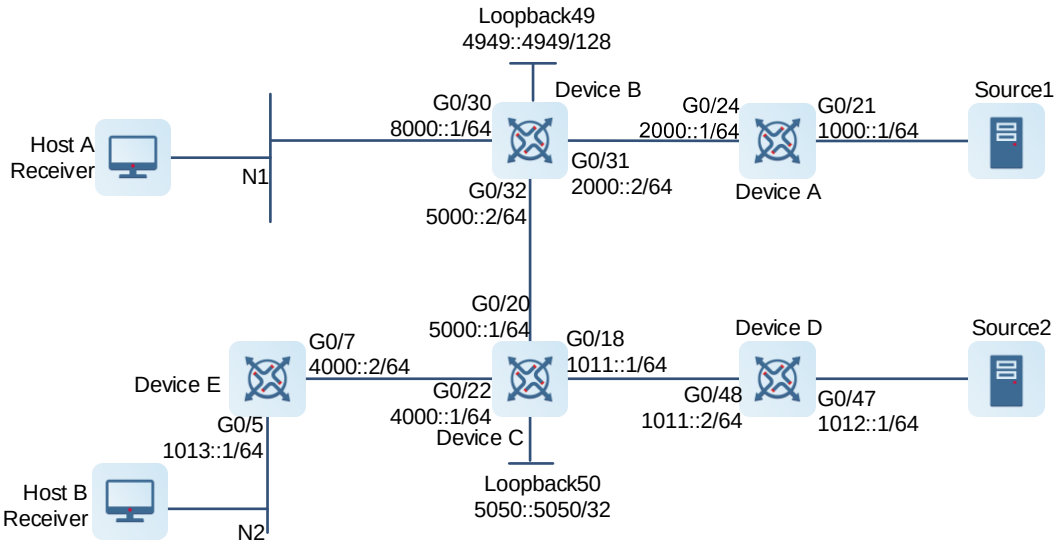
1.13.2 Configuring Basic Functions of PIM-SMv6 in the SSM Model

1. Requirements

As shown in Figure 1-6, the PIM-SMv6 protocol is enabled on multicast devices to provide SSM services for hosts in N1 and N2 so that the hosts can receive multicast data from specified multicast sources to the multicast group that the hosts join, for example, video-on-demand.

2. Topology

Figure 1-6 PIM-SMv6 Multicast Network Topology in the SSM Model



3. Notes

- Configure IPv6 addresses and unicast routing protocols for multicast devices. The PIMv6 protocol in a multicast domain depends on unicast routing protocols. Normal communication of unicast routing lays a foundation of multicast routing. Ensure that each PIMv6 node is reachable via unicast routes.
- Enable the multicast function on all multicast devices. Enabling the multicast function is a prerequisite of PIM-SMv6 configuration.

- Enable the basic functions of PIM-SMv6 on all interfaces of the multicast devices. The other functions of PIM-SMv6 can be enabled only after the basic functions of PIM-SMv6 are enabled.
- Enable the MLD function on the interface connected to hosts and set the MLD version to 2 so that the hosts can join or leave a multicast group by sending MLD messages based on the specified source. The multicast devices acting as leaf nodes maintain group membership based on the MLD protocol.
- Configure the same address range of SSM on all multicast devices. The SSM services provided by the multicast devices in the PIM-SMv6 domain are dedicated to specified groups, which realizes multicast controllability.

4. Procedure

(1) Configure IP addresses and enable unicast routing protocols on interfaces of different devices.

Configure IPv6 addresses and enable the OSPFv6 unicast routing protocol on interfaces of device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 address 1000::1/64
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 address 2000::1/64
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
```

Configure IPv6 addresses and enable the OSPFv6 unicast routing protocol on interfaces of device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 address 8000::1/64
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/31)# ipv6 address 2000::2/64
DeviceB(cfg-if-GigabitEthernet 0/31)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 address 5000::2/64
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ipv6 address 4949::4949/128
```

```
DeviceB(cfg-if-Loopback 49)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-Loopback 49)# exit
```

Configure IPv6 addresses and enable the OSPFv6 unicast routing protocol on interfaces of device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/18
DeviceC(cfg-if-GigabitEthernet 0/18)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/18)# ipv6 address 1011::1/64
DeviceC(cfg-if-GigabitEthernet 0/18)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/18)# exit
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 address 5000::1/64
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/22
DeviceC(cfg-if-GigabitEthernet 0/22)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/22)# ipv6 address 4000::1/64
DeviceC(cfg-if-GigabitEthernet 0/22)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/22)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ipv6 address 5050::5050/128
DeviceC(cfg-if-Loopback 50)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-Loopback 50)# exit
```

Configure IPv6 addresses and enable the OSPFv6 unicast routing protocol on interfaces of device D.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# interface GigabitEthernet 0/47
DeviceD(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/47)# ipv6 address 1012::1/64
DeviceD(cfg-if-GigabitEthernet 0/47)#    ipv6 ospf 4949 set-area 0
DeviceD(cfg-if-GigabitEthernet 0/47)# exit
DeviceD(cfg)# interface GigabitEthernet 0/48
DeviceD(cfg-if-GigabitEthernet 0/48)# unset switch-port
DeviceD(cfg-if-GigabitEthernet 0/48)# ipv6 address 1011::2/64
DeviceD(cfg-if-GigabitEthernet 0/48)# ipv6 ospf 4949 set-area 0
DeviceD(cfg-if-GigabitEthernet 0/48)# exit
```

Configure IPv6 addresses and enable the OSPFv6 unicast routing protocol on interfaces of device E.

```
DeviceE> privilege
DeviceE# config-system terminal
DeviceE(cfg)# interface GigabitEthernet 0/7
DeviceE(cfg-if-GigabitEthernet 0/7)# unset switch-port
DeviceE(cfg-if-GigabitEthernet 0/7)# ipv6 address 4000::2/64
DeviceE(cfg-if-GigabitEthernet 0/7)# ipv6 ospf 4949 set-area 0
DeviceE(cfg-if-GigabitEthernet 0/7)# exit
```

```
DeviceE(cfg)# interface GigabitEthernet 0/5
DeviceE(cfg-if-GigabitEthernet 0/5)# unset switch-port
DeviceE(cfg-if-GigabitEthernet 0/5)# ipv6 address 1013::1/64
DeviceE(cfg-if-GigabitEthernet 0/5)# ipv6 ospf 4949 set-area 0
DeviceE(cfg-if-GigabitEthernet 0/5)# exit
```

(2) Enable the multicast and PIM-SMv6 functions on all multicast devices, and enable the MLD function on all multicast devices connected to hosts.

Enable the IPv6 multicast and PIM-SMv6 functions on device A.

```
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/21
DeviceA(cfg-if-GigabitEthernet 0/21)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/21)# exit
DeviceA(cfg)# interface GigabitEthernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
```

Enable the IPv6 multicast, PIM-SMv6, and MLD functions on device B, and set the MLD version to 3.

```
DeviceB(cfg)# ipv6 mrouting-enable
DeviceB(cfg)# interface GigabitEthernet 0/30
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/30)# ipv6 mld specifiy-version 2
DeviceB(cfg-if-GigabitEthernet 0/30)# exit
DeviceB(cfg)# interface GigabitEthernet 0/31
DeviceB(cfg-if-GigabitEthernet 0/31)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/31)# exit
DeviceB(cfg)# interface GigabitEthernet 0/32
DeviceB(cfg-if-GigabitEthernet 0/32)# ipv6 pim sparse-mode
DeviceB(cfg-if-GigabitEthernet 0/32)# exit
DeviceB(cfg)# interface Loopback 49
DeviceB(cfg-if-Loopback 49)# ipv6 pim sparse-mode
DeviceB(cfg-if-Loopback 49)# exit
```

Enable the IPv6 multicast and PIM-SMv6 functions on device C.

```
DeviceC(cfg)# ipv6 mrouting-enable
DeviceC(cfg)# interface GigabitEthernet 0/18
DeviceC(cfg-if-GigabitEthernet 0/18)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/18)# exit
DeviceC(cfg)# interface GigabitEthernet 0/20
DeviceC(cfg-if-GigabitEthernet 0/20)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/20)# exit
DeviceC(cfg)# interface GigabitEthernet 0/22
DeviceC(cfg-if-GigabitEthernet 0/22)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/22)# exit
DeviceC(cfg)# interface Loopback 50
DeviceC(cfg-if-Loopback 50)# ipv6 pim sparse-mode
DeviceC(cfg-if-Loopback 50)# exit
```

```
DeviceC(cfg)#
```

Enable the IPv6 multicast and PIM-SMv6 functions on device D.

```
DeviceD(cfg)# ipv6 mrouting-enable
DeviceD(cfg)# interface GigabitEthernet 0/47
DeviceD(cfg-if-GigabitEthernet 0/47)# ipv6 pim sparse-mode
DeviceD(cfg-if-GigabitEthernet 0/47)# exit
DeviceD(cfg)# interface GigabitEthernet 0/48
DeviceD(cfg-if-GigabitEthernet 0/48)# ipv6 pim sparse-mode
DeviceD(cfg-if-GigabitEthernet 0/48)# exit
```

Enable the IPv6 multicast, PIM-SMv6, and MLD functions on Device E, and set the MLD version to 3.

```
DeviceE(cfg)# ipv6 mrouting-enable
DeviceE(cfg)# interface GigabitEthernet 0/7
DeviceE(cfg-if-GigabitEthernet 0/7)# ipv6 pim sparse-mode
DeviceE(cfg-if-GigabitEthernet 0/7)# exit
DeviceE(cfg)# interface GigabitEthernet 0/5
DeviceE(cfg-if-GigabitEthernet 0/5)# ipv6 mld specifiy-version 2
DeviceE(cfg-if-GigabitEthernet 0/5)# ipv6 pim sparse-mode
DeviceE(cfg-if-GigabitEthernet 0/5)# exit
```

(3) Enable the SSM function and configure an address range of multicast groups on all multicast devices.

Set the address range of an SSM multicast group to FF33::/64 on device A.

```
DeviceA(cfg)# ipv6 access-control-list ipv6-ssm-list
DeviceA(cfg-ipv6acl)# permit ipv6 any FF33::/64
DeviceA(cfg-ipv6acl)#exit
DeviceA(cfg)# ipv6 pim enable-ssm range ipv6-ssm-list
```

Set the address range of an SSM multicast group to FF33::/64 on device B.

```
DeviceB(cfg)# ipv6 access-control-list ipv6-ssm-list
DeviceB(cfg-ipv6acl)# permit ipv6 any FF33::/64
DeviceB(cfg-ipv6acl)#exit
DeviceB(cfg)# ipv6 pim enable-ssm range ipv6-ssm-list
```

Set the address range of an SSM multicast group to FF33::/64 on device C.

```
DeviceC(cfg)# ipv6 access-list ipv6-ssm-list
DeviceC(cfg-ipv6acl)# permit ipv6 any FF33::/64
DeviceC(cfg-ipv6acl)#exit
DeviceC(cfg)# ipv6 pim enable-ssm range ipv6-ssm-list
```

Set the address range of an SSM multicast group to FF33::/64 on device D.

```
DeviceD(cfg)# ipv6 access-list ipv6-ssm-list
DeviceD(cfg-ipv6acl)# permit ipv6 any FF33::/64
DeviceD(cfg-ipv6acl)#exit
DeviceD(cfg)# ipv6 pim enable-ssm range ipv6-ssm-list
```

Set the address range of an SSM multicast group to FF33::/64 on device E.

```
DeviceE(cfg)# ipv6 access-list ipv6-ssm-list
DeviceE(cfg-ipv6acl)# permit ipv6 any FF33::/64
```

```
DeviceE(cfg-ipv6acl)#exit
DeviceE(cfg)# ipv6 pim enable-ssm range ipv6-ssm-list
```

5. Verification

(1) Run the **display ipv6 pim-info sparse-mode interface** command to display the PIM-SMv6 running information on the interfaces of each device.

The PIM-SMv6 running information on the interfaces of device A is as follows. The PIM-SM running information on the interfaces of devices B, C, D, and E is the same.

```
DeviceA> privilege
DeviceA# display ipv6 pim-info sparse-mode interface
Interface                                VIFindex Ver/     Nbr       DR
                                                  Mode     Count    Prior
GigabitEthernet 0/21                     2        v2/S     0        1
     Address          : fe80::2d0:f8ff:fe22:358d
     Global Address: 1000::1
     DR               : this system
GigabitEthernet 0/24                     3        v2/S     1        1
     Address          : fe80::2d0:f8ff:fe22:358d
     Global Address: 2000::1
     DR               : fe80::2d0:f8ff:feee:8c1f
```

(2) Run the **display ipv6 pim-info sparse-mode mroute** command to display the multicast routing tables of PIMv6. Host A needs to receive packets sent from the source (1000::2/64) and the source (1012::2/64) to the multicast group (FF33::100). Host B needs to receive packets sent from the source (1000::2/64) to the multicast group (FF33::100). The information output is as follows:

Run the **display ipv6 pim-info sparse-mode mroute** command to display the multicast routing tables of device A.

```
DeviceA> privilege
DeviceA# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 1
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(1000::2, ff33::100)
RPF nbr: ::
RPF idx: None
SPT bit: 1
Upstream State: JOINED
kat expires in 135 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . . j . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
```

Run the **display ipv6 pim-info sparse-mode mroute** command to display the multicast routing tables of device B.

```
DeviceB> privilege
DeviceB# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 2
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0
```

```
(1000::2, ff33::100)
RPF nbr: fe80::2d0:f8ff:fe22:358d
RPF idx: GigabitEthernet 0/31
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 14 seconds
kat expires in 122 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  i  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  j  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  .  o  .  o  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
(1012::2, ff33::100)
RPF nbr: fe80::274:9cff:feee:f49f
RPF idx: GigabitEthernet 0/32
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 14 seconds
kat expires in 104 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  i  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Outgoing
0  .  .  .  .  o  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

Run the **display ipv6 pim-info sparse-mode mroute** command to display the multicast routing tables of device C.

```
DeviceC> privilege
DeviceC# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 2
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(1000::2, ff33::100)
RPF nbr: fe80::2d0:f8ff:feee:8c1f
RPF idx: GigabitEthernet 0/20
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 30 seconds
kat expires in 91 seconds
   00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Joined
0  .  .  .  .  .  j  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
7  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
Asserted
0  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
1  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
2  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
3  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
4  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
5  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
6  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .  .
```

```
7 . . . . . .
Outgoing
0 . . . . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

(1012::2, ff33::100)
RPF nbr: fe80::2d0:f8ff:fe22:363d
RPF idx: GigabitEthernet 0/18
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 40 seconds
kat expires in 11 seconds
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . . j . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
```

```
Outgoing
0 . . . o . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
```

Run the **display ipv6 pim-info sparse-mode mroute** command to display the multicast routing tables of device D.

```
DeviceE> privilege
DeviceD# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 1
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(1012::2, ff33::100)
RPF nbr: ::
RPF idx: None
SPT bit: 1
Upstream State: JOINED
kat expires in 37 seconds
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Joined
0 . . j . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
1 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
6 . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
```

```
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .

DeviceD#
```

Run the **display ipv6 pim-info sparse-mode mroute** command to display the multicast routing tables of device E.

```
DeviceE> privilege
DeviceE# display ipv6 pim-info sparse-mode mroute
IPv6 Multicast Routing Table
(*,*,RP) Entries: 0
(*,G) Entries: 0
(S,G) Entries: 1
(S,G,rpt) Entries: 0
FCR Entries: 0
REG Entries: 0

(1000::2, ff33::100)
RPF nbr: fe80::274:9cff:feee:f49f
RPF idx: GigabitEthernet 0/7
SPT bit: 1
Upstream State: JOINED
jt_timer expires in 24 seconds
kat expires in 54 seconds
  00 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31
Local
0 . i . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
```

```
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Joined
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Asserted
0 . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
Outgoing
0 . o . . . . . .
1 . . . . . .
2 . . . . . .
3 . . . . . .
4 . . . . . .
5 . . . . . .
6 . . . . . .
7 . . . . . .
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 access-control-list ipv6-ssm-list
 10 permit ipv6 any FF33::/64
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/21
 unset switch-port
 ipv6 address 1000::1/64
 ipv6 ospf 4949 set-area 0
```

```
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/24
 unset switch-port
 ipv6 address 2000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim enable-ssm range ipv6-ssm-list
!
```

- Device B configuration file

```
hostname DeviceB
!
ipv6 access-list ipv6-ssm-list
 10 permit ipv6 any FF33::/64
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/30
 unset switch-port
 ipv6 address 8000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 mld specifiy-version 2
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/31
 unset switch-port
 ipv6 address 2000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/32
 unset switch-port
 ipv6 address 5000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 49
 ipv6 address 4949::4949/128
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
```

```
 graceful-restart
!
ipv6 pim enable-ssm range ipv6-ssm-list
!
```

- Device C configuration file

```
hostname DeviceC
!
ipv6 access-contorl-list ipv6-ssm-list
 10 permit ipv6 any FF33::/64
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/18
 unset switch-port
 ipv6 address 1011::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/20
 unset switch-port
 ipv6 address 5000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/22
 unset switch-port
 ipv6 address 4000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface Loopback 50
 ipv6 address 5050::5050/128
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim enable-ssm range ipv6-ssm-list
!
```

- Device D configuration file

```
hostname DeviceD
!
ipv6 access-control-list ipv6-ssm-list
 10 permit ipv6 any FF33::/64
```

```
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/47
 unset switch-port
 ipv6 address 1012::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/48
 unset switch-port
 ipv6 address 1011::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim enable-ssm range ipv6-ssm-list
!
```

- Device E configuration file

```
hostname DeviceE
!
ipv6 access-control-list ipv6-ssm-list
 10 permit ipv6 any FF33::/64
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/5
 unset switch-port
 ipv6 address 1013::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 mld specifiy-version 2
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/7
 unset switch-port
 ipv6 address 4000::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim enable-ssm range ipv6-ssm-list
!
```

7. Common Errors

- IPv6 unicast routing is incorrectly configured.
- IPv6 multicast routing is not enabled on a multicast device.
- The SSM function is not enabled on a multicast device, or the SSM group address range on this multicast device is not consistent with the configuration on other multicast devices.
- The PIM-SMv6 function is not enabled on an interface, for example, the gateway interface connecting to hosts or a multicast source.
- The MLDv2 function is not enabled on an interface connected to hosts.

Contents

1 Configuring MLD Snooping 1
1.1 Introduction 1
1.1.2 Principles 1
1.1.3 Port Types 3
1.1.4 Working Mechanism 3
1.1.5 Protocols and Standards 4
1.2 Configuration Task Summary 4
1.3 Configuring Basic MLD Snooping Functions 5
1.3.1 Overview 5
1.3.2 Restrictions and Guidelines 5
1.3.3 Configuration Tasks 5
1.3.4 Notes 5
1.3.5 Configuring Basic MLD Snooping Functions in IVGL Mode 6
1.3.6 Configuring Basic MLD Snooping Functions in SVGL Mode 7
1.3.7 Configuring Basic MLD Snooping Functions in IVGL-SVGL Mode 8
1.4 Configuring Protocol Packet Processing Parameters 9
1.4.1 Overview 9
1.4.2 Restrictions and Guidelines 9
1.4.3 Configuration Tasks 9
1.4.4 Configuring a Static Multicast Router Port 9
1.4.5 Configuring a Static Member Port 10
1.4.6 Configuring Report Packet Suppression 10

1.4.7 Configuring Port Fast Leave 11
1.4.8 Configuring Dynamic Multicast Router Port Learning 11
1.4.9 Configuring the Aging Time for Dynamic Multicast Router Ports 12
1.4.10 Configuring the Aging Time for Dynamic Member Ports 12
1.4.11 Configuring the Maximum Response Time for Query Packets 13
1.5 Configuring Multicast Security Control 14
1.5.1 Overview 14
1.5.2 Prerequisites 14
1.5.3 Configuration Tasks 14
1.5.4 Configuring a Profile 14
1.5.5 Configuring Multicast Group Filtering on a Port 15
1.5.6 Configuring the Maximum Number of Multicast Groups That Can Be Dynamically Learned by a Port 16
1.5.7 Configuring Source Port Check 16
1.6 Configuring QinQ Processing 17
1.6.1 Overview 17
1.6.2 Restrictions and Guidelines 17
1.6.3 Prerequisites 18
1.6.4 Procedure 18
1.7 Monitoring 18
1.8 Configuration Examples 19
1.8.1 Configuring Basic MLD Snooping Functions in IVGL Mode 19
1.8.2 Configuring Basic MLD Snooping Functions in SVGL Mode 23
1.8.3 Configuring Basic MLD Snooping Functions in IVGL-SVGL Mode 28
1.8.4 Configuring Static Ports to Implement L2 Multicast 36

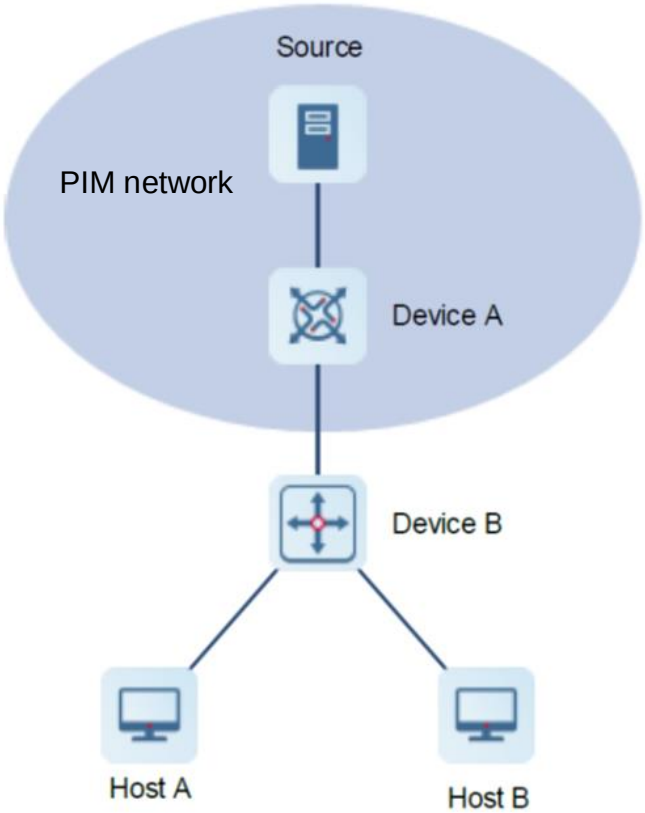
1 Configuring MLD Snooping

1.1 Introduction

Multicast Listener Discovery (MLD) snooping is a process of snooping protocol packets between an L3 multicast device and hosts to manage and control forwarding of IPv6 multicast traffic at the data link layer, implementing L2 multicast.

Generally, multicast packets need to pass through L2 switches, especially in some local area networks (LANs). As shown in Figure 1-1, the Protocol Independent Multicast (PIM) multicast device sends packets to hosts through an L2 switch. Because multicast group addresses cannot be learned on L2 devices, the packets will be broadcast in the VLAN. This wastes bandwidth and affects security. After the MLD snooping function is configured, the L2 switch can listen to MLD packets between hosts and the uplink PIM multicast device and establish L2 multicast forwarding entries to ensure that multicast data is forwarded only to specific hosts. This prevents multicast data from being broadcast in the L2 network.

Figure 1-1 Network Topology



1.1.2 Principles

MLD snooping is a basic L2 multicast function that forwards and controls multicast data at the data link layer. By listening to and analyzing protocol packets between hosts and the L3 multicast device, the L2 device running

MLD snooping establishes and maintains L2 multicast forwarding entries to ensure on-demand forwarding of multicast data at the link layer.

As shown in Figure 1-2, when MLD snooping is not configured on the L2 device, IPv6 multicast packets are broadcast in the VLAN. As shown in Figure 1-3, when MLD snooping is enabled on the L2 device, IPv6 multicast packets are forwarded only to the member hosts of the multicast group. This prevents multicast data from being broadcast in the L2 network.

Figure 1-2 Multicast Data Transmission When MLD Snooping Is Disabled

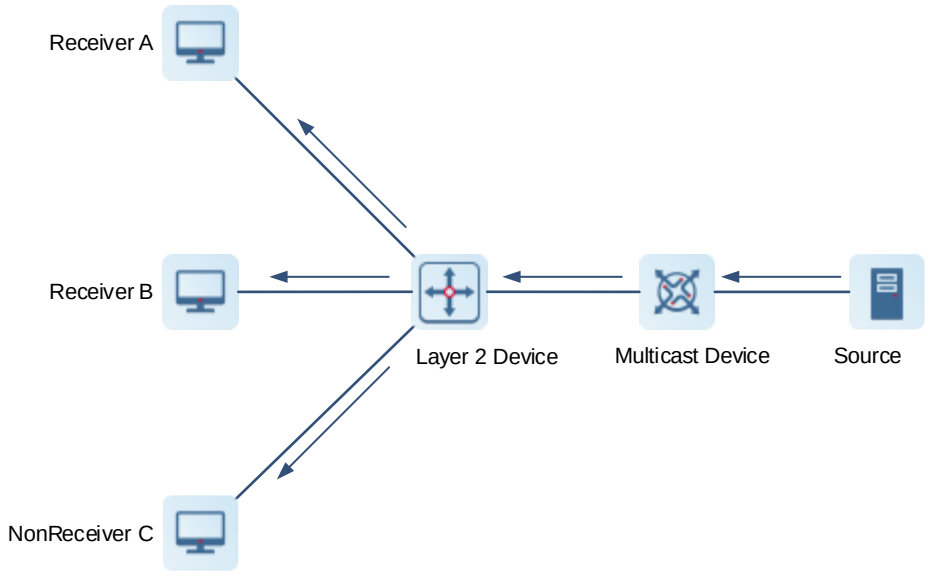
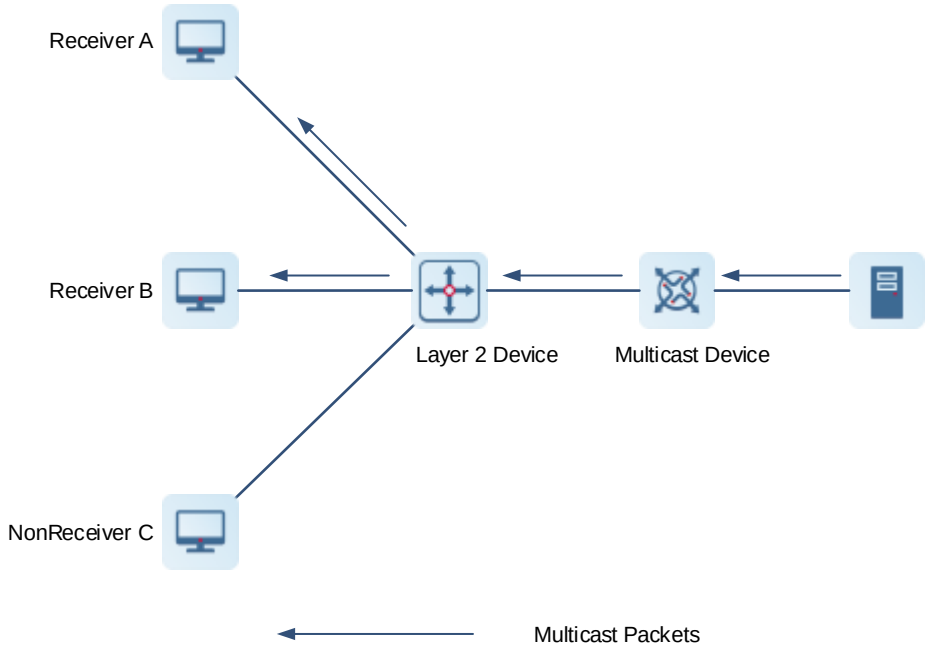


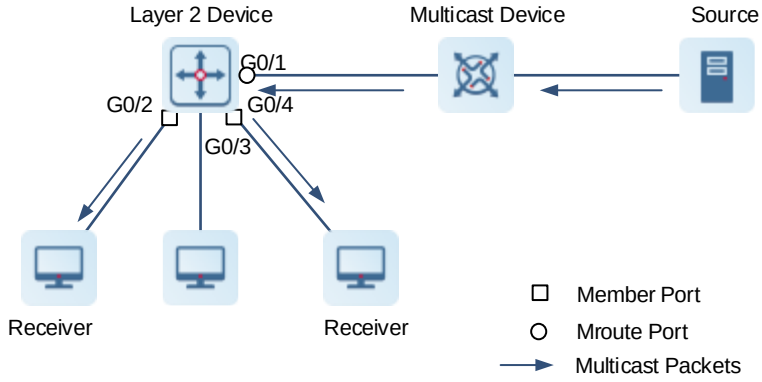
Figure 1-3 Multicast Data Transmission When MLD Snooping Is Enabled



1.1.3 Port Types

MLD snooping is enabled on a per-VLAN basis. MLD snooping ports are ports in a VLAN. The device running MLD snooping identifies ports in a VLAN as multicast router ports or member ports to manage and control forwarding of IP multicast data in the VLAN.

Figure 1-4 MLD Snooping Ports



- A multicast router port connects the L2 multicast device to an L3 multicast device, and indicates the direction of the multicast source. By listening to MLD packets, the L2 multicast device can automatically discover and maintain dynamic multicast router ports. You can also configure static multicast router ports.
- A member port connects the L2 multicast device to a host in a multicast group, and indicates the direction of a multicast group member. It is also called a listener port. By listening to MLD packets, the L2 multicast device can automatically discover and maintain dynamic member ports. You can also configure static member ports.

As shown in Figure 1-4, MLD snooping is enabled on the L2 multicast device, and the multicast data is received from a multicast router port and sent out from a member port.

1.1.4 Working Mechanism

A device running MLD snooping analyzes received MLD packets to discover and identify multicast router ports and member ports and establish and maintain MLD snooping forwarding entries. The device can identify and process the following types of MLD packets:

1. Query Packets

An L3 multicast device regularly sends an MLD general Query packet to all hosts and other multicast devices (with the address of FF02::1) in the local network segment to query the IPv6 multicast group members in this network segment. When receiving the MLD general Query packet, a multicast device forwards the packet through all ports in the VLAN except the one receiving the packet and processes the port receiving the packet as follows:

- Reset the aging timer of the port if the port is in the multicast router port list.
- Add the port to the multicast router port list if the port is not in the multicast router port list and start the aging timer of the port.
- When the L2 multicast device receives an MLD group-specific Query packet, it does not update the preceding timers.

2. Report Packets

When a member host receives a Query packet, the host will respond with a Report packet. If a host wants to join a multicast group, the host will proactively send a Report packet.

When receiving a Report packet, the device running MLD snooping forwards it to all multicast router ports in the VLAN, retrieves the address of the IPv6 multicast group that the host needs to join from the packet, and processes the port receiving the packet as follows:

- The device creates a forwarding entry for the IPv6 multicast group if no forwarding entry for the IPv6 multicast group exists, adds the port to the outbound port list as a dynamic member port, and starts the aging timer of the port.
- The device adds the port to the outbound port list as a dynamic member port and starts its aging timer if the forwarding entry of the IPv6 multicast group exists but the port is not contained in the outbound port list.
- The device resets the aging timer of the port if the forwarding entry of the IPv6 multicast group exists and the port is contained in the outbound port list.

3. Leave Packets

When a host leaves an IPv6 multicast group, it sends an MLD Leave packet (with the address of FF02::2) to notify the multicast device. When receiving a Leave packet, the device running MLD snooping performs the following actions in the VLAN:

- The device forwards the Leave packet through all multicast router ports.
- If the port receiving the Leave packet is a dynamic member port and the fast leave function is configured, the device immediately deletes the port from the MLD snooping forwarding entry of the corresponding multicast group. The port is no longer used as a dynamic member port of the group.
- If the port receiving the Leave packet is a dynamic member port but the fast leave function is not configured, the port status remains unchanged. When the aging timer of the port times out, the device deletes the port from the MLD snooping forwarding entry of the corresponding multicast group. The port is no longer used as a dynamic member port of the group.

1.1.5 Protocols and Standards

- RFC 4541: Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches

1.2 Configuration Task Summary

MLD snooping configuration includes the following tasks:

(1) Configuring Basic MLD Snooping Functions Select either of the following configuration tasks to configure.

- Configuring Basic MLD Snooping Functions in IVGL Mode
- Configuring Basic MLD Snooping Functions in SVGL Mode
- Configuring Basic MLD Snooping Functions in IVGL-SVGL Mode

(2) Configuring Protocol Packet Processing Parameters All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Static Multicast Router Port
- Configuring a Static Member Port

- Configuring Report Packet Suppression
- Configuring Port Fast Leave
- Configuring Dynamic Multicast Router Port Learning
- Configuring the Aging Time for Dynamic Multicast Router Ports
- Configuring the Aging Time for Dynamic Member Ports
- Configuring the Maximum Response Time for Query Packets

(3) Configuring Multicast Security Control All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Profile
- Configuring Multicast Group Filtering on a Port
- Configuring the Maximum Number of Multicast Groups That Can Be Dynamically Learned by a Port
- Configuring Source Port Check

(4) (Optional) Configuring QinQ Processing

1.3 Configuring Basic MLD Snooping Functions

1.3.1 Overview

The device running MLD snooping can provide independent or shared multicast services for user VLANs when operating in the following modes:

- Independent VLAN Group Learning (IVGL): provides independent multicast services for each user VLAN.
- Shared VLAN Group Learning (SVGL): provides shared multicast services for multiple user VLANs.
- IVGL-SVGL: provides both shared and independent multicast services for user VLANs.

1.3.2 Restrictions and Guidelines

- The SVGL mode and the IPv6 multicast function are mutually exclusive. If the IPv6 multicast function needs to be enabled on the device running MLD snooping, the device can operate only in IVGL mode.
- If the device running MLD snooping operates in SVGL or IVGL-SVGL mode, a profile must be defined to specify the multicast groups associated with the SVGL mode.

1.3.3 Configuration Tasks

Basic MLD snooping function configuration includes the following tasks. Please configure only one task.

- Configuring Basic MLD Snooping Functions in IVGL Mode
- Configuring Basic MLD Snooping Functions in SVGL Mode
- Configuring Basic MLD Snooping Functions in IVGL-SVGL Mode

Basic MLD snooping function configuration includes the following tasks.

1.3.4 Notes

The following table lists products supporting different basic MLD snooping functions.

Table 1-1 Products Supporting Different Basic MLD Snooping Functions

Function	Product Supporting the Function
Configuring Basic MLD Snooping Functions in IVGL Mode	S6980-64QC S6920-4C S6580-24DC8QC S6510-32CQ S6510-48VS8CQ S6250-48XS8CQ
Configuring Basic MLD Snooping Functions in SVGL Mode	S6980-64QC S6920-4C S6580-24DC8QC S6510-32CQ S6510-48VS8CQ S6250-48XS8CQ
Configuring Basic MLD Snooping Functions in IVGL-SVGL Mode	S6980-64QC S6920-4C S6580-24DC8QC S6510-32CQ S6510-48VS8CQ S6250-48XS8CQ

1.3.5 Configuring Basic MLD Snooping Functions in IVGL Mode

1. Overview

In IVGL mode, the device running MLD snooping provides independent multicast services for each user VLAN. Multicast traffic can be forwarded only in the belonged VLAN, and hosts can request multicast traffic only in the belonged VLAN.

2. Restrictions and Guidelines

- Unless otherwise specified, you are advised to enable MLD snooping globally on all L2 access devices connecting to hosts.
- After MLD snooping is enabled globally, it takes effect to all VLANs. You can disable MLD snooping on any VLAN. When the multicast service is disabled on a VLAN, the multicast services on other VLANs are not affected.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IPv6 MLD snooping function and run the IVGL mode.

ipv6 mld snooping-mode ivgl

MLD snooping is disabled by default.

(4) (Optional) Disable MLD snooping on a specified VLAN.

unset ipv6 mld snooping-mode vlan *vlan-id*

After MLD snooping is enabled globally, it takes effect to all VLANs by default. If MLD snooping is disabled globally, it is ineffective on all VLANs.

You can disable MLD snooping on a VLAN only after MLD snooping is enabled globally.

1.3.6 Configuring Basic MLD Snooping Functions in SVGL Mode

1. Overview

In SVGL mode, the device running MLD snooping can provide shared multicast services for multiple user VLANs. Shared multicast services are usually used to provide the same video on demand (VOD) service for users in multiple VLANs. Compared with independent multicast services, shared multicast services save bandwidth.

VLANs for shared multicast services are classified into a shared VLAN and sub VLANs. Multicast traffic of multicast groups applied in SVGL mode on the shared VLAN is forwarded from the shared VLAN to sub VLANs, and hosts on sub VLANs request multicast traffic of multicast groups applied in SVGL mode from the shared VLAN.

2. Restrictions and Guidelines

- Multicast services can be shared among VLANs. After configuring the SVGL mode, you need to configure multicast groups applied in SVGL mode.
- Shared multicast services apply only to the shared VLAN and sub VLANs and use group addresses applied in SVGL mode.

3. Prerequisites

Before configuring multicast groups associated with the SVGL mode, ensure that the corresponding profile is created and a range of multicast groups that are permitted or denied by the filter is specified. For details about profile configuration, see 1.5.4 Configuring a Profile.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IPv6 MLD snooping function and run the SVGL mode.

ipv6 mld snooping-mode svgl

MLD snooping is disabled by default.

(4) Configure multicast groups associated with the SVGL mode.

ipv6 mld snooping-mode svgl group-profile *profile-number*

No multicast group is associated with the SVGL mode by default.

(5) (Optional) Specify a shared VLAN.

ipv6 mld snooping-mode svgl shared-vlan *vlan-id*

The default shared VLAN in SVGL mode is VLAN 1.

1.3.7 Configuring Basic MLD Snooping Functions in IVGL-SVGL Mode

1. Overview

The IVGL-SVGL mode is also called the hybrid mode. In IVGL-SVGL mode, the device running MLD snooping can provide both shared and independent multicast services for user VLANs.

- In the shared VLAN and sub VLANs, shared multicast services apply to the multicast traffic of multicast groups applied in SVGL mode, and independent multicast services apply to other multicast traffic.
- In other VLANs (except the shared VLAN and sub VLANs), independent multicast services are provided.

2. Restrictions and Guidelines

- After configuring the IVGL-SVGL mode, you must specify the multicast groups associated with the SVGL mode.
- When a user VLAN is configured as a shared VLAN or sub VLAN, the user VLAN enjoys both shared and independent multicast services. When a user VLAN is configured as other VLANs, it enjoys only independent multicast services.

3. Prerequisites

Before configuring multicast groups associated with the SVGL mode, ensure that the corresponding profile is created and a range of multicast groups that are permitted or denied by the filter is specified. For details about profile configuration, see 1.5.4 Configuring a Profile.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the IPv6 MLD snooping function and run the IVGL-SVGL mode.

ipv6 mld snooping-mode ivgl-svgl

MLD snooping is disabled by default.

(4) (Optional) Specify a shared VLAN.

ipv6 mld snooping-mode svgl shared-vlan *vlan-id*

The default shared VLAN in SVGL mode is VLAN 1.

1.4 Configuring Protocol Packet Processing Parameters

1.4.1 Overview

By controlling protocol packet processing, an L2 multicast device can establish static or dynamic multicast forwarding entries. In addition, the device can adjust parameters to refresh dynamic multicast forwarding entries and MLD snooping membership quickly.

1.4.2 Restrictions and Guidelines

Configuring Basic MLD Snooping Functions Related configuration functions take effect only after basic MLD snooping functions are configured.

1.4.3 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Static Multicast Router Port
- Configuring a Static Member Port
- Configuring Report Packet Suppression
- Configuring Port Fast Leave
- Configuring Dynamic Multicast Router Port Learning
- Configuring the Aging Time for Dynamic Multicast Router Ports
- Configuring the Aging Time for Dynamic Member Ports
- Configuring the Maximum Response Time for Query Packets

1.4.4 Configuring a Static Multicast Router Port

1. Overview

A multicast router port is used to receive uplink multicast data and forward MLD Report and Leave packets. Static multicast router ports never age and can forward MLD Report and Leave packets to the uplink MLD querier stably.

2. Restrictions and Guidelines

- In SVGL mode, if no sub VLAN is configured, the configuration of static multicast router ports is valid only in the shared VLAN. In other VLANs, static multicast router ports can be configured but do not take effect. If sub VLANs are configured, the configuration of static multicast router ports is valid in the shared VLAN and non-sub VLANs. In sub VLANs, static multicast router ports can be configured but do not take effect.
- In SVGL-IVGL mode, if no sub VLAN is configured, the configuration of static multicast router ports is valid in all VLANs. If sub VLANs are configured, the configuration of static multicast router ports is valid in the shared VLAN and non-sub VLANs. In sub VLANs, static multicast router ports can be configured but do not take effect.
- In IVGL mode, the configuration of static multicast router ports is valid in all VLANs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static multicast router port.

ipv6 mld snooping-mode vlan *vlan-id* **multicast-router interface** *interface-type interface-number*

No static multicast router port is configured by default.

1.4.5 Configuring a Static Member Port

1. Overview

When a member port connecting to a member host is configured as a static member port, the host can receive multicast traffic from the specified multicast group no matter whether the host joins the multicast group. Static member ports never age.

2. Restrictions and Guidelines

- In SVGL mode, if no sub VLAN is configured, the configuration of static member ports is valid only in the shared VLAN. In other VLANs, static member ports can be configured but do not take effect. If sub VLANs are configured, the configuration of static member ports is valid in the shared VLAN and non-sub VLANs. In sub VLANs, static member ports can be configured but do not take effect.
- In SVGL-IVGL mode, if no sub VLAN is configured, the configuration of static member ports is valid in all VLANs. If sub VLANs are configured, the configuration of static member ports is valid in the shared VLAN and non-sub VLANs. In sub VLANs, static member ports can be configured but do not take effect.
- In IVGL mode, the configuration of static member ports is valid in all VLANs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static member port.

ipv6 mld snooping-mode vlan *vlan-id* **static-member** *ipv6-group-address* **interface** *interface-type interface-number*

No static member port is configured by default.

1.4.6 Configuring Report Packet Suppression

1. Overview

MLD Query and Report packets are exchanged periodically to maintain the group membership. When many hosts in a network join the same multicast group, many MLD packets will be sent to the MLD multicast device, which wastes network bandwidth and affects the performance of the MLD multicast device. Report packet suppression can optimize this situation.

When Report packet suppression is configured, the MLD multicast device forwards only the first Report packet from a specific VLAN for a multicast group to the multicast router port and suppresses subsequent Report

packets for the same multicast group during one query interval. This function helps reduce the number of packets in the network.

2. Restrictions and Guidelines

Only MLDv1 Report packets can be suppressed, and MLDv2 Report packets cannot be suppressed.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure Report packet suppression.

ipv6 mld snooping-mode suppress enable

Report packet suppression is disabled by default.

1.4.7 Configuring Port Fast Leave

1. Overview

When the port fast leave function is enabled and a port receives a Leave packet (including the MLDv1 Leave packet and MLDv2 Report packet of the INCLUDE type without carrying any source address), the port is directly deleted from the member port list of the corresponding multicast forwarding entry. When receiving group-specific Query packets and multicast data, the device does not forward the packets to this port.

2. Restrictions and Guidelines

The port fast leave function is applicable when only one host is connected to each port. This function helps save bandwidth and resources.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the port fast leave function.

ipv6 mld snooping-mode inf-fast-leave enable

The port fast leave function is disabled by default.

1.4.8 Configuring Dynamic Multicast Router Port Learning

1. Overview

In some test scenarios, a user host may send Query packets and PIMv6 neighbor messages. When the L2 device receives such Query packets or PIM neighbor messages, the L2 device will set the port receiving the packets as a dynamic multicast router port. All multicast packets in the VLAN will be forwarded to this port, and the host receives a large number of useless multicast packets. In addition, Query packets and PIMv6 neighbor messages sent by user hosts may affect the status of the L3 multicast routing protocol, such as the querier and

designated router (DR) election. In this case, you can disable dynamic multicast router port learning to resolve this problem and improve network security.

2. Restrictions and Guidelines

Disabling dynamic multicast router port learning and enabling a static multicast router port do not affect each other.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure dynamic multicast router port learning.

ipv6 mld snooping-mode [**vlan** *vlan-id*] **multicast-router learn**

Dynamic multicast router port learning is enabled by default.

1.4.9 Configuring the Aging Time for Dynamic Multicast Router Ports

1. Overview

A multicast router port is used to receive uplink multicast data and forward MLD Report and Leave packets. After enabling MLD snooping, an L2 device can dynamically learn multicast router ports to forward multicast packets for the uplink device. If a dynamic multicast router port does not receive a Query packet or neighbor message from the uplink device within the aging time due to unstable network or packet congestion, the L2 device will delete the dynamic multicast router port, which may cause multicast data interruption.

2. Restrictions and Guidelines

You can adjust the aging time of dynamic multicast router ports based on the network load. When the network load is heavy, set the aging time to a larger value. A too short aging time may cause dynamic multicast router ports to be added and deleted frequently, resulting in data interruption.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the aging time for dynamic multicast router ports.

ipv6 mld snooping-mode dyn-mr-aging-interval *dynamic-mroute-aging-interval*

The default aging time of dynamic multicast router ports is 300s.

1.4.10 Configuring the Aging Time for Dynamic Member Ports

1. Overview

When the device running MLD snooping receives an MLD Join packet from a host to join an IP multicast group, the device adds the port receiving the packet to the member port list and sets an aging time for the port. If the

port is already in the member port list, the device resets the aging timer of the port. If the timer times out, it is deemed that no user host receives multicast packets through this port, and the multicast device deletes the port from the MLD snooping member port list.

2. Restrictions and Guidelines

- You can adjust the aging time of dynamic member ports based on the network load. When the network load is heavy, set the aging time to a larger value. A too short aging time may cause dynamic member ports to be added and deleted frequently, resulting in data interruption.
- After this command is configured, the aging timer value of dynamic member ports is *host-aging-interval* for subsequent MLD Join packets. The aging time takes effect immediately after configuration and the started member port timers are updated.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the aging time for dynamic member ports.

ipv6 mld snooping-mode host-aging-interval *host-aging-interval*

The default aging time of dynamic member ports is 260s.

1.4.11 Configuring the Maximum Response Time for Query Packets

1. Overview

When hosts in the directly-connected network segment of the device serving as the MLD snooping querier receive a Query packet from the device, the hosts need to return a Report packet within the maximum response time. This function allows you to configure the maximum response time for Query packets on the device. If no host returns a Report packet within the maximum response time, the device considers that no group member exists in the directly-connected network segment and deletes the group information.

2. Restrictions and Guidelines

- When receiving an MLD general Query packet, the multicast device will update the aging timers of all member ports. The timer time is the maximum response time for Query packets. After the timer expires, the device regards that no group member receives multicast traffic through a port and deletes the port from the MLD snooping forwarding table.
- When receiving an MLD group-specific Query packet, the multicast device will start the aging timers of all member ports of the specific group. The timer time is the maximum response time for Query packets. After the timer expires, the device regards that no group member receives multicast traffic through a port and deletes the port from the MLD snooping forwarding table.
- For MLDv2 group-specific Query packets, the multicast device does not update the timers.
- The configured maximum response time for Query packets takes effect when the next Query packet is received.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the maximum response time for Query packets.

ipv6 mld snooping-mode max-response-time *max-response-time*

The maximum response time for Query packets is 10s by default.

1.5 Configuring Multicast Security Control

1.5.1 Overview

After multicast security control is configured, the device running MLD snooping can control the multicast service scope and load to prevent invalid multicast traffic and improve L2 multicast network security.

1.5.2 Prerequisites

Configuring Basic MLD Snooping Functions

1.5.3 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Profile
- Configuring Multicast Group Filtering on a Port
- Configuring the Maximum Number of Multicast Groups That Can Be Dynamically Learned by a Port
- Configuring Source Port Check

1.5.4 Configuring a Profile

1. Overview

A profile is used to define a range of multicast groups that permit or deny user access for reference by other functions.

- When the SVGL mode is enabled, a profile is used to define a range of multicast groups applied in SVGL mode.
- When multicast filtering is configured on a port, a profile is used to define a range of multicast groups that permit or deny user access through the port.
- When multicast filtering is configured on a VLAN, a profile is used to define a range of multicast groups that permit or deny user access on the VLAN.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a profile.

ipv6 mld group-profile *profile-number*

No profile is created by default.

(4) Define a range of multicast groups for a profile.

range *start-ipv6-address* [*end-ipv6-address*]

No multicast group range is defined for a profile by default.

Multiple multicast group ranges can be configured at the same time. If two multicast group ranges overlap, the ranges will be merged automatically.

(5) Configure the filtering action for a profile. The configuration steps below are mutually exclusive. Please configure only one task.

- Set the filtering action of a profile to Deny.

  deny

  If only the Deny action is configured and no multicast group range is configured, no group is denied. The effect is the same as permitting all groups.

- Set the filtering action of a profile to Permit.

  permit

  If only the Permit action is configured and no multicast group range is configured, no group is permitted. The effect is the same as denying all groups.

The Deny action is performed for a profile by default.

1.5.5 Configuring Multicast Group Filtering on a Port

1. Overview

Generally, the device running ports can join any multicast group. By configuring a range of multicast groups that permit or deny user access, you can customize the multicast service scope for users to guarantee the interest of operators and prevent invalid multicast traffic.

2. Restrictions and Guidelines

By applying a profile to a port, you can restrict the multicast groups that users can join on the port.

3. Prerequisites

Before configuring a range of multicast groups for a profile, ensure that the corresponding profile is created and a range of multicast groups that are permitted or denied by the filter is specified. For details about profile configuration, see 1.5.4 Configuring a Profile.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 2 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*

(4) Configure multicast group filtering on a port.

ipv6 mld snooping-mode group-filter *profile-number*

The multicast group filtering function is disabled on a port by default.

1.5.6 Configuring the Maximum Number of Multicast Groups That Can Be Dynamically Learned by a Port

1. Overview

If too much multicast traffic is requested concurrently, the multicast device will be severely burdened. Configuring the maximum number of multicast groups allowed for concurrent request can guarantee the bandwidth. You can limit the number of multicast groups allowed for concurrent request on a port.

2. Restrictions and Guidelines

Only dynamically learned multicast groups on a port are counted. Statically configured multicast groups are excluded.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.
  - **interface** *ethernet-type interface-number*
- Enter the Layer 2 link aggregation configuration mode.
  - **interface aggregateport** *interface-number*

(4) Configure the maximum number of multicast groups that can be dynamically learned by a port.

ipv6 mld snooping-mode max-member-record *max-groups-number*

The maximum number of multicast groups that can be dynamically learned by a port is 64,000 by default.

1.5.7 Configuring Source Port Check

1. Overview

The source port check function restricts users to receive only multicast traffic on multicast router ports to prevent users from sending invalid multicast traffic.

- After the source port check function is enabled, only multicast traffic received on multicast router ports is

valid. Multicast traffic received on other ports is invalid and will be discarded.

- When the source port check function is disabled, multicast traffic received on any port is valid.

2. Restrictions and Guidelines

- To restrict users to receive multicast traffic only on multicast router ports, configure this function.
- If no multicast router port exists after the source port check function is enabled, the device will discard received multicast traffic.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure source port check.

ipv6 mld snooping-mode source-inf validate

Source port check is disabled by default.

1.6 Configuring QinQ Processing

Note

This command is not supported by the S6980-64QC, or S6580-24DC8QC.

1.6.1 Overview

On a device with MLD snooping enabled and a dot1q-tunnel (QinQ) port configured, MLD snooping will process the MLD packets received on the QinQ port using the following two modes:

- Mode 1: Create multicast entries on the VLAN where the MLD packets are located. Forwarding MLD packets on the VLAN where these packets are located is called transparent transmission. For example, MLD snooping is enabled for a device, Port A on the device is designated as the QinQ port, the default VLAN of this port is VLAN 1, and Port A allows the passage of VLAN 1 and VLAN 10 packets. When an MLD Report packet is sent from VLAN 10 to Port A, MLD snooping establishes a multicast entry for VLAN 10 and forwards the MLD Report packet to the multicast router port of VLAN 10.
- Mode 2: Create multicast entries on the default VLAN of the QinQ port. Encapsulate the MLD packets with the VLAN tag of the default VLAN where the QinQ port is located and forward the packets within the default VLAN. For example, MLD snooping is enabled for a device, Port A on the device is designated as the QinQ port, the default VLAN of this port is VLAN 1, and Port A allows the passage of VLAN 1 and VLAN 10 packets. When an MLD Report packet is sent from VLAN 10 to Port A, MLD snooping establishes a multicast entry for VLAN 1, encapsulates the MLD Report packet with the tag of VLAN 1, and forwards the packet to the multicast router port of VLAN 1.

1.6.2 Restrictions and Guidelines

If the QinQ port needs to forward multicast packets on the VLANs specified by the VLAN IDs (VIDs) of the packets, transparent packet transmission needs to be configured on the QinQ port.

1.6.3 Prerequisites

- Configuring Basic MLD Snooping Functions.
- Configure basic QinQ functions. For details about QinQ configuration, see *Configuring QinQ*.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure transparent packet transmission on the QinQ port.

ipv6 mld snooping-mode tunnel

The function of transparent MLD packet transmission on the QinQ port is disabled by default.

1.7 Monitoring

Run the display command to check the running status of a configured function to verify the configuration effect.

Run the cleanup commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-2 Monitoring

Command	Purpose
cleanup ipv6 mld-cache snooping gda-table	Clears MLD snooping multicast forwarding entries.
cleanup ipv6 mld-cache snooping statistics	Clears MLD snooping statistics.
display ipv6 mld-info snooping	Displays the current MLD snooping mode.
display ipv6 mld-info snooping gda-table	Displays MLD snooping forwarding entries.
display ipv6 mld-info snooping statistics	Displays MLD snooping statistics.
display ipv6 mld-info snooping mrouter	Displays MLD snooping multicast router ports.
display ipv6 mld-info snooping interfaces [*interface-type interface-type interface-number*]	Displays MLD snooping interface information, interface filtering profile, and maximum number of multicast groups that can be dynamically learned by a port.

Command	Purpose
display ipv6 mld-info snooping vlan *vlan-id*	Displays multicast information about a single VLAN, on which MLD snooping is configured.
display ipv6 mld-info group-profile *profile-number*	Displays an MLD profile.

1.8 Configuration Examples

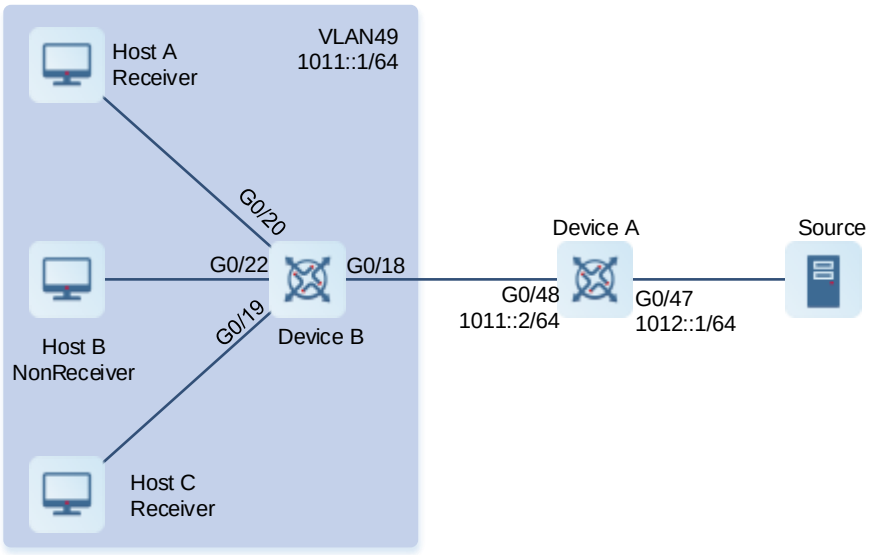
1.8.1 Configuring Basic MLD Snooping Functions in IVGL Mode

1. Requirements

In the multicast network shown in Figure 1-5, the router (Device A) connects to the user network through the switch (Device B), and MLD snooping is run on Device A. The multicast source sends data to multicast group FF14::10. There are three hosts (Host A, Host B, and Host C) in the network. Host A and Host C join multicast group FF14::10. Only Host A and Host B can receive multicast data that the multicast source sends to multicast group FF14::10.

2. Topology

Figure 1-5 Topology of Basic MLD Snooping Functions in IVGL Mode



3. Notes

Configure basic MLD snooping functions on Device B.

- Configure the IPv6 addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.
- Enable multicast routing and PIM-SMv6 related functions on Device A.

- Enable MLD snooping and run the IVGL mode on Device B.

4. Procedure

(1) Configure the IPv6 addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.

Configure the IPv6 address and unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 address 1012::1/64
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/48)# ipv6 address 1011::2/64
DeviceA(cfg-if-GigabitEthernet 0/48)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# ipv6 router-protocol ospf 4949
DeviceA(cfg-router)# exit
```

Configure the IPv6 address, VLAN, and unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan 49
DeviceB(cfg-vlan)# exit
DeviceB(cfg)# interface vlan 49
DeviceB(cfg-if-VLAN 49)# ipv6 address 1011::1/64
DeviceB(cfg-if-VLAN 49)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 49)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/19)# exit
DeviceB(cfg)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/20)# exit
DeviceB(cfg)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/22)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/22)# exit
DeviceB(cfg)# ipv6 router-protocol ospf 4949
DeviceB(cfg-router)# exit
```

(2) Enable multicast routing and PIM-SMv6 related functions on Device A.

```
DeviceA(cfg)# ipv6 mrouting-enable
```

```
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# ipv6 pim c-bsr GigabitEthernet 0/47
DeviceA(cfg)# ipv6 pim c-rp GigabitEthernet 0/47
```

(3) Enable MLD snooping and run the IVGL mode on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 mld snooping-mode ivgl
```

5. Verification

Send packets from the multicast source (1012::2) to multicast group G (FF14::10). Enable Host A and Host C to join G.

Run the **display ipv6 mld-info snooping-mode gda-table** command on Device B to display the MLD snooping forwarding entry and check whether the member port list contains only GigabitEthernet 0/19 and GigabitEthernet 0/20.

```
DeviceB# display ipv6 mld-info snooping-mode gda-table
Multicast Switching Cache Table
  D: DYNAMIC
  S: STATIC
  M: MROUTE
(*, FF14::10, 49):
  VLAN(49) 3 OPORTS:
    GigabitEthernet 0/18(M)
    GigabitEthernet 0/19(M)
    GigabitEthernet 0/20(DM)
```

Run the **display ipv6 mld-info snooping-mode** command on Device B and check whether the MLD snooping working mode is IVGL.

```
DeviceB# display ipv6 mld-info snooping-mode
MLD-snooping mode: IVGL
Source port check: Disable
MLD Inf-Fast-Leave: Disable
MLD Report suppress: Disable
Query Max Response Time: 10 (Seconds)
Dynamic Mroute Aging Time: 300(Seconds)
Dynamic Host Aging Time: 260(Seconds)
MLD Snooping version: 1
MLD Tunnel: Disable

vlan 49
-------------
```

```
MLD Snooping state: Enabled
Multicast router learning mode: Enable
MLD Inf-Fast-Leave: Enabled
MLD VLAN Mode: STATIC
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/47
 unset switch-port
 ipv6 address 1012::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/48
 unset switch-port
 ipv6 address 1011::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 pim c-bsr GigabitEthernet 0/47
ipv6 pim c-rp GigabitEthernet 0/47
!
```

- Device B configuration file

```
hostname DeviceB
!
vlan range 1,49
!
interface GigabitEthernet 0/18
 switch-port access vlan 49
!
interface GigabitEthernet 0/19
 switch-port access vlan 49
!
interface GigabitEthernet 0/20
 switch-port access vlan 49
!
interface GigabitEthernet 0/22
 switch-port access vlan 49
!
interface VLAN 49
 ipv6 address 1011::1/64
```

```
 ipv6 ospf 4949 set-area 0
!
Ipv6 mld snooping-mode ivgl
!
```

7. Common Errors

- The working mode of MLD snooping is improper.

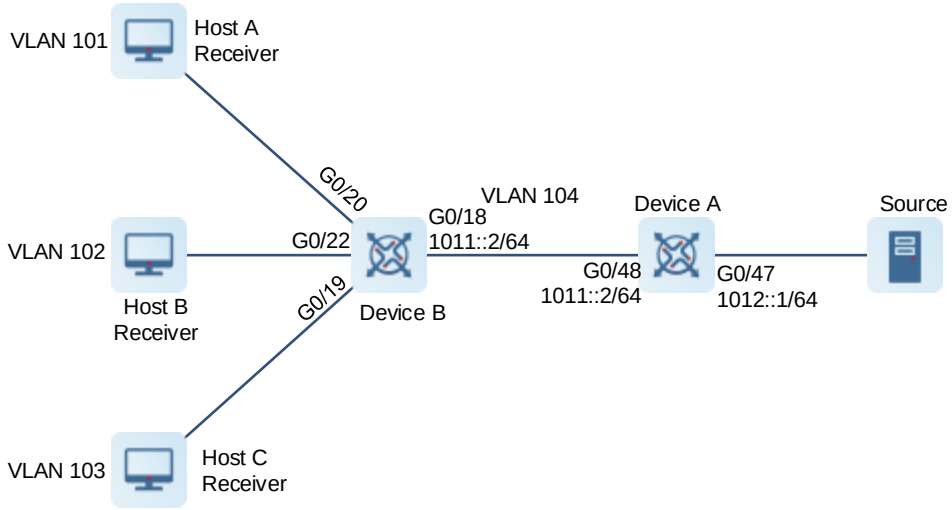
1.8.2 Configuring Basic MLD Snooping Functions in SVGL Mode

1. Requirements

In the multicast network shown in Figure 1-6, the router (Device A) connects to the user network through the switch (Device B), MLD snooping is run on Device A, and Device A directly connects to the multicast source. The multicast source sends data to multicast group FF14::10. Host A, Host B, and Host C are connected to VLAN 101, VLAN 102, and VLAN 103, respectively, and the three hosts join multicast group FF14::10. Host A, Host B, and Host C can receive multicast data that the multicast source sends to multicast group FF14::10.

2. Topology

Figure 1-6 Topology of Basic MLD Snooping Functions in SVGL Mode



3. Notes

Configure basic MLD snooping functions on Device B.

- Configure the IPv6 addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.
- Enable multicast routing and PIM-SMv6 related functions on Device A.
- Enable MLD snooping and run the SVGL mode on Device B.
- Configure the multicast groups associated with the MLD snooping SVGL mode on Device B and specify the shared VLAN and sub VLANs.

4. Procedure

(1) Configure the IPv6 addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.

Configure the IP address and unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 address 1012::1/64
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# vlan 104
DeviceA(cfg)# interface vlan 104
DeviceA(cfg-if-VLAN 104)# ipv6 address 1011::2/64
DeviceA(cfg-if-VLAN 104)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-VLAN 104)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet 0/48)# switch-port trunk native vlan 104
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# ipv6 router-protocol ospf 4949
DeviceA(cfg-router)# exit
```

Configure the IPv6 address, VLAN, and unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan range 101-104
DeviceB(cfg-vlan-range)# exit
DeviceB(cfg)# interface VLAN 101
DeviceB(cfg-if-VLAN 101)# ipv6 address 1010::1/64
DeviceB(cfg-if-VLAN 101)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 101)# exit
DeviceB(cfg)# interface VLAN 102
DeviceB(cfg-if-VLAN 102)# ipv6 address 1020::1/64
DeviceB(cfg-if-VLAN 102)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 102)# exit
DeviceB(cfg)# interface VLAN 103
DeviceB(cfg-if-VLAN 103)# ipv6 address 1030::1/64
DeviceB(cfg-if-VLAN 103)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 103)# exit
DeviceB(cfg)# interface VLAN 104
DeviceB(cfg-if-VLAN 104)# ipv6 address 1011::2/64
DeviceB(cfg-if-VLAN 104)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 104)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port mode trunk
```

```
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port trunk native vlan 104
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 103
DeviceB(cfg-if-GigabitEthernet 0/19)# exit
DeviceB(cfg)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 101
DeviceB(cfg-if-GigabitEthernet 0/20)# exit
DeviceB(cfg)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/22)# switch-port access vlan 102
DeviceB(cfg-if-GigabitEthernet 0/22)# exit
DeviceB(cfg)# ipv6 router-protocol ospf 4949
DeviceB(cfg-router)# exit
```

(2) Enable multicast routing and PIM-SM related functions on Device A.

```
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface vlan 104
DeviceA(cfg-if-VLAN 104)# ipv6 pim sparse-mode
DeviceA(cfg-if-VLAN 104)# exit
DeviceA(cfg)# ipv6 pim c-bsr GigabitEthernet 0/47
DeviceA(cfg)# ipv6 pim c-rp GigabitEthernet 0/47
```

(3) Enable MLD snooping and run the IVGL mode on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 mld snooping-mode svgl
```

(4) Configure the multicast groups associated with the MLD snooping SVGL mode on Device B and specify the shared VLAN.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 mld group-profile 1
DeviceB(cfg-profile)# permit
DeviceB(cfg-profile)# range ff13::1 ff14::fff
DeviceB(cfg-profile)# exit
DeviceB(cfg)# ipv6 mld snooping-mode svgl group-profile 1
DeviceB(cfg)# ipv6 mld snooping-mode svgl shared-vlan 104
```

5. Verification

Send packets from the multicast source (1012::2) to multicast group G (FF14::10). Enable Host A, Host B, and Host C to join G.

Run the **display ipv6 mld-info snooping gda-table** command on Device B to display the MLD snooping forwarding entry and check whether the member port list contains GigabitEthernet 0/19, GigabitEthernet 0/20, and GigabitEthernet 0/22.

```
DeviceB# display ipv6 mld-info snooping gda-table
Multicast Switching Cache Table
   D: DYNAMIC
   S: STATIC
   M: MROUTE
(*, FF14::10, 104):
   VLAN(104) 1 OPORTS:
      GigabitEthernet 0/18(M)

   VLAN(101) 1 OPORTS:
      GigabitEthernet 0/20(D)

   VLAN(103) 1 OPORTS:
      GigabitEthernet 0/19(D)

   VLAN(102) 1 OPORTS:
      GigabitEthernet 0/22(D)
```

Run the **display ipv6 mld-info snooping** command on Device B and check whether the MLD snooping working mode is SVGL.

```
DeviceB# display ipv6 mld-info snooping
MLD-snooping mode: SVGL
SVGL vlan: 104
SVGL profile number: 1
Source port check: Disable
MLD Fast-Leave: Disable
MLD Report suppress: Disable
Query Max Response Time: 10 (Seconds)
Dynamic Mroute Aging Time: 300(Seconds)
Dynamic Host Aging Time: 260(Seconds)
MLD Snooping version: 1
MLD Tunnel: Disable
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 mrouting-enable
!
vlan 104
!
interface GigabitEthernet 0/47
 unset switch-port
 ipv6 address 1012::1/64
 ipv6 ospf 4949 set-area 0
```

```
 ipv6 pim sparse-mode
interface GigabitEthernet 0/48
 switch-port mode trunk
 switch-port trunk native vlan 104
!
interface VLAN 104
 ipv6 address 1011::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode!
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim c-bsr GigabitEthernet 0/47
ipv6 pim c-rp GigabitEthernet 0/47
!
```

- Device B configuration file

```
hostname DeviceB
!
ipv6 mld group-profile 1
 permit
 range FF13::1 FF14::FFF
!
vlan range 1,101-104
!
interface GigabitEthernet 0/18
 switch-port mode trunk
 switch-port trunk native vlan 104
!
interface GigabitEthernet 0/19
 switch-port access vlan 103
!
interface GigabitEthernet 0/20
 switch-port access vlan 101
!
interface GigabitEthernet 0/22
 switch-port access vlan 102
!
interface VLAN 101
 ipv6 address 1010::1/64
 ipv6 ospf 4949 set-area 0
!
interface VLAN 102
 ipv6 address 1020::1/64
 ipv6 ospf 4949 set-area 0
!
```

```
interface VLAN 103
 ipv6 address 1030::1/64
 ipv6 ospf 4949 set-area 0
!
interface VLAN 104
 ipv6 address 1011::2/64
 ipv6 ospf 4949 set-area 0
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 mld snooping-mode svgl shared-vlan 104
ipv6 mld snooping-mode svgl group-profile 1
ipv6 mld snooping-mode svgl
!
```

7. Common Errors

- The multicast groups associated with the SVGL mode are not configured.
- The address of the sent multicast traffic is not within the range of multicast groups associated with the SVGL mode.

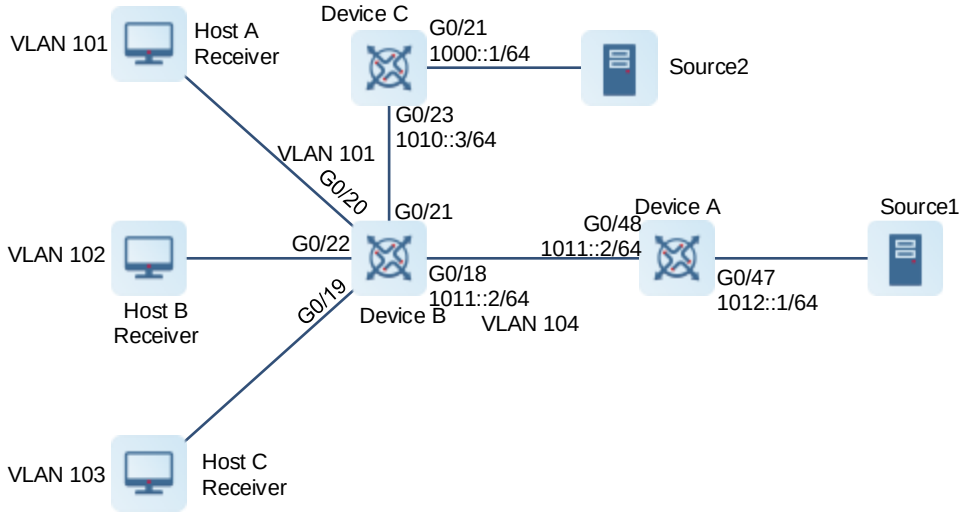
1.8.3 Configuring Basic MLD Snooping Functions in IVGL-SVGL Mode

1. Requirements

In the multicast network shown in Figure 1-7, the routers (Device A and Device C) connect to the user network through the switch (Device B), MLD snooping is run on Device A and Device C, and Device A and Device C directly connect to the multicast sources. Multicast source 1 sends data to multicast group FF14::10, and multicast source 2 sends data to multicast group FF15::11. Host A, Host B, and Host C are connected to VLAN 101, VLAN 102, and VLAN 103, respectively. Host B and Host C join multicast group FF14::10, and Host A joins multicast groups FF14::10 and FF15::11. Host B and Host C can receive multicast data that source 1 sends to multicast group FF14::10, and Host A can receive multicast data that source 1 sends to multicast group FF14::10 and multicast data that source 2 sends to multicast group FF15::11. Host A, Host B, and Host C enjoy shared multicast services, and Host A also enjoy independent multicast services.

2. Topology

Figure 1-7 Topology of Basic MLD Snooping Functions in IVGL-SVGL Mode



3. Notes

Configure basic MLD snooping functions in IVGL-SVGL mode on Device B.

- Configure the IPv6 addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.
- Enable multicast routing and PIM-SMv6 related functions on Device A and Device C.
- Enable MLD snooping and run the IVGL-SVGL mode on Device B.
- Configure the multicast groups associated with the MLD snooping SVGL mode on Device B and specify the shared VLAN and sub VLANs.

4. Procedure

(1) Configure the IPv6 addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.

Configure the IPv6 address and unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 address 1012::1/64
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# vlan 104
DeviceA(cfg)# interface vlan 104
DeviceA(cfg-if-VLAN 104)# ipv6 address 1011::2/64
DeviceA(cfg-if-VLAN 104)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-VLAN 104)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
```

```
DeviceA(cfg-if-GigabitEthernet 0/48)# switch-port mode trunk
DeviceA(cfg-if-GigabitEthernet 0/48)# switch-port trunk native vlan 104
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# ipv6 router-protocol ospf 4949
DeviceA(cfg-router)# exit
```

Configure the IPv6 address, VLAN, and unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan range 101-104
DeviceB(cfg-vlan-range)# exit
DeviceB(cfg)# interface VLAN 101
DeviceB(cfg-if-VLAN 101)# ipv6 address 1010::1/64
DeviceB(cfg-if-VLAN 101)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 101)# exit
DeviceB(cfg)# interface VLAN 102
DeviceB(cfg-if-VLAN 102)# ipv6 address 1020::1/64
DeviceB(cfg-if-VLAN 102)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 102)# exit
DeviceB(cfg)# interface VLAN 103
DeviceB(cfg-if-VLAN 103)# ipv6 address 1030::1/64
DeviceB(cfg-if-VLAN 103)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 103)# exit
DeviceB(cfg)# interface VLAN 104
DeviceB(cfg-if-VLAN 104)# ipv6 address 1011::2/64
DeviceB(cfg-if-VLAN 104)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 104)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port mode trunk
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port trunk native vlan 104
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 103
DeviceB(cfg-if-GigabitEthernet 0/19)# exit
DeviceB(cfg)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 101
DeviceB(cfg-if-GigabitEthernet 0/20)# exit
DeviceB(cfg)# interface GigabitEthernet 0/21
DeviceB(cfg-if-GigabitEthernet 0/21)# switch-port access vlan 101
DeviceB(cfg-if-GigabitEthernet 0/21)# exit
DeviceB(cfg)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/22)# switch-port access vlan 102
DeviceB(cfg-if-GigabitEthernet 0/22)# exit
DeviceB(cfg)# ipv6 router-protocol ospf 4949
DeviceB(cfg-router)# exit
```

Configure the IPv6 address and unicast routing protocol on Device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 address 1000::1/64
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface GigabitEthernet 0/23
DeviceC(cfg-if-GigabitEthernet 0/23)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/23)# ipv6 address 1010::3/64
DeviceC(cfg-if-GigabitEthernet 0/23)# ipv6 ospf 4949 set-area 0
DeviceC(cfg-if-GigabitEthernet 0/23)# exit
DeviceC(cfg)# ipv6 router-protocol ospf 4949
DeviceC(cfg-router)# exit
```

(2) Enable IPv6 multicast routing and PIM-SMv6 related functions on Device A and Device C.

Enable IPv6 multicast routing and PIM-SMv6 related functions on Device A.

```
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface vlan 104
DeviceA(cfg-if-VLAN 104)# ipv6 pim sparse-mode
DeviceA(cfg-if-VLAN 104)# exit
DeviceA(cfg)# ipv6 pim c-bsr GigabitEthernet 0/47
DeviceA(cfg)# ipv6 pim c-rp GigabitEthernet 0/47
```

Enable multicast routing and PIM-SMv6 related functions on Device C.

```
DeviceC(cfg)# ipv6 mrouting-enable
DeviceC(cfg)# interface GigabitEthernet 0/21
DeviceC(cfg-if-GigabitEthernet 0/21)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/21)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/21)# exit
DeviceC(cfg)# interface GigabitEthernet 0/23
DeviceC(cfg-if-GigabitEthernet 0/23)# ipv6 pim sparse-mode
DeviceC(cfg-if-GigabitEthernet 0/23)# exit
```

(3) Enable MLD snooping and run the IVGL-SVGL mode on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 mld snooping-mode ivgl-svgl
```

(4) Configure the multicast groups associated with the MLD snooping SVGL mode on Device B and specify the shared VLAN.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 mld group-profile 1
DeviceB(cfg-profile)# permit
```

```
DeviceB(cfg-profile)# range ff13::1 ff14::fff
DeviceB(cfg-profile)# exit
DeviceB(cfg)# ipv6 mld snooping-mode svgl group-profile 1
DeviceB(cfg)# ipv6 mld snooping-mode svgl shared-vlan 104
```

5. Verification

Run the **display ipv6 mld-info snooping gda-table** command on Device B to display the MLD snooping forwarding entries and check whether the member port list of entry (*, FF14::10, 104) contains GigabitEthernet 0/22, GigabitEthernet 0/20, and GigabitEthernet 0/19 and the member port list of entry (*, FF15::11, 101) contains GigabitEthernet 0/20.

```
DeviceB# display ipv6 mld-info snooping gda-table
Multicast Switching Cache Table
  D: DYNAMIC
  S: STATIC
  M: MROUTE
(*, FF15::11, 101):
  VLAN(101) 2 OPORTS:
    GigabitEthernet 0/20(D)
    GigabitEthernet 0/21(M)

(*, FF14::10, 104):
  VLAN(104) 1 OPORTS:
    GigabitEthernet 0/18(M)

  VLAN(103) 1 OPORTS:
    GigabitEthernet 0/19(D)

  VLAN(102) 1 OPORTS:
    GigabitEthernet 0/22(D)

  VLAN(101) 1 OPORTS:
    GigabitEthernet 0/20(D)
```

Run the **display ipv6 mld-info snooping** command on Device B and check whether the MLD snooping working mode is IVGL-SVGL.

```
DeviceB# display ipv6 mld-info snooping
MLD-snooping mode: IVGL-SVGL
SVGL vlan: 104
SVGL profile number: 1
Source port check: Disable
MLD Fast-Leave: Disable
MLD Report suppress: Disable
Query Max Response Time: 10 (Seconds)
Dynamic Mroute Aging Time: 300(Seconds)
Dynamic Host Aging Time: 260(Seconds)
```

```
MLD Snooping version: 1
MLD Tunnel: Disable

vlan 1
-------------
MLD Snooping state: Enabled
Multicast router learning mode: Enable
MLD Fast-Leave: Enabled
MLD VLAN Mode: STATIC

vlan 101
-------------
MLD Snooping state: Enabled
Multicast router learning mode: Enable
MLD Fast-Leave: Enabled
MLD VLAN Mode: STATIC

vlan 102
-------------
MLD Snooping state: Enabled
Multicast router learning mode: Enable
MLD Fast-Leave: Enabled
MLD VLAN Mode: STATIC

vlan 103
-------------
MLD Snooping state: Enabled
Multicast router learning mode: Enable
MLD Fast-Leave: Enabled
MLD VLAN Mode: STATIC

vlan 104
-------------
MLD Snooping state: Enabled
Multicast router learning mode: Enable
MLD Fast-Leave: Enabled
MLD VLAN Mode: STATIC
```

6. Configuration Files

```
Device A configuration file
hostname DeviceA
!
ipv6 mrouting-enable
!
vlan 104
!
```

```
interface GigabitEthernet 0/47
 unset switch-port
 ipv6 address 1012::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
interface GigabitEthernet 0/48
 switch-port mode trunk
 switch-port trunk native vlan 104
!
interface VLAN 104
ipv6 address 1011::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode!
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim c-bsr GigabitEthernet 0/47
ipv6 pim c-rp GigabitEthernet 0/47
!
```

- Device B configuration file

```
hostname DeviceB
!
ipv6 mld group-profile 1
 permit
 range FF13::1 FF14::FFF
!
vlan range 1,101-104
!
interface GigabitEthernet 0/18
 switch-port mode trunk
 switch-port trunk native vlan 104
!
interface GigabitEthernet 0/19
 switch-port access vlan 103
!
interface GigabitEthernet 0/20
 switch-port access vlan 101
!
interface GigabitEthernet 0/21
 switch-port access vlan 101
!
interface GigabitEthernet 0/22
 switch-port access vlan 102
!
interface VLAN 101
```

```
 ipv6 address 1010::1/64
 ipv6 ospf 4949 set-area 0
!
interface VLAN 102
 ipv6 address 1020::1/64
 ipv6 ospf 4949 set-area 0
!
interface VLAN 103
 ipv6 address 1030::1/64
 ipv6 ospf 4949 set-area 0
!
interface VLAN 104
 ipv6 address 1011::2/64
 ipv6 ospf 4949 set-area 0
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 mld snooping-mode svgl shared-vlan 104
ipv6 mld snooping-mode svgl group-profile 1
ipv6 mld snooping-mode ivgl-svgl
!
```

- Device C configuration file

```
hostname DeviceC
!
ipv6 mrouting-enable
!
interface GigabitEthernet 0/21
 unset switch-port
 ipv6 address 1000::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet 0/23
 unset switch-port
 ipv6 address 1010::3/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
ipv6 router-protocol ospf 4949
 graceful-restart
!
```

7. Common Errors

- The multicast groups associated with the SVGL mode are not configured.
- The address of sent multicast traffic is not within the range of multicast groups associated with the SVGL

mode.

- The multicast group addresses of multicast traffic in IVGL mode are within the multicast group range in SVGL mode. As a result, IVGL forwarding cannot be implemented.

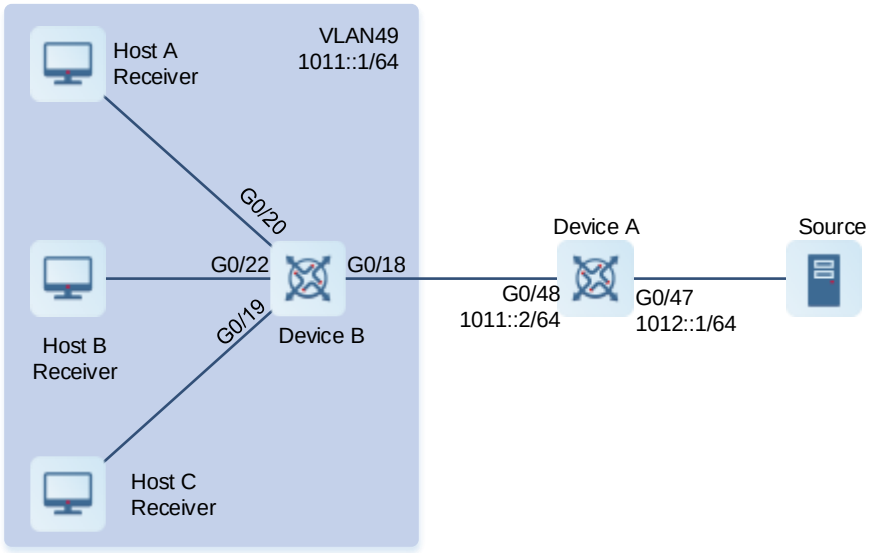
1.8.4 Configuring Static Ports to Implement L2 Multicast

1. Requirements

In the multicast network shown in Figure 1-8, Host A, Host B, and Host C need to receive data that a multicast source sends to a specific multicast group stably. The router (Device A) connects to the user network through the L2 device (Device B). The static MLD multicast groups FF14::10 to FF14::12 are configured on the user-side port on Device A. Host A and Host B want to receive data of multicast group FF14::10 stably, and Host C wants to receive data of multicast group FF14::11.

2. Topology

Figure 1-8 Topology of Static Ports to Implement L2 Multicast



3. Notes

Configure MLD snooping static multicast router ports and member ports on Device B.

- Configure the IPv6 addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.
- Enable multicast routing and MLD static multicast groups on Device A.
- Enable MLD snooping on Device B.
- Configure static multicast router ports and member ports on Device B.

4. Procedure

(1) Configure the IPv6 addresses, VLANs, and unicast routing protocols of the network nodes to ensure reachability of the network nodes using unicast routing.

Configure the IPv6 address and unicast routing protocol on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 address 1012::1/64
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/48)# ipv6 address 1011::2/64
DeviceA(cfg-if-GigabitEthernet 0/48)# ipv6 ospf 4949 set-area 0
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# ipv6 router-protocol ospf 4949
DeviceA(cfg-router)# exit
```

Configure the IPv6 address, VLAN, and unicast routing protocol on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# vlan 49
DeviceB(cfg-vlan)# exit
DeviceB(cfg)# interface vlan 49
DeviceB(cfg-if-VLAN 49)# ipv6 address 1011::1/64
DeviceB(cfg-if-VLAN 49)# ipv6 ospf 4949 set-area 0
DeviceB(cfg-if-VLAN 49)# exit
DeviceB(cfg)# interface GigabitEthernet 0/18
DeviceB(cfg-if-GigabitEthernet 0/18)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/18)# exit
DeviceB(cfg)# interface GigabitEthernet 0/19
DeviceB(cfg-if-GigabitEthernet 0/19)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/19)# exit
DeviceB(cfg)# interface GigabitEthernet 0/20
DeviceB(cfg-if-GigabitEthernet 0/20)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/20)# exit
DeviceB(cfg)# interface GigabitEthernet 0/22
DeviceB(cfg-if-GigabitEthernet 0/22)# switch-port access vlan 49
DeviceB(cfg-if-GigabitEthernet 0/22)# exit
DeviceB(cfg)# ipv6 router-protocol ospf 4949
DeviceB(cfg-router)# exit
```

(2) Enable IPv6 multicast routing and PIM-SMv6 related functions on Device A.

```
DeviceA(cfg)# ipv6 mrouting-enable
DeviceA(cfg)# interface GigabitEthernet 0/47
DeviceA(cfg-if-GigabitEthernet 0/47)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/47)# exit
DeviceA(cfg)# interface GigabitEthernet 0/48
DeviceA(cfg-if-GigabitEthernet 0/48)# ipv6 mld static-group-entries ff14::10
DeviceA(cfg-if-GigabitEthernet 0/48)# ipv6 mld static-group-entries ff14::11
```

```
DeviceA(cfg-if-GigabitEthernet 0/48)# ipv6 mld static-group-entries ff14::12
DeviceA(cfg-if-GigabitEthernet 0/48)# ipv6 pim sparse-mode
DeviceA(cfg-if-GigabitEthernet 0/48)# exit
DeviceA(cfg)# ipv6 pim c-bsr GigabitEthernet 0/47
DeviceA(cfg)# ipv6 pim c-rp GigabitEthernet 0/47
```

(3) Enable MLD snooping functions on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 mld snooping-mode ivgl
```

(4) Configure static multicast router ports and member ports on Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ipv6 mld snooping-mode vlan 49 static-member ff14::11 interface GigabitEthernet 0/19
DeviceB(cfg)# ipv6 mld snooping-mode vlan 49 static-member ff14::10 interface GigabitEthernet 0/22
DeviceB(cfg)# ipv6 mld snooping-mode vlan 49 static-member ff14::10 interface GigabitEthernet 0/20
DeviceB(cfg)# ipv6 mld snooping-mode vlan 49 multicast-router interface GigabitEthernet 0/18
```

5. Verification

Run the **display ipv6 mld-info snooping multicast-router** command and check whether the static router port is GigabitEthernet 0/18.

```
DeviceB# display ipv6 mld-info snooping multicast-router
Multicast Switching Mroute Port
  D: DYNAMIC
  S: STATIC
(*, *, 49):
  VLAN(49)   1 MROUTES:
    GigabitEthernet 0/18(DS)

(*, *, 101):
  VLAN(101)   1 MROUTES:
    GigabitEthernet 0/21(D)
```

Run the **display ipv6 mld-info snooping gda-table** command and check whether the static member ports of multicast group FF14::10 are GigabitEthernet 0/20 and GigabitEthernet 0/22 and the static member port of multicast group FF14::11 is GigabitEthernet 0/19.

```
DeviceB# display ipv6 mld-info snooping gda-table
Multicast Switching Cache Table
  D: DYNAMIC
  S: STATIC
  M: MROUTE
(*, FF14::10, 49):
  VLAN(49) 3 OPORTS:
    GigabitEthernet 0/18(M)
    GigabitEthernet 0/20(S)
```

```
      GigabitEthernet 0/22(S)

(*, FF14::11, 49):
   VLAN(49) 2 OPORTS:
      GigabitEthernet 0/18(M)
      GigabitEthernet 0/19(S)
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ipv6 mrouting-enable
!
interface GigabitEthernet0/47
 unset switch-port
 ipv6 address 1012::1/64
 ipv6 ospf 4949 set-area 0
 ipv6 pim sparse-mode
!
interface GigabitEthernet0/48
 unset switch-port
 ipv6 address 1011::2/64
 ipv6 ospf 4949 set-area 0
 ipv6 mld static-group-entries ff14::10
 ipv6 mld static-group-entries ff14::11
 ipv6 mld static-group-entries ff14::12
 ipv6 pim sparse-mode
!
Ipv6 router-protocol ospf 4949
 graceful-restart
!
ipv6 pim c-bsr GigabitEthernet 0/47
ipv6 pim c-rp GigabitEthernet 0/47
!
```

- Device B configuration file

```
hostname DeviceB
!
vlan range 1,49
!
interface GigabitEthernet 0/18
 switch-port access vlan 49
!
interface GigabitEthernet 0/19
 switch-port access vlan 49
```

```
!
interface GigabitEthernet 0/20
 switch-port access vlan 49
!
interface GigabitEthernet 0/22
 switch-port access vlan 49
!
interface VLAN 49
 ipv6 address 1011::1/64
 ipv6 ospf 4949 set-area 0
!
router-protocol ospf 49
 graceful-restart
!
ipv6 mld snooping-mode ivgl
!
ipv6 mld snooping-mode vlan 49 static-member FF14::11 interface GigabitEthernet 0/19
ipv6 mld snooping-mode vlan 49 static-member FF14::10 interface GigabitEthernet 0/22
ipv6 mld snooping-mode vlan 49 static-member FF14::10 interface GigabitEthernet 0/20
ipv6 mld snooping-mode vlan 49 multicast-router interface GigabitEthernet 0/18
!
```

7. Common Errors

- Basic MLD snooping functions are not configured or fail to be configured.

ACL and QoS Configuration

1. Configuring ACL
2. Configuring QoS
3. Configuring MMU

Contents

1 Configuring ACL 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 14
1.3 Configuring an IP Standard ACL 15
1.3.1 Overview 15
1.3.2 Restrictions and Guidelines 15
1.3.3 Configuration Tasks 15
1.3.4 Creating an IP Standard ACL 15
1.3.5 Applying an IP Standard ACL 17
1.4 Configuring an IP Extended ACL 18
1.4.1 Overview 18
1.4.2 Restrictions and Guidelines 18
1.4.3 Configuration Tasks 18
1.4.4 Creating an IP Extended ACL 18
1.4.5 Applying an IP Extended ACL 20
1.5 Configuring a MAC Extended ACL 22
1.5.1 Overview 22
1.5.2 Restrictions and Guidelines 22
1.5.3 Configuration Tasks 22
1.5.4 Creating a MAC Extended ACL 22
1.5.5 Applying a MAC Extended ACL 24

1.6 Configuring an Expert Extended ACL....25
1.6.1 Overview....25
1.6.2 Restrictions and Guidelines....25
1.6.3 Configuration Tasks....25
1.6.4 Creating an Expert Extended ACL....25
1.6.5 Applying an Expert Extended ACL....28
1.7 Configuring an IPv6 ACL....29
1.7.1 Overview....29
1.7.2 Restrictions and Guidelines....29
1.7.3 Configuration Tasks....29
1.7.4 Creating an IPv6 ACL....29
1.7.5 Applying an IPv6 ACL....31
1.8 Configuring an Expert Advanced ACL (ACL80)....33
1.8.1 Overview....33
1.8.2 Restrictions and Guidelines....33
1.8.3 Configuration Tasks....33
1.8.4 Creating an ACL80....33
1.8.5 Applying an ACL80....34
1.9 Configuring a Redirect ACL....35
1.9.1 Overview....35
1.9.2 Restrictions and Guidelines....35
1.9.3 Prerequisite....35
1.9.4 Procedure....35
1.10 Configuring a Global Security ACL....36

1.10.1 Overview 36
1.10.2 Prerequisite 36
1.10.3 Procedure 36
1.11 Configuring an SVI Router ACL 36
1.11.1 Overview 36
1.11.2 Prerequisite 36
1.11.3 Procedure 36
1.12 Synchronizing All ACEs from Software to Hardware 37
1.12.1 Overview 37
1.12.2 Procedure 37
1.13 Monitoring 37
1.14 Examples 39
1.14.1 Configuring an IP Standard ACL 39
1.14.2 Configuring an IP Extended ACL 40
1.14.3 Configuring a MAC Extended ACL 45
1.14.4 Configuring an Expert Extended ACL 47
1.14.5 Configuring an IPv6 ACL 50
1.14.6 Configuring an ACL80 51
1.14.7 Configuring ACL Rules Based on Time Range 54
1.14.8 Configuring an SVI Router ACL 56
1.14.9 Configuring the ACL Packet Counting Function 61

1 Configuring ACL

1.1 Introduction

1.1.1 Overview

The access control list (ACL) is also called access list or packet filtering in some documents. An ACL defines a series of "permit" or "deny" rule statements, and applies these rules to a device interface to control the data packets entering/leaving the interface, so as to improve the security of network devices.

Configuring ACL can guarantee network security, reliability, and stability, for example:

- Preventing packet attacks: An ACL can be configured to deny the attacks launched using the Internet Protocol (IP), Transmission Control Protocol (TCP), or Internet Control Message Protocol (ICMP) packets.
- Controlling network access: An ACL can be configured to restrict users to some services, for example, users are allowed to access only the WWW and email services, but not allowed to access other services such as telnet; some users are allowed to access services only within a given time period, or only certain hosts are allowed to access the network.
- Controlling network traffic: An ACL can be used together with quality of service (QoS) to guarantee prioritized services for some important data flows. For QoS configuration, see *Configuring QoS*.

1.1.2 Principles

1. Basic Concepts

- ACL

ACLs are classified into basic ACLs and dynamic ACLs.

Users can select a basic ACL or a dynamic ACL as needed. In general, a basic ACL can meet security requirements. However, attackers may counterfeit source addresses by using some software to spoof devices and access the network. Before users access the network, the dynamic ACL requires them to pass identity authentication, which makes it hard for attackers to access the network. Therefore, the dynamic ACL can be used in some sensitive areas to ensure network security.

Note

Spoofing devices by counterfeiting source addresses, namely, spoofing, is an inherent problem of all ACLs. Spoofing also occur on the dynamic ACL: Attackers may fake users' addresses to access the network during the effective access period after the users pass the identity authentication. This problem can be solved in two ways. One is to set shorter idle time of user access. The other is to use the IP Security (IPSec) encryption protocol to encrypt the network data, and ensure that all the data is encrypted when the data is transmitted to a device.

The ACL is generally configured on the following network devices:

- A device between an internal network and an external network (such as the Internet)

A device at the boundary of two networks

- A device connecting to a control port

- ACE

  An access control entry (ACE) is a statement containing the "permit" or "deny" action and filtering rules. Each ACE has a sequence number, which is automatically allocated by the device or manually configured. An ACL contains one or more ACEs. The ACL identifies and filters packets by using ACEs.

  The sequence of ACEs in an ACL determines the matching priority of the ACEs in the ACL. When processing packets, a network device performs rule matching in the ascending order of ACE sequence numbers. When finding a matched ACE, the network device stops matching subsequent ACEs.

  For example, create ACE 10 to deny all data flows.

```
10 deny ip any any
20 permit tcp 192.168.12.0 0.0.0.255 eq telnet any
```

  Since ACE 10 denies all the IP packets, the telnet packet from the host in the 192.168.12.0/24 network is denied even if it matches ACE 20. The reason is that, after finding that the packet matches the first ACE 10, the device stops matching the packet with the subsequent ACE 20.

  Here is another example: Create ACE 10 to permit all IPv6 data flows to pass through.

```
10 permit ipv6 any any
20 deny ipv6 host 200::1 any
```

  Since ACE 10 permits all IPv6 packets to pass through, the IPv6 packet sent from the host with IP address 200::1 can pass through even if it matches ACE 20. The reason is that, after finding that the packet matches the first ACE, the device stops matching the packet with the subsequent ACE 20.

- Step

  When the device automatically allocates sequence numbers to ACEs, the difference between two adjacent ACE sequence numbers is called a step. For example, if the step is set to 5, the device automatically allocates sequence numbers to ACEs in the ascending order of 5, 10, 15.... See the example below.

```
5 deny ip any any
10 permit tcp 192.168.12.0 0.0.0.255 eq telnet any
```

  After the step is changed, the ACE sequence numbers will be automatically rearranged based on the new step value. For example, when the step is changed to 10, the original ACE sequence numbers are changed from 5, 10, and 15 to 5, 15, and 25.

  A new ACE can be inserted between two ACEs by changing the step. For example, four ACEs are created, and the ACEs are manually numbered 1, 2, 3, and 4 in turn. To insert a new ACE after ACE 1, first change the step to 2 (then, the sequence numbers of the original four ACEs automatically change to 1, 3, 5, and 7), and then insert a manually configured ACE with sequence number 2.

- Filter field template

  Filter fields mean the fields used to identify and classify packets when an ACE is generated. A filter field template is a combination of such fields. An ACE identifies Ethernet packets by some fields in them, including:

  Layer 2 fields:

  - 48-bit source MAC address (the 48-bit source MAC address must be matched completely)
  - 48-bit destination MAC address (the 48-bit destination MAC address must be matched completely)

- 16-bit Layer 2 type field

Layer 3 fields:

- Source IP address field (all the source IP address values can be matched or a subnet can be used to define a type of flow)
- Destination IP address field (all the destination IP address values can be matched or a subnet can be used to define a type of flow)
- Protocol type field

Layer 4 fields:

- You can define a TCP source port, destination port, or both for matching; you can also define the range of the source port or destination port for matching.
- You can define a User Datagram Protocol (UDP) source port, destination port, or both for matching; you can also define the range of the source port or destination port for matching.

For example, you can create an ACE to identify and classify packets based on their destination IP address field. You can also create another ACE to identify and classify packets based on their source IP address field and UDP source port field. These two ACEs use different filter field templates.

- Rules

Rules refer to the values of fields in an ACE filter field template. For example, an ACE reads as follows:

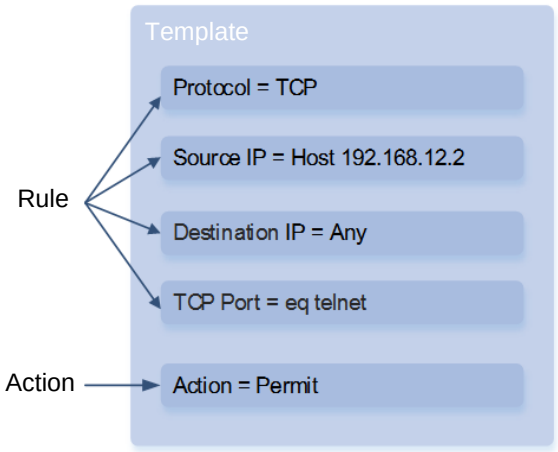
```
10 permit tcp host 192.168.12.2 any eq telnet
```

In this ACE, the filter field template is a collection of the following fields: source IP address field, destination IP address field, IP protocol field, and TCP destination port field. The corresponding values (namely, rules) are as follows: The source IP address is **Host 192.168.12.2**, the destination IP address is **Any** (namely, all the hosts), the IP protocol is **TCP**, and the TCP destination port is **Telnet**. See Figure 1-1.

- Action

Action is specified in ACE. The actions include permit and deny. Permit means that the traffic matching the rules can pass, and deny means that the traffic matching the rules is rejected.

Figure 1-1 Analysis of ACE: permit tcp host 192.168.12.2 any eq telnet



Note

- A filter field template can be a collection of Layer 3 fields and Layer 4 fields, or a collection of multiple Layer 2 fields. However, the filter field templates of the standard and extended ACLs cannot be a collection of Layer 2 and Layer 3 fields, or Layer 2 and Layer 4 fields, or Layer 2, Layer 3, and Layer 4 fields. To use a collection of Layer 2, Layer 3, and Layer 4 fields, you can use an expert extended ACL.
- Outbound ACLs to be associated with switch virtual interfaces (SVIs) can be IP standard, IP extended, MAC extended, or expert ACLs.
- If a destination MAC address is configured for matching in MAC extended and expert ACLs and such ACLs are applied in the outbound direction of an SVI, the ACEs containing the destination MAC address can be configured but fail to take effect. If you configure a destination IP address for matching in the IP extended and expert ACLs but the destination IP address is not within the subnet IP address range of the associated SVI, the configured ACLs do not take effect. For example, if VLAN 1 has an IP address 192.168.64.1 255.255.255.0 and an IP extended ACL is created, with the ACE **deny udp any 192.168.65.1 0.0.0.255 eq 255**, this ACL does not take effect when it is applied to the egress of VLAN 1. The cause is that the destination IP address is not within the subnet IP address range of VLAN 1. If the ACE is **deny udp any 192.168.64.1 0.0.0.255 eq 255**, the ACL takes effect because the destination IP address complies with the requirements.
- ACL resources (ternary content addressable memory (TCAM), keys, port groups, and ranges) are dynamically allocated resources. When services are delivered, the policy resource integration module allocates ACL resources based on the current status. ACL resources are first allocated to services accessed first. If ACL resources are insufficient for later accessed services, an ACL resource allocation failure occurs and a syslog error message is displayed. When data synchronization is triggered by device restart, hot swapping, or process restart, services may not be synchronized in their original time sequence. As a result, due to different time sequences of the services, no ACL resources are not allocated to services that should have ACL resources allocated. In this case, a syslog error message is displayed indicating insufficient ACL resources. You can delete related configuration to avoid data inconsistency.

Specification

- The outbound ACL acting on the physical port and Layer 3 aggregation port (AP) only supports matching well-known packets (unicast and multicast), but does not support matching unknown unicast packets, that is, the outbound ACL configured on a port does not take effect for the unknown packets or broadcast packets.
- When the input ACL, IEEE 802.1x, global IP and MAC address binding, port security, and IP source guard are all configured, the permit and default deny ACEs do not take effect but other deny ACEs take effect normally.
- When the input ACL and QoS are configured, the permit ACE does not take effect, but other deny ACEs take effect normally; the default deny ACE takes effect after the QoS entry.
- If an ACE is added to an ACL after the ACL is applied to the inbound direction of multiple SVIs, and the device is restarted after the configuration is saved, the ACLs configured on several SVIs may fail to take effect due to the hardware capacity causes.

Note

- When an expert ACL is configured and applied to the outbound direction of a port, if some ACEs in the ACL contain Layer 3 matching information (such as IP address and Layer 4 port), the non-IP packets received by the port cannot be subjected to the permit and deny ACEs of the ACL.
- When ACLs are applied, if an ACE in the ACLs (including the IP ACL and expert extended ACL) contains non-Layer 2 fields for matching (such as source IP (SIP) and destination IP (DIP)), the ACLs are ineffective for the tagged multiprotocol label switching (MPLS) packets.

2. IP ACL

IP ACL implements refined control over the IPv4 packets entering/leaving the device. You can deny or permit specific IPv4 packets destined to a network according to the actual needs, so as to control the access of IP users to network resources.

You can define a series of access rules in an IP ACL, and apply the ACL to the inbound or outbound direction of a port. You can also apply an IP ACL globally. When IPv4 packets enter or leave a device, the device determines whether to forward or block them by judging whether they match the rules.

To configure an IP ACL on a device, you must specify a unique name or number for each ACL to uniquely identify the ACL.

IP ACLs are classified into IP standard ACLs and IP extended ACLs. Table 1-1 lists the ranges of numbers that can be used for IP standard ACLs and IP extended ACLs.

Table 1-1 Number Ranges for IP Standard ACLs and IP Extended ACLs

Type	Number Range	Matching Field
IP standard ACL	1 to 99, 1300 to 1999	Source IP address
IP extended ACL	100 to 199, 2000 to 2699	● Source IP address ● Destination IP address ● IP protocol number ● Layer 4 source port ID or ICMP type ● Layer 4 destination port ID or ICMP code

The IP standard ACL controls the forwarding or blocking of packets based on the source IP address. The IP extended ACL controls the forwarding or blocking of packets by combining the matching fields in the table above.

A single ACL can use multiple independent ACL statements to define multiple rules, and all the statements use the same number or name so that the statements are bound to the same ACL.

Note

The ICMP code matching field in ACL rules is invalid for the ICMP packets with ICMP type 3. If the code field for matching ICMP packets is configured in an ACL rule, when ACL matching is performed on the ICMP packets with type 3 on the routers, the matching result may be different from the expected result.

One rule statement of denying all data flows is hidden at the end of each IP ACL. If a packet fails to match any rules, it will be denied. For example:

```
access-control-list 1 permit host 192.168.4.12
```

This ACL only permits the packets with source host address 192.168.4.12 to pass through, and denies the packets of the other hosts. The reason is that this ACL contains a rule statement at its end as follows:

```
access-control-list 1 deny any
```

Here is another example:

```
access-control-list 1 deny host 192.168.4.12
```

If an ACL contains the above statement only, the packets of all the hosts are denied wen passing through the port.

Caution

When defining an ACL, you must consider the route update packets. The rule statement of denying all data flows at the end of the ACL may result in blocking of all the route update packets.

The reflexive ACL automatically generates a temporary ACL according to the Layer 3 and Layer 4 information of the traffic originated from the internal network. A temporary ACL is created according to the following principles: The IP protocol number remains unchanged, the source IP address and destination IP address are strictly interchanged, and the TCP/UDP source port and destination port are strictly interchanged Only when the Layer 3 and Layer 4 information of the returned traffic strictly matches that in the temporary ACL previously created based on the outbound traffic, will the device permit the traffic to enter the internal network.

3. MAC Extended ACL

The MAC extended ACL implements refined control over the packets entering/leaving the device based on the Layer 2 information of the packets. You can deny or permit specific Layer 2 packets destined to a network according to the actual needs, so as to protect network resources against attacks or control the access of users to network resources.

You can define a series of access rules in a MAC extended ACL, and apply the ACL to the inbound or outbound direction of a port. When packets enter or leave a device, the device determines whether to forward or block them by judging whether they match the rules.

To configure a MAC extended ACL on a device, you must specify a unique name or number for each ACL to uniquely identify the ACL. Table 1-2 lists the range of numbers for the MAC extended ACLs.

Table 1-2 Number Range for MAC Extended ACLs

Type	Number Range	Matching Field
MAC extended ACL	700 to 799	● Source MAC address ● Destination MAC address ● Ethernet protocol type

The MAC extended ACL controls the forwarding or blocking of packets based on the source or destination MAC address and the Ethernet type of packets.

A single MAC extended ACL can use multiple independent ACL statements to define multiple rules, and all the statements use the same number or name so that the statements are bound to the same ACL.

Note

If a MAC extended ACL rule does not specify IPv6 packets, that is, the Ethernet type field is not defined or the defined Ethernet type field value is not 0x86dd, then the MAC extended ACL is not used to match IPv6 packets. If you need to match IPv6 packets, define an IPv6 ACL.

One rule statement of denying all data flows is hidden at the end of each MAC extended ACL. If a packet fails to match any rules, it will be denied. For example:

```
access-control-list 700 permit host 00d0.f800.0001 any
```

This ACL permits only the packets sent from the host with MAC address 00d0.f800.0001 to pass through, and denies the packets from all the other hosts. The reason is that this ACL contains a rule statement at its end as follows:

```
access-control-list 700 deny any any
```

4. Expert Extended ACL

An expert extended ACL implements refined control over the packets entering/leaving a device based on the Layer 2 and Layer 3 information of the packets. The expert extended ACL can be regarded as a combination and enhancement of the IP ACL and MAC extended ACL. An expert extended ACL can contain IP ACL rules and MAC extended ACL rules, and specify VLAN IDs for packet matching.

You can define a series of access rules in an expert extended ACL, and apply the ACL to the inbound or outbound direction of a port. When packets enter or leave a device, the device determines whether to forward or block them by judging whether they match the access rules.

To configure an expert extended ACL on a device, you must specify a unique name or number for each ACL to uniquely identify the ACL. Table 1-3 lists the range of numbers that can be used for expert extended ACLs.

Table 1-3 Number Range for Expert Extended ACLs

Type	Number Range	Matching Field
Expert extended ACL	2700 to 2899	● Source IP address ● Destination IP address ● IP protocol number ● Layer 4 source port ID or ICMP type ● Layer 4 destination port ID or ICMP code ● Source MAC address ● Destination MAC address ● Ethernet protocol type ● VLAN ID

The expert extended ACL controls the forwarding or blocking of packets by combining the matching fields in the table above.

A single expert extended ACL can use multiple independent ACL statements to define multiple rules, and all the statements must use the same number or name so that the statements are bound to the same ACL.

Note

If an expert extended ACL rule does not specify IPv6 packets, that is, the Ethernet type field is not defined or the defined Ethernet type field is not 0x86dd, then the expert extended ACL does not match IPv6 packets. If you need to match IPv6 packets, define an IPv6 ACL.

Specification

- The virtual extensible LAN (VXLAN) field in an expert extended ACL used for a data center product is used to match the inner packets of VXLAN. Therefore, you can apply an expert extended ACL in VXLAN mode to match the inner IP field of VXLAN.
- When a device needs to match VXLAN packets, you can specify the destination port ID of the VXLAN protocol to confirm VXLAN packets. You can also specify whether matched VXLAN packets carry tags. The **ip-flag** and **dscp** fields of VXLAN packets can be matched against ACLs

Specification

The UDF option in an expert extended ACL used for a data center product is a user-defined offset field for matching, in which users need to specify the protocol layer, offset value, data, and mask to be matched.

Note

The MPLS option in an expert extended ACL used for a data center product can provide the matching rules for the IPv4 field and tag field contained in MPLS packets.

One rule statement of denying all data flows is hidden at the end of each expert extended ACL. If a packet fails to match any rules, it will be denied. For example:

```
access-control-list 2700 permit 0x0806 any any any any any
```

This ACL only permits the packets with Ethernet type 0x0806 (namely, ARP) to pass through, and denies the packets of the other types. The reason is that this ACL contains a rule statement at its end as follows:

```
access-control-list 2700 deny any any any any
```

5. IPv6 ACL

The IPv6 ACL implements refined control over the IPv6 packets entering/leaving a device. You can deny or permit specific IPv6 packets destined to a network according to the actual needs, so as to control the access of IPv6 users to network resources.

You can define a series of access rules in an IPv6 ACL, and apply the ACL to the inbound or outbound direction of a port. When IPv6 packets enter or leave a device, the device determines whether to forward or block them by judging whether they match the rules.

To configure an ACL on a device, you must specify a unique name for the ACL of a protocol.

One rule statement of denying all IPv6 data flows is hidden at the end of each IPv6 ACL. If a packet fails to match any rules, it is denied. For example:

```
ipv6 access-control-list ipv6_acl
 10 permit ipv6 host 200::1 any
```

This ACL only permits the IPv6 packets coming from source host 200::1 to pass through, and denies the IPv6 packets from all the other hosts. The reason is that this ACL contains a rule statement at its end as follows:

```
deny ipv6 any any
```

6. ACL80

ACL80 is the expert advanced ACL, also known as user-defined ACL. ACL80 supports matching the specified bytes in the first 80 bytes of a packet by bit.

There are three elements in ACL80 matching: matching field content, matching field mask, and the start position (that is, the offset) of matching. The bits of the matching field content map to those of the matching field mask. The matching field content specifies the field value to be matched, and the matching field mask specifies whether the corresponding bits need to be matched. When a bit needs to be matched, the corresponding bit in the matching field mask must be set to **1**. If a bit in a matching field mask is set to **0**, the corresponding bit in the matching field content will not be matched regardless of the bit value in the matching field content. For example:

```
10 permit 00d0f8123456 ffffffffffff 0
20 deny 00d0f8654321 ffffffffffff 6
```

In ACE 10, the matching field content is 00d0f8123456, the matching field mask is ffffffffffff, and the offset is 0. This rule indicates that, if the destination MAC address of a packet is 00d0f8123456, the packet is permitted to be forwarded.

In ACE 20, the matching field content is 00d0f8654321, the matching field mask is ffffffffffff, and the offset is 6. This rule indicates that, if the source MAC address of a packet is 00d0f8654321, the packet is blocked.

You need to deeply understand the Layer 2 data frame structure so as to use the user-defined ACL correctly. Figure 1-2 shows the first 64 bytes of an Layer 2 data frame. In the figure, each letter represents one hexadecimal number and every two letters represent one byte.

Figure 1-2 First 64 Bytes of an Layer 2 Data Frame

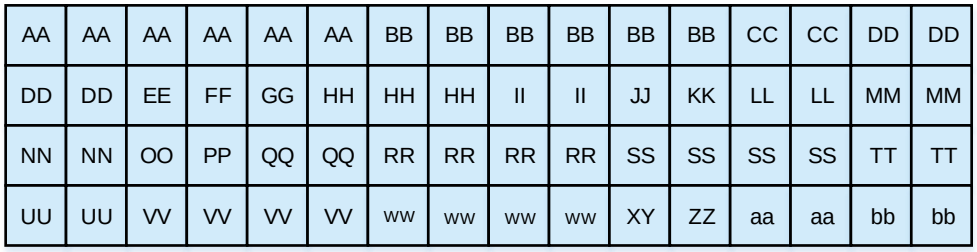


Table 1-4 lists the meaning and offset value of each letter.

Table 1-4 Meanings and Offset Values of Letters

Letter	Description	Offset	Letter	Description	Offset
A	Destination MAC address	0	O	TTL field	34
B	Source MAC address	6	P	Protocol number	35

Letter	Description	Offset	Letter	Description	Offset
C	VLAN tag field	12	Q	IP checksum	36
D	Data frame length field	16	R	Source IP address	38
E	Destination service access point (DSAP) field	18	S	Destination IP address	42
F	Source service access point (SSAP) field	19	T	TCP source port	46
G	Control field	20	U	TCP destination port	48
H	Org code field	21	V	Serial number	50
I	Encapsulated data type	24	W	Acknowledgment field	54
J	IP version No.	26	XY	IP header length and reserved bits	58
K	ToS field	27	Z	Reserved bits and flags bits	59
L	IP packet length	28	a	Windows size field	60
M	ID	30	b	Others	62
N	Flags field	32			

The offset of each field in the table is their offset in the IEEE 802.3 data frame containing the SNAP+Tag fields. The user-defined ACL extracts specific fields from the first 80 bytes of a data frame according to offsets listed in Table 1-5, and then compare the fields with the matching field masks so as to determine the matching field content. For example, if you permit all TCP packets to be forwarded, you can set the matching field content to **06**, the matching field mask to **ff**, and the offset to **35**. Create ACE 10 as follows:

```
10 permit 06 ff 35
```

Apply the ACL to the inbound or outbound direction of a port. When a packet enters or leaves a device, ACL80 extracts the content of the TCP protocol number field from the data frame, and then compares the content with the matching field content to find out all the matched TCP packets for forwarding.

7. Redirect ACL

Redirect ACL enables a device to redirect the packets matching the rules to a specified port for forwarding. Or the device obtains and analyzes the packets in specified ports.

Redirect ACL binds an ACL on an input port and specifies an output port. When the input interface receives a packet, it queries ACL rules bound to the port one by one. If the packet conforms to the characteristics described in a rule, it is forwarded from the output port specified by the policy.

Note

- Redirect ACL takes effect only in the inbound direction of a port.
- When a global redirect ACL is configured with a destination port. The port will not be affected by the global redirect ACL.

8. Global Security ACL

There are various virus packets in the network, and the identification characteristics of virus packets are the same or similar on each port. A port-based security ACL is often configured to filter out packets that comply with some characteristics (such as forged TCP attack packets) and prevent virus packets. You can create an ACL, add ACEs that match the characteristics of various virus packets, and apply the ACL to each port of a device to filter out virus packets.

When the port-based security ACL is used in the attack-defense scenarios such as virus filtering, there are much inconvenience.

- Ports must be configured one by one. Duplicate configuration, poor operating performance, and excessive consumption of ACL resources can occur.
- The access control function of the security ACL is weakened. Because the security ACL is used for virus filtering, the basic functions of the security ACL including restricting route updates and restricting network access become unavailable.

The global security ACL can be used for global anti-virus deployment and defense without affecting the port-based security ACL. A global security ACL can take effect on all the Layer 2 ports after one command is configured.

When both the global security ACL and port-based security ACL are configured, they take effect at the same time. The packets that match rules in the global security ACL will be directly filtered out as virus packets, and the packets that fail to match the rules of the global security ACL are still controlled by the port-based security ACL. To exempt some ports from the control of the global security ACL, disable the global security ACL function on these ports. When the global, port-based, and VLAN-based security ACLs are applied at the same time, the priority is port-based security ACL, VLAN-based security ACL, and global security ACL in the descending order.

To avoid misconfiguration of the global security ACL, the global security ACL disabling function is provided. If you disable the global security ACL function and then configure a global security ACL, the system prompts a configuration failure. If you have configured a global security ACL and then disable the global security ACL function, the system deletes all the current global security ACLs and displays a log prompt.

9. SVI Router ACL

An ACL applied to an SVI (namely, SVI ACL) takes effect for both the packets forwarded on Layer 2 in a VLAN and the inter-VLAN routed packets. As a result, different users in the same VLAN cannot make normal communication. The SVI router ACL function enables the ACL applied to an SVI to take effect only for the inter-VLAN routed packets.

The SVI router ACL function is disabled by default. The SVI ACL takes effect for the inter-VLAN Layer 3 forwarded packets and the intra-VLAN packets forwarded by bridges. After the SVI router ACL function is enabled, the SVI ACL is effective only for the inter-VLAN Layer 3 forwarded packets.

10. Packet Matching Logging

The packet matching logging is used to monitor the running status of ACL rules and provide necessary information for routine network maintenance and network optimization.

To better grasp the running status of ACL in the device, when adding ACEs, you can decide whether to specify the packet matching logging option as needed. If this option is specified, the matching logs are output when

ACEs match packets. ACL prints log information based on ACE, that is, the device periodically prints the ACEs that match packets and the number of matched packets. See the example below:

```
*Sep  9 16:23:06: %ACL-6-MATCH: ACL 100 ACE 10 permit icmp any any, match 78 packets, match 1000 bytes.
```

To reasonably control the quantity and frequency of log output, ACL allows configuring the log output interval.

Caution

- An ACL rule with the log option uses more hardware resources. If all the configured rules have the log option, the hardware policy capacity of the device is halved.
- The default output interval of packet matching logs is 0 minutes, that is, no ACL matching log is output. After specifying the log option when configuring ACL rules, you also need to configure the output interval, otherwise no matching log is output.
- For a rule with the log option, if no packet matches the rule within the specified time interval, no packet matching log related to the rule is output; if a packet matches the rule within the specified time interval, the packet matching log related to the rule is output after the time interval expires. The number of matched packets is the total number of packets matching the rule within the time interval, that is, the number of matched packets from the last log output to the current log output.

Specification

- The log option can be configured for the IP ACL and IPv6 ACL rules only.

11. Packet Matching Counting

For the needs of network management, you may want to know whether an ACL rule matches packets and the number of matched packets. ACL provides the rule-based packet matching counting function. You can enable or disable the packet matching counting function for all the rules in an ACL. When a packet matches a rule, the corresponding matching count increases accordingly. You can run the statistics clearing command for an ACL to clear the packet matching counts of all rules in the ACL and start new counting.

Caution

Enabling the packet matching counting function of ACL requires more hardware entries, and in extreme cases, halves the available hardware policy capacity of the device.

Specification

For IP ACL, MAC extended ACL, expert extended ACL, and IPv6 ACL, the packet matching counting function can be configured.

12. ACL Effective Time Range

If you need to control some flows within a specified time range, for example, disable chat tools during working hours, you can configure effective time ranges for ACEs to control the pass-through time of flows. The time ranges are classified into absolute time ranges and periodic time ranges.

An absolute time range indicates the time range from the specified start time to the end time. This time range does not occur cyclically nor has a period, for example, "12:00:00 on January 1, 2000 to 12:00:00 on January 1, 2001".

A periodic time range indicates a periodic time interval, for example, "8:00 on each Monday to 17:00 on each Friday".

For the configuration of time ranges, see *Configuring Time Range*.

13. Global Control Plane Security ACL

In some application scenarios, an ACL needs to be bound to restrict the source IP host to processing the initial packets in a TCP handshake, instead of conducting restriction after a TCP connection is established. The global control plane security ACL filters packets only through software. It not only reduces the hardware resource consumption but also meets the requirement for processing the initial TCP packets. If you apply a security ACL globally by running a control plane application command, the ACL takes effect on software only.

The global control plane security ACL takes effect on all Layer 2 ports. ACEs are not applied to hardware but only to software, which reduces the hardware resource consumption. During a TCP handshake, the software ACL checks initial TCP packets and filters out the TCP packets that match the ACL.

Note

- The global control plane security ACL is effective to software-based filtering and can be applied only to the inbound direction.
- The global control plane security ACL is not limited by the configuration of global ACL excluded ports. After the excluded ports are configured, the global control plane security ACL remains effective on the excluded ports.
- The global control plane security ACL takes effect on both Layer 2 ports and Layer 3 ports. That is, it can take effect on the following types of ports: access port, trunk port, hybrid port, Layer 3 Ethernet port, Layer 2 or Layer 3 aggregation port. It does not take effect on the SVIs or member ports of aggregation ports.
- The global control plane security ACL can be associated only with the IP standard ACL.

14. ACL hit-log

To allow users to know the content of packets matching ACEs, you can enable the device to output hit-logs. The hit-logs are output based on ACEs and packets, that is, the device outputs the information and number of hit-logs periodically.

```
Feb 23 22:46:53 s6510-2: %ACLD-6-HITLOG: acl(ext_acl) ace_sn(220) Permit igmp 112.112.0.21
(HundredGigabitEthernet 0/7)->100.11.11.134, (5) packet
Feb 23 22:46:57 s6510-2: %ACLD-6-HITLOG: acl(mac_acl) ace_sn(10) Permit udp
1234:1234:1234:1234:1234:1234:1234:1404(0) (HundredGigabitEthernet 0/7)->4100::123(0), (1) packet
Feb 23 22:46:59 s6510-2: %ACLD-6-HITLOG: acl(mac_acl) ace_sn(10) Permit packet 0000.0001.0106
(HundredGigabitEthernet 0/7)->ecb9.70b2.ae1c, (480) packet
```

There are three types of source and destination addresses of hit-logs: IPv4 address, IPv6 address, and MAC address.

To control output number and frequency of hit-logs, you can specify the output interval.

Warning

ACL hit-log is implemented based on the functions on the control plane and forwarding plane. Therefore, ACL must be applied on both the control plane and forwarding plane to activate hit-log configuration.

Note

- The default output interval of hit-logs is 1s, that it, up to 10 hit-logs are output per second. To enable log output, the output interval must be specified. To disable hit-log, set the output interval to 0.
- If the ACE containing the hit-log option does not match any packets within the specified interval, no hit-log will be output. If there are matched packets, the hit-log will be output when the interval expires.
- Hit-log is supported only on the switch.

15. ACL miss-log

If the packets match ACEs frequently, there are numerous hit-logs, which floods the CLI interface and affects performance. Furthermore, logs may be lost. Therefore, not all hit-logs are output.

The number of hit-logs that are not output is collected and output in the form of the miss-log.

The miss-logs are output periodically. The following example is for your reference:

```
Feb 23 22:46:53 s6510-2: %ACLD-6-MISSLOG: ACL logging rate-limited or missed 9652 packets
```

To control output number and frequency of miss-logs, you can specify the output interval.

To activate miss-log output, enable miss-log, hit-log, and ACL counter.

Note

- The default output interval of miss-logs is 1 min, that it, up to 1 miss-log is output per minute. To enable log output, the output interval must be specified. To disable miss-log, set the output interval to 0.
- Miss-log is supported by only the switch.

1.2 Configuration Task Summary

ACL configuration tasks are as follows.

To configure an ACL, select at least one of the following tasks to configure:

- Configuring an IP Standard ACL
- Configuring an IP Extended ACL
- Configuring a MAC Extended ACL
- Configuring an Expert Extended ACL
- Configuring an IPv6 ACL
- Configuring an Expert Advanced ACL (ACL80)
- Synchronizing All ACEs from Software to Hardware

1.3 Configuring an IP Standard ACL

1.3.1 Overview

This section describes how to create and apply an IP standard ACL to control the IPv4 packets entering/leaving a port. You can deny or permit specific IPv4 packets destined to a network so as to control the access of IP users to network resources.

1.3.2 Restrictions and Guidelines

- If a device needs to control users' access to network resources by checking the source IP addresses of packets, you can configure an IP standard ACL.
- The IP standard ACL can be configured on access, convergence, or core devices, depending on user distribution. The IP standard ACL takes effect only on the configured device, and does not affect the other devices in the network.

1.3.3 Configuration Tasks

IP standard ACL configuration includes the following tasks:

(1) Creating an IP Standard ACL

(2) Applying an IP Standard ACL

1.3.4 Creating an IP Standard ACL

1. Overview

This section describes how to create an IP standard ACL and configure rules.

2. Restrictions and Guidelines

- An IP standard ACL can contain no rules. When no rules are configured in an ACL, one rule statement of denying all data flows is hidden in the ACL to prohibit all IPv4 packets from entering the device.
- To make some ACL rules take effect at specified time or expire at specified time, for example, make an ACL effective within some time ranges in a week, configure an ACL rule with the **time-range** option.
- When configuring an ACL rule with the **time-range** option, configure the time range option. For the configuration of time ranges, see *Configuring Time Range*.
- For the hit-log function, see ACL hit-log and
- ACL miss-log for details. The ACE with the **log** option consumes more hardware resources. If all ACEs contain the **log** option, the hardware-based policy capacity will be cut by half.
- The default output interval of packet matching logs is 0 minutes, that is, no ACL matching log is output. After specifying the **log** option when configuring ACL rules, you also need to configure an output interval; otherwise no matching log is output.
- For a rule with the **log** option, if no packet is matched within the specified time interval, no packet matching log related to the rule is output; if a packet is matched within the specified time interval, the packet matching log related to the rule is output after the time interval expires. The number of matched packets is the total number of packets matching the rule within the time interval, that is, the number of matched packets from the last log output to the current log output.
- During the actual network maintenance, if many ACLs or rules are configured, but no remark is configured

for these ACLs or rules, it is difficult to distinguish the purposes of these ACLs or rules. Configuring remarks for ACLs or rules can help you understand the purpose of the ACLs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an IP standard ACL and rules. Please configure only one task.

- Create a numerically indexed IP standard ACL and rules.

  access-control-list *acl-number* { **deny** | **permit** } { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } [**time-range** *time-range-name*] [**hit-log**] [**log**]

  No IP standard ACL or rule is configured by default.

- Create a numerically indexed or named IP standard ACL and rules. Run the following commands in turn to create an IP standard ACL and rules.

  ip access-control-list standard { *acl-name* | *acl-number* }

  No IP standard ACL is configured by default.

  [*sequence-number*] { **deny | permit** } { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } [**time-range** *time-range-name*] [**hit-log**] [**log**]

  There is an ACL rule of deny type in an IP standard ACL by default.

(4) (Optional) Configure an update interval for packet matching logs.

ip access-control-list log-output interval *time-value*

The default update interval is 0 minutes. That is, no ACL matching log is output.

(5) (Optional) Configure a remark for the ACL. Please configure only one task.

- Configure a remark for the numerically indexed IP standard ACL.

  access-control-list *acl-number* **list-comment** *text*

- Configure remarks for rules in the numerically indexed or named IP standard ACL.

  list-comment *text*

No remark is configured for an ACL by default.

(6) (Optional) Configure remarks for rules in the IP standard ACL. Please configure only one task.

- Configure remarks for rules in the numerically indexed IP standard ACL.

  access-control-list *acl-number* **comment** *text*

- Configure remarks for rules in the numerically indexed or named IP standard ACL.

  comment *text*

No remark is configured for ACL rules by default.

(7) (Optional) Enable the packet matching counting function of an IP standard ACL.

ip access-control-list counter { *acl-name* | *acl-number* }

The packet matching counting function of an IP standard ACL is disabled by default.

(8) (Optional) Configure a step for rule sequence numbers in the IP standard ACL.

ip access-control-list resequence { *acl-name* | *acl-number* } *start-value step-value*

Both the start value and step of an IP standard ACL rule sequence numbers are **10** by default.

1.3.5 Applying an IP Standard ACL

1. Overview

Apply an IP standard ACL in the global configuration mode, interface configuration mode, or SVI configuration mode to make the IP standard ACL take effect.

2. Restrictions and Guidelines

Only one IP ACL, one MAC extended ACL, or one expert extended ACL can be applied to the inbound or outbound direction of a device port. Besides, an IPv6 ACL can be applied.

- Based on the user distribution, on the access, convergence, or core devices, in specified interface configuration mode, VXLAN interface configuration mode, or VLAN interface configuration mode, you can apply an IP standard ACL.
- Configuring the **in** or **out** option indicates that the IP standard ACL takes effect for the packets entering the device or those forwarded by the device.
- Configure the **counter-only** option to count the packets with some characteristics. The counting function takes effect only for permit rules but not for deny rules in an ACL.
- After an ACL is used as a **counter-only** ACL, the counting function cannot be globally enabled for the ACL, and the ACL cannot be applied globally or on the interface as a common ACL. That is, an ACL with the specified *acl-number* or *acl-name* cannot be used as a **counter-only** ACL and common ACL at the same time
- If the **control-plane** option is configured, the ACL is effective for software only so as to save hardware resources.
- If the **forward-plane** option is configured, the ACL is effective for hardware only.
- If the **forward-control-plane** option is configured, the ACL is effective for both software and hardware.
- The **port-counter** option indicates port-based packet counting.
- To apply the ACL with the **port-counter** option on all interfaces, the entries beyond the capacity may fail to be installed. The entries within the limit take effect normally while the entries beyond the limit are installed randomly on some ports.
- The **port-counter** option cannot be applied to member ports automatically. To count packets on member ports, apply the ACL to the member ports manually.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Apply an IP standard ACL globally.

ip access-apply { *acl-name* | *acl-number* } { **in** | **out** } [**control-plane** | **counter-only** | **forward-control-plane** | **forward-plane**] [**port-counter**]

No IP standard ACL is applied globally by default.

(4) Enter the interface configuration mode.

- Enter the Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the SVI configuration mode.

  interface vlan interface-*number*

(5) Apply an IP standard ACL to an interface.

ip access-apply { *acl-name* | *acl-number* } { **in** | **out** } [**control-plane** | **counter-only** | **forward-control-plane** | **forward-plane** | **priority**]

No IP standard ACL is applied to an interface by default.

1.4 Configuring an IP Extended ACL

1.4.1 Overview

This section describes how to create and apply an IP extended ACL to control the IPv4 packets entering/leaving a port. You can deny or permit specific IPv4 packets destined to a network so as to control the access of IP users to network resources.

1.4.2 Restrictions and Guidelines

- If a device needs to control users' access to network resources by checking the source IP addresses, destination IP addresses, protocol numbers, TCP/UDP source or destination port IDs of packets, you can configure an IP extended ACL.
- The IP extended ACL can be configured on access, convergence, or core devices, depending on user distribution. The IP extended ACL takes effect only on a configured device, but does not affect the other devices in the network.

1.4.3 Configuration Tasks

IP extended ACL configuration includes the following tasks:

(1) Creating an IP Extended ACL

(2) Applying an IP Extended ACL

1.4.4 Creating an IP Extended ACL

1. Overview

This section describes how to create an IP extended ACL and configure its rules.

2. Restrictions and Guidelines

- An IP extended ACL can contain no rules. When no rules are configured in an ACL, one rule statement of

denying all data flows is hidden in the ACL to prohibit all IPv4 packets from entering the device.

- To make some ACL rules take effect at specified time or expire at specified time, for example, make an ACL effective within some time ranges in a week, you can configure an ACL with the **time-range** option. If the specified *time-range-name* does not exist, the ACL takes effec.
- When configuring an ACL rule with the **time-range** option, see *Configuring Time Range*.
- For the hit-log function, see ACL hit-log and
- ACL miss-log for details
- Configuring an ACL rule with the **log** option consumes more hardware resources. If all the configured rules have the **log** option, the hardware policy capacity of the device is halved.
- The default output interval of packet matching logs is 0 minutes, that is, no ACL matching log is output. After specifying the **log** option when configuring ACL rules, you also need to configure an output interval; otherwise no matching log is output.
- For a rule with the **log** option, if no packet is matched within the specified time interval, no packet matching log related to the rule is output; if a packet is matched within the specified time interval, the packet matching log related to the rule is output after the time interval expires. The number of matched packets is the total number of packets matching the rule within the time interval, that is, the number of matched packets from the last log output to the current log output.
- During the actual network maintenance, if many ACLs or rules are configured, but no remark is configured for these ACLs or rules, it is difficult to distinguish the purposes of these ACLs or rules. Configuring remarks for ACLs or rules can help you understand the purpose of the ACLs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an IP extended ACL and its rules. Please configure only one task.

- Create a numerically indexed IP extended ACL and its rules.

  access-control-list *acl-number* { **deny** | **permit** } *protocol* { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } { *destination-ipv4-address destination-ipv4-wildcard* | **host** *destination-ipv4-address* | **any** } [[**precedence** *precedence*] [**tos** *tos*] | [**dscp** *dscp*] [**ecn** *ecn*]] [**fragment**] [**time-range** *time-range-name*] [**hit-log**] [**log**]

  No IP extended ACL or rule is configured by default.

- Create a numerically indexed or named IP extended ACL and its rules. Run the following commands in turn to configure an IP extended ACL and its rules.

  ip access-control-list extended { *acl-name* | *acl-number* }

  No IP extended ACL is configured by default.

  [*sequence-number*] { **deny** | **permit** } *protocol* { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } { *destination-ipv4-address destination-ipv4-wildcard* | **host**

destination-ipv4-address | **any** } [[**precedence** *precedence*] [**tos** *tos*] | [**dscp** *dscp*] [**ecn** *ecn*]] [**fragment**] [**time-range** *time-range-name*] [**hit-log**] [**log**]

There is an ACL rule of deny type in an IP extended ACL by default.

(4) (Optional) Configure an update interval for packet matching logs.

ip access-control-list log-output interval *time-value*

The default update interval is 0 minutes. That is, no ACL matching log is output.

(5) (Optional) Configure a remark for the IP extended ACL. Please configure only one task.

- Configure a remark for a numerically indexed IP extended ACL.

  access-control-list *acl-number* **list-comment** *comment text*

- Configure a remark for a numerically indexed or named IP extended ACL.

  list-comment *text*

No remark is configured for an ACL by default.

(6) (Optional) Configure remarks for rules in the IP extended ACL. Please configure only one task.

- Configure remarks for rules in the numerically indexed IP extended ACL.

  access-control-list *acl-number* **comment** *text*

- Configure remarks for rules in the numerically indexed or named ACL.

  comment *text*

No remark is configured for ACL rules by default.

(7) (Optional) Enable the packet matching counting function of an IP extended ACL.

ip access-control-list counter { *acl-name* | *acl-number* }

The packet matching counting function of an IP extended ACL is disabled by default.

(8) (Optional) Configure the start value and the step for rule sequence numbers in the IP extended ACL.

ip access-control-list resequence { *acl-name* | *acl-number* } *start-value step-value*

Both the start value and step of IP extended ACL rule sequence numbers are **10** by default.

1.4.5 Applying an IP Extended ACL

1. Overview

Apply an IP extended ACL in the global configuration mode, interface configuration mode, or SVI configuration mode to make the IP extended ACL take effect.

2. Restrictions and Guidelines

Only one IP ACL, one MAC extended ACL, or one expert extended ACL can be applied to the inbound or outbound direction of a device port. Besides, an IPv6 ACL can be applied.

- Based on the user distribution, on the access, convergence, or core devices, in specified interface configuration mode, VXLAN interface configuration mode, or VLAN interface configuration mode, you can apply an IP extended ACL.
- Configuring the **in** or **out** option indicates that the ACL takes effect for the packets entering the device or those forwarded by the device.
- Configuring the **reflect** option enables the reflexive ACL.

- Configure the **counter-only** option to count the packets with some characteristics. The counting function takes effect only for permit rules but not for deny rules in an ACL.
- After an ACL is used as a **counter-only** ACL, the counting function cannot be globally enabled for the ACL, and the ACL cannot be applied globally or on ports as a common ACL. That is, ACLs with the same *acl-number* or *acl-name* cannot be used as a **counter-only** ACL and a common ACL at the same time.
- If the **control-plane** option is configured, the ACL is effective for software only so as to save hardware resources.
- If the **forward-plane** option is configured, the ACL is effective for hardware only.
- If the **forward-control-plane** option is configured, the ACL is effective for both software and hardware.
- The **port-counter** option indicates port-based packet counting.
- To apply the ACL with the **port-counter** option on all interfaces, the entries beyond the capacity may fail to be installed. The entries within the limit take effect normally while the entries beyond the limit are installed randomly on some ports.
- The **port-counter** option cannot be applied to member ports automatically. To count packets on member ports, apply the ACL to the member ports manually.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Apply an IP extended ACL globally.

ip access-apply { *acl-name* | *acl-number* } { **in** | **out** } [**control**-**plane** | **forward-control-plane** | **forward-plane**] [**port-counter**]

No IP extended ACL is applied globally by default.

(4) Enter the interface configuration mode.

- Enter the Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the aggregate interface configuration mode.
  interface aggregateport *interface-number*

- Enter the SVI configuration mode.

  interface vlan interface-*number*

(5) Apply an IP extended ACL to an interface.

ip access-apply { *acl-name* | *acl-number* } { **in** | **out** } [**control-plane** | **counter-only** | **forward-control-plane** | **forward-plane** | **priority**]

No IP extended ACL is applied to an interface by default.

1.5 Configuring a MAC Extended ACL

1.5.1 Overview

This section describes how to create and apply a MAC extended ACL to control the Layer 2 packets entering/leaving a port. You can deny or permit specific Layer 2 packets destined to a network so as to control the access of users to network resources based on the Layer 2 packet headers.

1.5.2 Restrictions and Guidelines

- If a device needs to control users' access to network resources based on the Layer 2 packet information (such as the MAC addresses of users' PCs), you can configure a MAC extended ACL.
- The MAC extended ACL can be configured on access, convergence, or core devices, depending on user distribution. The MAC extended ACL takes effect only on the configured device, but does not affect other devices in the network.

1.5.3 Configuration Tasks

MAC extended ACL configuration includes the following tasks:

(1) Creating a MAC Extended ACL

(2) Applying a MAC Extended ACL

1.5.4 Creating a MAC Extended ACL

1. Overview

This section describes how to create a MAC extended ACL and configure its rules.

2. Restrictions and Guidelines

- An MAC extended ACL can contain no rules. When no rules are configured in an ACL, one rule statement of denying all data flows is hidden in the ACL to prohibit all Ethernet Layer 2 packets from entering the device.
- To make some ACL rules take effect at specified time or expire at specified time, for example, make an ACL effective within some time ranges in a week, you can configure an ACL with the **time-range** option. If the specified *time-range-name* does not exist, the ACL takes effect.
- When configuring an ACL rule with the **time-range** option, configure the time range option. For the configuration of time ranges, see *Configuring Time Range*.
- For the hit-log function, see ACL hit-log and
- ACL miss-log for details.
- During the actual network maintenance, if many ACLs or rules are configured, but no remark is configured for these ACLs or rules, it is difficult to distinguish the purposes of these ACLs or rules. Configuring remarks for ACLs or rules can help you understand the purpose of the ACLs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a MAC extended ACL and its rules. Please configure only one task.

- Create a numerically indexed MAC extended ACL and its rules.

  access-control-list *acl-number* { **deny** | **permit** } { *source-mac-address source-mac-wildcard* | **host** *source-mac-address* | **any** } { *destination-mac-address destination-mac-wildcard* | **host** *destination-mac-address* | **any** } [*ethernet-type*] [**cos** [*cos-value*] [**inner** *cos-value*]] [**time-range** *time-range-name*] [**hit-log**]

  No MAC extended ACL or rule is configured by default.

- Create a numerically indexed or named MAC extended ACL and its rules. Run the following commands in turn to configure a MAC extended ACL and its rules.

  mac access-control-list extended { *acl-name* | *acl-number* }

  No MAC extended ACL is configured by default.

  Run the command to configure a MAC extended ACL and enter the MAC extended ACL configuration mode.

  [*sequence-number*] { **deny** | **permit** } { *source-mac-address source-mac-wildcard* | **host** *source-mac-address* | **any** } { *destination-mac-address destination-mac-wildcard* | **host** *destination-mac-address* | **any** } [*ethernet-type*] [**cos** [*cos-value*] [**inner** *cos-value*]] [**time-range** *time-range-name*] [**hit-log**]

  There is an ACL rule of deny type in a MAC extended ACL by default.

(4) (Optional) Configure a remark for the ACL. Please configure only one task.

- Configure a remark for the numerically indexed ACL.

  access-control-list *acl-number* **list-comment** *text*

- Configure remarks for rules in the numerically indexed or named ACL.

  list-comment *text*

No remark is configured for an ACL by default.

(5) (Optional) Configure remarks for ACL rules. Please configure only one task.

- Configure remarks for rules in the numerically indexed ACL.

  access-control-list *acl-number* **comment** *text*

- Configure remarks for rules in the numerically indexed or named ACL.

  comment *text*

No remark is configured for ACL rules by default.

(6) (Optional) Enable the packet matching counting function of a MAC extended ACL.

mac access-control-list counter { *acl-name* | *acl-number* }

The packet matching counting function of a MAC extended ACL is disabled by default.

(7) (Optional) Configure the start value and the step for rule sequence numbers in the MAC extended ACL.

mac access-control-list resequence { *acl-name* | *acl-number* } *start-value step-value*

Both the start value and step of MAC extended ACL rule sequence numbers are **10** by default.

1.5.5 Applying a MAC Extended ACL

1. Overview

Apply a MAC extended ACL in the global configuration mode, interface configuration mode, or SVI configuration mode to make the MAC ACL take effect.

2. Restrictions and Guidelines

Only one IP ACL, one MAC extended ACL, or one expert extended ACL can be applied to the inbound or outbound direction of a device port. Besides, an IPv6 ACL can be applied.

- Based on the user distribution, on the access, convergence, or core devices, in specified interface configuration mode, VXLAN interface configuration mode, or VLAN interface configuration mode, you can apply a MAC extended ACL.
- Configuring the **in** or **out** option indicates that the ACL takes effect for the packets entering the device or those forwarded by the device.
- onfigure the **counter-only** option to count the packets with some characteristics. The counting function takes effect only for permit rules but not for deny rules in an ACL.
- After an ACL is used as a **counter-only** ACL, the counting function cannot be globally enabled for the ACL, and the ACL cannot be applied globally or on ports as a common ACL. That is, an ACL with the specified *acl-number* or *acl-name* cannot be used as a **counter-only** ACL and common ACL at the same time .
- If the **control-plane** option is configured, the ACL is effective for software only so as to save hardware resources.
- If the **forward-plane** option is configured, the ACL is effective for hardware only.
- If the **forward-control-plane** option is configured, the ACL is effective for both software and hardware.
- The **port-counter** option indicates port-based packet counting.
- To apply the ACL with the **port-counter** option on all interfaces, the entries beyond the capacity may fail to be installed. The entries within the limit take effect normally while the entries beyond the limit are installed randomly on some ports.
- The **port-counter** option cannot be applied to member ports automatically. To count packets on member ports, apply the ACL to the member ports manually.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Apply a MAC extended ACL globally.

mac access-apply { *acl-name* | *acl-number* } { **in** | **out** } [**control-plane** | **forward-control-plane** | **forward-plane**] [**port-counter**]

No MAC extended ACL is applied globally by default.

(4) Enter the interface configuration mode.

- Enter the Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the aggregate interface configuration mode.
  interface aggregateport *interface-number*

- Enter the SVI configuration mode.

  interface vlan interface-*number*

(5) Apply a MAC extended ACL to an interface.

mac access-apply { *acl-name* | *acl-number* } { **in** | **out** } [**control-plane** | **counter-only**| **forward-control**-**plane** | **forward-plane**]

No MAC extended ACL is applied to an interface by default.

1.6 Configuring an Expert Extended ACL

1.6.1 Overview

This section describes how to create and apply an expert extended ACL to control the packets entering/leaving the port. You can deny or permit specific packets destined to a network.

1.6.2 Restrictions and Guidelines

- If a device needs to control users' access to network resources by using IP ACL rules, MAC extended ACL rules, and VLANs in a mixed manner, you can configure an expert extended ACL.
- You can configure an expert extended ACL on an access, convergence, or core device based on the distribution of users. The configuration takes effect only on a configured device, but does not affect other devices in the network.

1.6.3 Configuration Tasks

Expert extended ACL configuration includes the following tasks:

(1) Creating an Expert Extended ACL

(2) Applying an Expert Extended ACL

1.6.4 Creating an Expert Extended ACL

1. Overview

This section describes how to create an expert extended ACL and configure its rules.

2. Restrictions and Guidelines

- An expert extended ACL can contain no rules. When no rules are configured in an ACL, one rule statement of denying all data flows is hidden in the ACL to prohibit all packets from entering the device.
- To make some ACL rules take effect at specified time or expire at specified time, for example, make an ACL effective within some time ranges in a week, , you can configure an ACL with the **time-range** option. If the specified *time-range-name* does not exist, the ACL takes effect.
- When configuring an ACL rule with the **time-range** option, configure the time range option. For the configuration of time ranges, see *Configuring Time Range*.

- For the hit-log function, see ACL hit-log and
- ACL miss-log for details.
- MPLS packets support matching labels from Layer 1 to Layer 6, for example, **permit mpls exp** 0 1 2 3 4 5 6 7 is used to match the EXP values in label 1 to label 6 from left to right respectively (the label closest to the Layer 2 header is in Layer 1). Another example is **permit mpls exp any** 1, which means that only the EXP value of Layer 2 labels is matched.

- During the actual network maintenance, if many ACLs or rules are configured, but no remark is configured for these ACLs or rules, it is difficult to distinguish the purposes of these ACLs or rules. Configuring remarks for ACLs or rules can help you understand the purpose of the ACLs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an expert extended ACL and its rules. Please configure only one task.

- Create a numerically indexed expert extended ACL and its rules.

  access-control-list *acl-number* { **deny** | **permit** } [*protocol* | [*ethernet-type*] [**cos** [*cos-value*] [**inner** *cos-value*]]] [**VID** [*vlan-id*] [**inner** *vlan-id*]] { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } { *source-mac-address source-mac-wildcard* | **host** *source*-mac-*address* | **any** } { *destination-ipv4-address destination-ipv4-wildcard* | **host** *destination-ipv4-address* | **any** } { *destination-mac-address destination-mac-wildcard* | **host** *destination-mac-address* | **any** } [[**precedence** *precedence*] [**tos** *tos*] | [**dscp** *dscp*] [**ecn** *ecn*]] [**fragment**] [[**udf** *udf-id header pos value mask*]] [**time-range** *time-range-name*] [**hit-log**]

  No expert extended ACL or rule is configured by default.

- Create a numerically indexed or named expert extended ACL and its rules. Run the following commands in turn to configure an expert extended ACL and its rules.

  expert access-control-list extended { *acl-name* | *acl-number* }

  No expert extended ACL is configured by default.

  Run the command to configure an expert extended ACL and enter the expert extended ACL configuration mode.

  [*sequence-number*] { **deny** | **permit** } [*protocol* | [*ethernet-type*] [**cos** [*cos-value*] [**inner** *cos-value*]]] [**VID** [*vlan-id*] [**inner** *vlan-id*]] { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } { *source-mac-address source-mac-wildcard* | **host** *source*-mac-*address* | **any** } { *destination-ipv4-address destination-ipv4-wildcard* | **host** *destination-ipv4-address* | **any** } { *destination-mac-address destination-mac-wildcard* | **host** *destination-mac-address* | **any** } [[**precedence** *precedence*] [**tos** *tos*] | [**dscp** *dscp*] [**ecn** *ecn*]] [**fragment**] [[**udf** *udf-id header pos value mask*]] [**time-range** *time-range-name*] [**hit-log**]

  There is an ACL rule of deny type in an expert extended ACL by default.

- Create a numerically indexed expert extended ACL and MPLS packet matching rules.

access-control-list *acl-number* { **deny** | **permit** } **mpls {** *source-ipv4-address source-ipv4-wildcard* **|** **host** *source-ipv4-address* | **any** } { *destination-ipv4-address destination-ipv4-wildcard* | **host** *destination-ipv4-address* | **any** } [**hit-log**]

No expert extended ACL or is configured by default.

- Create a numerically indexed or named expert extended ACL. Run the following commands in turn to configure an expert extended ACL.

  expert access-control-list extended { *acl-number* | *acl-name* }

  No expert extended ACL is configured by default.

  No expert extended ACL is configured by default.

- Create a numerically indexed or named expert extended ACL and MPLS label matching rules. Run the following commands in turn to configure an expert extended ACL and MPLS label matching rules.

  expert access-control-list extended { *acl-name* | *acl-number* }

  No expert extended ACL is configured by default.

  Run the command to configure an expert extended ACL and enter the expert extended ACL configuration mode.

  [*sequence-number*] { **deny** | **permit** } { **vxlan | vxlan-ignore-dport** } *protocol* { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } [**eq** *port*] { *destination-ipv4-address destination-ipv4-wildcard* | **host** *destination-ipv4-address* | **any** } [**eq** *port*] [**tagged**] [**udp-dport** *dport*] [**ip-flag** *flag*] [**dscp** *dscp*] [**time-range** *time-range-name*] [**hit-log**]

  There is an ACL rule of deny type in an expert extended ACL by default.

(4) Create an expert extended ACL and VXLAN inner 5-tuple rules.

- Run the following commands in turn to configure an expert extended ACL and VXLAN inner 5-tuple rules.

  expert access-control-list extended { *acl-number* | *acl-name* }

  No expert extended ACL is configured by default.

  Run the command to configure an expert extended ACL and enter the expert extended ACL configuration mode.

  [*sn*] { **permit** | **deny** } [**vxlan | vxlan-ignore-dport**] *protocol* { *source-ipv4-address source-ip–wildcard* | **host** *source-ipv4-address* | **any** } { *destination-ipv4-address destination-ipv4-wildcard* | **host** *destination-ipv4-address* | **any** } [**tagged**] [**udp-dport** *dport*] [**time-range** *time-name*]] [**hit-log**]

(5) (Optional) Configure a remark for the ACL. Please configure only one task.

- Configure a remark for the numerically indexed ACL.

  access-control-list *acl-number* **list-comment** *comment*

- Configure remarks for rules in the numerically indexed or named ACL.

  list-comment *comment*

No remark is configured for an ACL by default.

(6) (Optional) Configure remarks for ACL rules. Please configure only one task.

- Configure remarks for rules in the numerically indexed ACL.

access-control-list *acl-number* **comment** *comment*

- Configure remarks for rules in the numerically indexed or named ACL.

  comment *comment*

No remark is configured for ACL rules by default.

(7) (Optional) Configure the start value and the step for rule sequence numbers in the expert extended ACL.

expert access-control-list resequence { *acl-name* | *acl-number* } *start-value step-value*

Both the start value and step of an expert ACL rule sequence numbers are **10** by default.

1.6.5 Applying an Expert Extended ACL

1. Overview

Apply an expert extended ACL in the global configuration mode, interface configuration mode, or SVI configuration mode to make the expert extended ACL take effect.

2. Restrictions and Guidelines

Only one IP ACL, one MAC extended ACL, or one expert extended ACL can be applied to the inbound or outbound direction of a device port. Besides, an IPv6 ACL can be applied.

- Configuring the **in** or **out** option indicates that the ACL takes effect for the packets entering the device or those forwarded by the device.
- Configuring the **reflect** option enables the reflexive ACL.
- Configure the **counter-only** option to count the packets with some characteristics. The counting function takes effect only for permit rules but not for deny rules in an ACL.
- After an ACL is used as a **counter-only** ACL, the counting function cannot be globally enabled for the ACL, and the ACL cannot be applied globally or on ports as a common ACL. That is, ACLs with the same *acl-number* or *acl-name* cannot be used as a **counter-only** ACL and a common ACL at the same time.
- If the **control-plane** option is configured, the ACL is effective for software only so as to save hardware resources.
- If the **forward-plane** option is configured, the ACL is effective for hardware only.
- If the **forward-control-plane** option is configured, the ACL is effective for both software and hardware.
- The **port-counter** option indicates port-based packet counting.
- To apply the ACL with the **port-counter** option on all interfaces, the entries beyond the capacity may fail to be installed. The entries within the limit take effect normally while the entries beyond the limit are installed randomly on some ports.
- The **port-counter** option cannot be applied to member ports automatically. To count packets on member ports, apply the ACL to the member ports manually.
- The rule for matching the IP field of MPLS packets in an expert extended ACL cannot be applied in the outbound direction, namely, it does not take effect after applied.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Apply an expert extended ACL globally.

expert access-apply { *acl-name* | *acl-number* } { **in** | **out** } [**control-plane** | **counter-only** | **forward-control-plane** | **forward-plane**] [**port-counter**]

No expert extended ACL is applied globally by default.

(4) Enter the interface configuration mode.

- Enter the Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the aggregate interface configuration mode.
  interface aggregateport *interface-number*

- Enter the SVI configuration mode.

  interface vlan interface-*number*

(5) Apply an expert extended ACL to an interface.

expert access-apply { *acl-name* | *acl-number* } { **in** | **out** } [**control-plane** | **counter-only**| **forward-control-plane** | **forward-plane** | **priority**]

No expert extended ACL is applied to an interface by default.

1.7 Configuring an IPv6 ACL

1.7.1 Overview

This section describes how to create and apply an IPv6 ACL to control the IPv6 packets entering/leaving the port. You can deny or permit specific IPv6 packets destined to a network so as to control the access of IPv6 users to network resources.

1.7.2 Restrictions and Guidelines

- To control IPv6 users' access to network resources, configure an IPv6 ACL.
- You can configure an IPv6 ACL on an access, convergence, or core device based on the distribution of users. The IPv6 ACL takes effect only on the local device, but does not affect other devices in the network.

1.7.3 Configuration Tasks

IPv6 ACL configuration includes the following tasks:

(1) Creating an IPv6 ACL

(2) Applying an IPv6 ACL

1.7.4 Creating an IPv6 ACL

1. Overview

This section describes how to create an IPv6 ACL and configure its rules.

2. Restrictions and Guidelines

- You cannot specify the name of an IPv6 ACL but its number.
- An IPv6 ACL can contain no rules. When no rules are configured in an ACL, one rule statement of denying all data flows is hidden in the IPv6 ACL to prohibit all IPv6 packets except ND packets from entering the device.
- To make some ACL rules take effect at specified time or expire at specified time, for example, make an ACL effective within some time ranges in a week, you can configure an ACL with the **time-range** option. If the specified *time-range-name* does not exist, the ACL takes effect.
- When configuring an ACL rule with the **time-range** option, configure the time range option. For the configuration of time ranges, see *Configuring Time Range*.
- For the hit-log function, see ACL hit-log and
- ACL miss-log for details.
- Configuring an ACL rule with the **log** option consumes more hardware resources. If all the configured rules have the **log** option, the hardware policy capacity of the device is halved.
- The default output interval of packet matching logs is 0 minutes, that is, no ACL matching log is output. After specifying the **log** option when configuring ACL rules, you also need to configure an output interval; otherwise no matching log is output.
- For a rule with the **log** option, if no packet is matched within the specified time interval, no packet matching log related to the rule is output; if a packet is matched within the specified time interval, the packet matching log related to the rule is output after the time interval expires. The number of matched packets is the total number of packets matching the rule within the time interval, that is, the number of matched packets from the last log output to the current log output.
- During the actual network maintenance, if many ACLs or rules are configured, but no remark is configured for these ACLs or rules, it is difficult to distinguish the purposes of these ACLs or rules. Configuring remarks for ACLs or rules can help you understand the purpose of the ACLs.
- Before configuring a step for ACL rule sequence numbers, you must first configure an ACL.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an IPv6 ACL and enter the IPv6 ACL configuration mode.

ipv6 access-control-list *acl-name*

No IPv6 ACL is configured by default.

(4) Configure IPv6 ACL rules. Please configure only one task.

- Configure IPv6 ACL rules of TCP or UDP.

  [*sn*] { **permit** | **deny** } [*protocol* { *source-ipv6-prefix* / *prefix-length* | *source-ipv6-address source-ipv6-mask* | **host** *source-ipv6-address* | **any** } { { *destination-ipv6-prefix* / *prefix-length* | *destination-ipv6-address destination-ipv6-mask* | **host** *destination-ipv6-address* | **any** } [*op dstport* | **range** *lower upper*]] [**cos** *value* [**inner** *value*]] [{ **any** | **host** *source-mac-address* |

source-mac-address source-mac-wildcard } { **any** | **host** *destination–mac-address* | *destination-mac-address destination-mac-wildcard* }] [**dscp** *dscp*] [**flow-label** *flow-label*] [**fragment**] [**vid** *vlan-id*] [**time-range** *time-name*] [**hit-log**] [**log**]

- Configure IPv6 ACL rules of protocols other than TCP or UDP.

  [*sn*] { **permit** | **deny** } [*protocol* { *source-ipv6-prefix* / *prefix-length* | *source-ipv6-address source-ipv6-mask* | **host** *source-ipv6-address* | **any** } { *destination-ipv6-prefix* / *prefix-length* | *destination-ipv6-address destination-ipv6-mask* | **host** *destination-ipv6-address* | **any** }] [{ **any** | **host** *source-mac-address* | *source-mac-address source-mac-wildcard* } { **any** | **host** *destination–mac-address* | *destination–mac-address destination-mac-wildcard* }] [**dscp** *dscp*] [**flow-label** *flow-label*] [**fragment**] [**vid** *vlan-id*] [**time-range** *time-name*] [**log**] [**hit-log**]

(5) (Optional) Configure a remark for the ACL.

list-comment *comment*

No remark is configured for an ACL by default.

(6) (Optional) Configure remarks for ACL rules.

comment *comment*

No remark is configured for ACL rules by default.

(7) (Optional) Configure an update interval of packet matching logs.

ipv6 access-control-list log-output interval *time*

The default update interval of packet matching logs is 0 minutes, that is, no ACL matching log is output.

(8) (Optional) Enable the packet matching counting function of an IP extended ACL.

ip access-control-list counter { *acl-name* | *acl-number* }

The packet matching counting function of an IP extended ACL is disabled by default.

(9) (Optional) Configure the start value and the step for rule sequence numbers in the IPv6 extended ACL.

ipv6 access-control-list resequence *acl-name start-value step-value*

Both the start value and step of an IPv6 ACL rule sequence numbers are **10** by default.

1.7.5 Applying an IPv6 ACL

1. Overview

Apply an IPv6 ACL in the global configuration mode, interface configuration mode, or SVI configuration mode to make the IPv6 ACL take effect.

2. Restrictions and Guidelines

- Based on the user distribution, on the access, convergence, or core devices, in specified interface configuration mode, VXLAN interface configuration mode, or VLAN interface configuration mode, you can apply an IPv6 ACL.
- Configuring the **in** or **out** option indicates that the ACL takes effect for the packets entering the device or those forwarded by the device.
- Configuring the **reflect** option enables the reflexive ACL.

- Configure the **counter-only** option to count the packets with some characteristics. The counting function takes effect only for permit rules but not for deny rules in an ACL.
- After an ACL is used as a **counter-only** ACL, the counting function cannot be globally enabled for the ACL, and the ACL cannot be applied globally or on ports as a common ACL. That is, an ACL with the specified *acl-number* or *acl-name* cannot be used as a **counter-only** ACL and common ACL at the same time.
- If the **control-plane** option is configured, the ACL is effective for software only so as to save hardware resources.
- If the **forward-plane** option is configured, the ACL is effective for hardware only.
- If the **forward-control-plane** option is configured, the ACL is effective for both software and hardware.
- The **port-counter** option indicates port-based packet counting.
- To apply the ACL with the **port-counter** option on all interfaces, the entries beyond the capacity may fail to be installed. The entries within the limit take effect normally while the entries beyond the limit are installed randomly on some ports.
- The **port-counter** option cannot be applied to member ports automatically. To count packets on member ports, apply the ACL to the member ports manually.
- Neither of the ACE used for matching the IP field of MPLS packets and the ACE used for matching the label field of MPLS packets in an IPv6 ACL can be applied in the outbound direction, namely, it does take effect after applied.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Apply an Ipv6 standard ACL globally.

ipv6 access-apply { *acl-name* | *acl-number* } { **in** | **out** } [**control-plane** | **counter-only** | **forward-control-plane** | **forward-plane**] [**port-counter**]

No Ipv6 standard ACL is applied globally by default.

(4) Enter the interface configuration mode.

- Enter the interface configuration mode.

  interface *interface-type interface-number*

- Enter the aggregate interface configuration mode.
  interface aggregateport *interface-number*

- Enter the VLAN interface configuration mode.

  interface vlan *vlan-id*

(5) Apply an IPv6 ACL to an interface.

ipv6 traffic-filter *acl-name* { **in** | **out** } [**counter-only** | **control-plane** | **forward-control-plane** | **forward-plane** | **priority**]

No IPv6 ACL is applied to an interface by default.

1.8 Configuring an Expert Advanced ACL (ACL80)

1.8.1 Overview

When the fixed matching fields of the IP standard ACL, IP extended ACL, MAC extended ACL, expert extended ACL, and IPv6 ACL all fail to meet requirements, you can configure an expert advanced ACL (ACL80) to define the packet fields to be matched.

1.8.2 Restrictions and Guidelines

- You can configure an ACL80 on an access, convergence, or core device based on the distribution of users. The ACL80 takes effect only on the configured device, but does not affect other devices in the network.

1.8.3 Configuration Tasks

The ACL80 configuration includes the following tasks:

(1) Creating an ACL80

(2) Applying an ACL80

1.8.4 Creating an ACL80

1. Overview

This section describes how to create an ACL80 and configure its rules.

2. Restrictions and Guidelines

- An ACL80 supports matching Ethernet II packets, IEEE 802.2 LLC SNAP packets, and IEEE 802.2 packets. If the values of the DSAP field to the Control field to be matched are AAAA03, IEEE 802.2 SNAP packets are matched. If the values of the DSAP field to the Control field to be matched are E0E003, IEEE 802.2 LLC packets are matched. If an ACL80 tries to match Ethernet II packets, you cannot set the values of the DSAP field to the Control field.
- Due to hardware, currently ACL80 cannot implement random byte matching for the first 80 bytes of packets, but only supports matching the fields of the destination MAC address, source MAC address, VLAN ID, ETYPE, IP protocol number, source IPv4 address, destination IPv4 address, source port, destination port, ICMP_TYPE, ICMP_CODE, and PPPOE_IPTYPE in packets.
- When an ACL80 matches IP, ARP, and other information, the encapsulated data type and data type mask need to be configured first, that is, the field with offset 24 needs to be configured first, and the mask needs to be set to all Fs. For example, to allow the packets with the source IP address 192.168.1.2 to pass through, run the **permit 0800 FFFF 24 C0A80102 FFFFFFFF 38** command.
- An ACL80 can contain no rules. When no rules are configured in an ACL, one rule statement of denying all data flows is hidden in the ACL to prohibit all packets from entering the device.
- During the actual network maintenance, if many ACLs or rules are configured, but no remark is configured for these ACLs or rules, it is difficult to distinguish the purposes of these ACLs or rules. Configuring remarks for ACLs or rules can help you understand the purpose of the ACLs.
- For the hit-log function, see ACL hit-log and
- ACL miss-log for details.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an expert advanced ACL and enter the expert advanced ACL configuration mode.

expert access-control-list advanced *acl-name*

No expert advanced ACL is configured by default.

(4) (Optional) Configure rules in the expert advanced ACL.

[*sequence-number*] { **deny** | **permit** } *hex hex-mask offset* [**hit-log**]

No expert advanced ACL is configured.

(5) (Optional) Configure a remark for the ACL.

list-comment *text*

No remark is configured for an ACL by default.

(6) (Optional) Configure remarks for ACL rules.

comment *text*

No remark is configured for ACL rules by default.

1.8.5 Applying an ACL80

1. Overview

Apply an ACL80 in the global configuration mode, interface configuration mode, SVI configuration mode, or VXLAN configuration mode to make the ACL80 take effect.

2. Restrictions and Guidelines

- Configuring the **in** or **out** option indicates that the ACL takes effect for the packets entering the device or those forwarded by the device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the aggregate interface configuration mode.
  interface aggregateport *interface-number*

- Enter the SVI configuration mode.

  interface vlan interface-*number*

- Enter the VXLAN configuration mode.

  vxlan *vni-number*

(4) Apply an expert advanced ACL to an interface.

expert access-apply { *acl-name* | *acl-number* } { **in** | **out** }

No expert advanced ACL is applied to an interface by default.

1.9 Configuring a Redirect ACL

1.9.1 Overview

This section describes how to configure the ACL-based redirection function on a specified port or globally, to redirect the matched packets entering the port or all the ports to a specified port for forwarding.

1.9.2 Restrictions and Guidelines

- You can configure the ACL-based redirection function on an access, aggregation, or core device based on the distribution of users. The configuration takes effect only on the local device, but does not affect other devices in the network.
- The ACL-based redirection function can be configured on the Layer 2 Ethernet interface and the Layer 2 aggregate interface only.
- The packet to be redirected must be an Layer 2 forwarded packet (that is, if a packet is forwarded from VLAN 2 to VLAN 3, it cannot be redirected), and the destination port of redirection must be in the same VLAN as the source port.
- If no rule is configured in an ACL, the ACL-based redirection function is unavailable.

1.9.3 Prerequisite

You must configure an ACL first to enable the ACL-based redirection function.

1.9.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

- Configure a redirect ACL.

  global redirect destination interface *interface-type interface-number* **acl** { *acl-name* | *acl-number* } **in**

  No redirect ACL is configured by default.

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

(4) Configure a redirect ACL. Please configure only one task.

redirect destination interface *interface-type interface-number* **acl** { *acl-name* | *acl-number* } **in**

No redirect ACL is configured on a port by default.

1.10 Configuring a Global Security ACL

1.10.1 Overview

Disable the global security ACL function to forbid configuring global security ACLs.

1.10.2 Prerequisite

You must configure an ACL first to enable the global security ACL function.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Disable the global security ACL function.

global access-apply disable

The global security ACL function is enabled by default.

1.11 Configuring an SVI Router ACL

1.11.1 Overview

The SVI router ACL function can be configured to enable the ACL applied to an SVI to take effect only for the inter-VLAN routed packets.

1.11.2 Prerequisite

To implement this function, first configure an ACL.

1.11.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an SVI router ACL.

svi router-acls enable

1.12 Synchronizing All ACEs from Software to Hardware

1.12.1 Overview

When the entry capacity of the device software is greater than that supported by the hardware, no ACE can be added. When the entry capacity of the device software drops below that supported by the hardware, the ACEs that fail to be added before will not be added again. This command can be used to refresh the configuration to trigger the re-adding of ACEs, thus synchronizing ACEs from software to hardware.

1.12.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Synchronize all ACEs from software to hardware.

acl ref synchronize all

1.13 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may interrupt services due to loss of important information.

Run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-5 ACL Monitoring

Command	Purpose
display access-control-lists [*acl-name* | *acl-number*] [**summary**] [**port-counter** [**interface** *interface-type interface-number*] [**in** | **out**]]	Displays a basic ACL.
display redirect [**interface {** *interface-type interface-number* | **vlan** *vlan-id* }]	Displays the redirection entries bound to a specified port. If no port is specified, the redirection entries bound to all the ports are displayed.

Command	Purpose
display access-apply [**interface {** *interface-type interface-number* | **vlan** *vlan-id* }]	Displays the configuration of an ACL applied to a port.
display ip access-apply [**interface** { *interface-type interface-number* | **vlan** *vlan-id* }]	Displays the configuration of an IP standard ACL and an extended ACL applied to a port.
display mac access-apply [**interface** *interface-type interface-number* | **vlan** *vlan-id* | **wlan** *wlan-id*]	Displays the configuration of a MAC extended ACL applied to a port.
display expert access-apply [**interface** { *interface-type interface-number* | **vlan** *vlan-id* }]	Displays the configuration of an expert extended ACL applied to a port.
display ipv6 traffic-filter [**interface {** *interface-type interface-number* }]	Displays the configuration of an IPv6 ACL applied to a port.
display acl res [**dev** *dev-number* [**slot** *slot-number*]]	Displays information about all ternary content-addressable memory (TCAMs) or a specified TCAM.
display acl capability	Displays device capability.
display svi router-acls state	Checks whether an ACL applied to an SVI takes effect on Layer 2 and Layer 3 networks.
display acl res detail [**dev** *dev-num* [**slot** *slot-num*]]	Displays usage details about all TCAMs or a specified TCAM.
cleanup acl res	Clears historical peak information of TCAM resource usage.
cleanup counters access-control-list [*acl-name* | *acl-number*]	Clears the number of packets matching an ACL.
cleanup access-control-list counters [*acl-name* | *acl-number*]	Clears the number of packets matching ACL deny rules.
debugging acl acld event	Debugs the ACL running process.
debugging acl acld client-show	Displays ACL client information.
debugging acl acld acl-show	Displays ACLs created on all ACL clients.

1.14 Examples

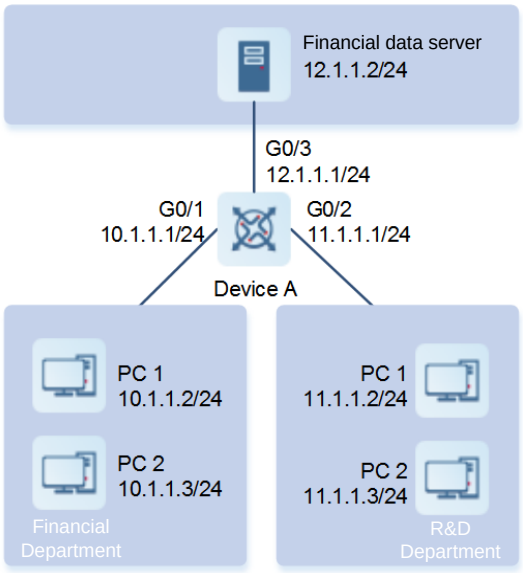
1.14.1 Configuring an IP Standard ACL

1. Requirements

An IP standard ACL needs to be configured to prevent the departments other than the financial department from accessing the financial data server.

2. Topology

Figure 1-3 Configuring a Standard IP ACL



3. Notes

- Configure an IP standard ACL on device A and add access rules.
- Apply the IP standard ACL to the outbound direction of the port connected to the financial data server on device A.

4. Procedure

(1) Configure an IP standard ACL and add access rules.

On device A, configure an IP standard ACL and add access rules.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip access-control-list standard 1
DeviceA(cfg-stdacl)# permit 10.1.1.0 0.0.0.255
DeviceA(cfg-stdacl)# deny 11.1.1.1 0.0.0.255
DeviceA(cfg-stdacl)# exit
```

(2) Apply the IP standard ACL to a port.

On device A, apply the IP standard ACL to the outbound direction of the port connected to the financial data server.

```
DeviceA(cfg)# interface gigabitethernet 0/3
DeviceA(cfg-if-GigabitEthernet 0/3)# ip access-apply 1 out
```

5. Verification

Check whether the ACL configuration commands are correct on device A.

```
DeviceA# display access-control-lists

ip access-control-list standard 1
10 permit 10.1.1.0 0.0.0.255
20 deny 11.1.1.0 0.0.0.255

DeviceA# display access-apply
ip access-apply 1 out
Applied On interface GigabitEthernet 0/3
```

Ping the financial data server from a PC of the R&D department and confirm that the ping operation fails.

Ping the financial data server from a PC of the financial department and confirm that the ping operation succeeds.

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
ip access-control-list standard 1
 10 permit 10.1.1.0 0.0.0.255
 20 deny 11.1.1.0 0.0.0.255
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 10.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 ip address 11.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/3
 unset switch-port
 ip access-apply 1 out
 ip address 12.1.1.1 255.255.255.0
!
```

1.14.2 Configuring an IP Extended ACL

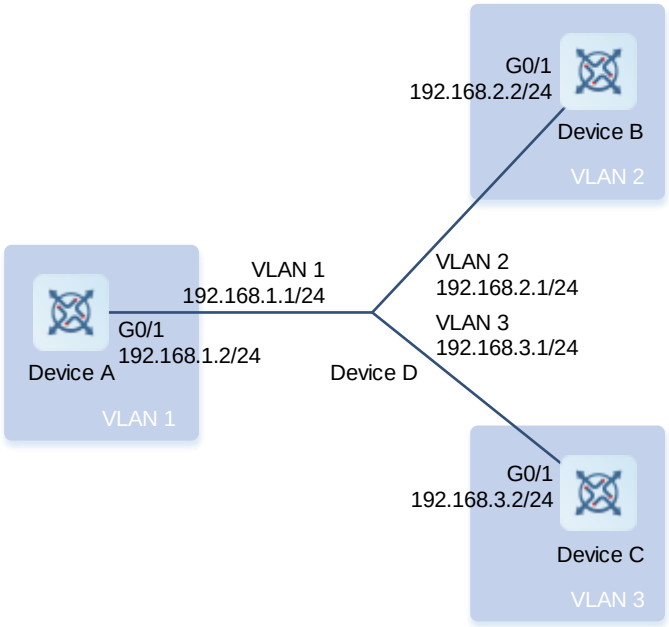
1. Requirements

Device A (VLAN 1), device B (VLAN 2), and device C (VLAN 3) are directly connected to device 4, which functions as the gateway of all the hosts. Requirement 1: The ping operation fails between VLAN 2 and VLAN 3,

but succeeds between VLAN 1 and VLAN 2, and between VLAN 1 and VLAN 3. Requirement 2: The DHCP packets of VLAN 1 and VLAN 2 are unreachable to each other, but the communication using other protocols is normal. Requirement 3: The host in VLAN 1 cannot to access VLAN 3 through telnet or SSH, but the communication using other protocols is normal.

2. Topology

Figure 1-4 Configuring an Extended IP ACL



3. Notes

- On device D, configure an IP extended ACL and add an access rule of filtering out packets received by UDP port 67 or 68 to meet requirement 2. On device C, configure an IP extended ACL and add an access rule of filtering out packets received by ACL ports 23 and 22 to meet requirement 3.
- On device A, apply the IP extended ACL to VLAN 1 interface, VLAN 2 interface, and VLAN 3 interface. On device C, apply the IP extended ACL to the line connected to the device D.

4. Procedure

(1) Configure IP addresses for all the device ports (omitted).

(2) Configure an IP extended ACL and add access rules.

On device D, configure an IP extended ACL and add access rules.

```
DeviceD> privilege
DeviceD# config-system terminal
DeviceD(cfg)# ip access-control-list extended inter_vlan_access1
DeviceD(cfg-extacl)# deny udp any eq bootps any eq bootpc
DeviceD(cfg-extacl)# deny udp any eq bootpc any eq bootps
DeviceD(cfg-extacl)# comment //Deny DHCP packets.
DeviceD(cfg-extacl)# permit ip any any
DeviceD(cfg-extacl)# comment //Permit the communication using other packets.
```

```
DeviceD(cfg-extacl)# exit
DeviceD(cfg)# ip access-control-list extended inter_vlan_access2
DeviceD(cfg-extacl)# deny ip 192.168.2.0 0.0.0.255 192.168.3.0 0.0.0.255
DeviceD(cfg-extacl)# comment //Deny the mutual access between VLAN 2 and VLAN 3.
DeviceD(cfg-extacl)# deny udp any eq bootpc any eq bootps
DeviceD(cfg-extacl)# deny udp any eq bootps any eq bootpc
DeviceD(cfg-extacl)# comment //Deny DHCP packets.
DeviceD(cfg-extacl)# permit ip any any
DeviceD(cfg-extacl)# comment //Permit the communication using other packets.
DeviceD(cfg-extacl)# exit
DeviceD(cfg)# ip access-control-list extended inter_vlan_access3
DeviceD(cfg-extacl)# deny ip 192.168.3.0 0.0.0.255 192.168.2.0 0.0.0.255
DeviceD(cfg-extacl)# comment //Deny the mutual access between VLAN 2 and VLAN 3.
DeviceD(cfg-extacl)# permit ip any any
DeviceD(cfg-extacl)# comment //Permit the communication using other packets.
DeviceD(cfg-extacl)# exit
```

On device C, configure an IP extended ACL and add access rules.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ip access-control-list extended access_deny
DeviceC(cfg-extacl)# deny tcp 192.168.1.0 0.0.0.255 eq telnet any eq telnet
DeviceC(cfg-extacl)# comment //Deny VLAN 1 access to VLAN 3 through telnet.
DeviceC(cfg-extacl)# deny tcp 192.168.1.0 0.0.0.255 eq 22 any eq 22
DeviceC(cfg-extacl)# comment //Deny VLAN 1 access to VLAN 3 through SSH.
DeviceC(cfg-extacl)# exit
```

(3) Apply the IP extended ACLs.

On device D, apply the IP extended ACL to the corresponding ports.

```
DeviceD(cfg)# interface vlan 1
DeviceD(cfg-if-VLAN 1)# ip access-apply inter_vlan_access1 in
DeviceD(cfg-if-VLAN 1)# exit
DeviceD(cfg)# interface vlan 2
DeviceD(cfg-if-VLAN 2)# ip access-apply inter_vlan_access2 in
DeviceD(cfg-if-VLAN 2)# exit
DeviceD(cfg)# interface vlan 3
DeviceD(cfg-if-VLAN 3)# ip access-apply inter_vlan_access3 in
DeviceD(cfg-if-VLAN 3)# exit
```

On device C, apply the IP extended ACL to the line connected to the switch.

```
hostC(cfg)# line vty 0
hostC(cfg-line)# access-class access_deny in
hostC(cfg-line)# exit
```

5. Verification

(1) Verify the connectivity.

The ping operation succeeds between VLAN 1 and VLAN 2, and between VLAN 1 and VLAN 3.

```
DeviceA# ping 192.168.2.2
Sending 5, 100-byte ICMP Echoes to 192.168.2.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/8/10 ms
DeviceA#
DeviceA# ping 192.168.3.2
Sending 5, 100-byte ICMP Echoes to 192.168.3.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/8/10 ms
```

The ping operation fails between VLAN 2 and VLAN 3.

```
DeviceB# ping 192.168.3.2
Sending 5, 100-byte ICMP Echoes to 192.168.3.2, timeout is 2 seconds:
   < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5)
```

(2) The host in VLAN 1 cannot access VLAN 3 through telnet.

```
DeviceA# ping 192.168.3.2
Sending 5, 100-byte ICMP Echoes to 192.168.3.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/8/10 ms
DeviceA#
DeviceA# telnet 192.168.3.2
Trying 192.168.3.2, 23...
% Destination unreachable; gateway or host down
```

6. Configuration Files

- Device D configuration file

```
hostname DeviceD
!
vlan 1
!
vlan 2
!
vlan 3
!
ip access-control-list extended inter_vlan_access1
 10 deny udp any eq bootps any eq bootpc
 20 deny udp any eq bootpc any eq bootps
 comment //Deny DHCP packets.
 30 permit ip any any
```

```
 comment //Permit the communication using other packets.
!
ip access-control-list extended inter_vlan_access2
 10 deny ip 192.168.2.0 0.0.0.255 192.168.3.0 0.0.0.255
 comment //Deny the mutual access between VLAN 2 and VLAN 3.
 20 deny udp any eq bootpc any eq bootps
 30 deny udp any eq bootps any eq bootpc
 comment //Deny DHCP packets.
 40 permit ip any any
 comment //Permit the communication using other packets.
!
ip access-control-list extended inter_vlan_access3
 10 deny ip 192.168.3.0 0.0.0.255 192.168.2.0 0.0.0.255
 comment //Deny the mutual access between VLAN 2 and VLAN 3.
 20 permit ip any any
 comment //Permit the communication using other packets.
!
interface GigabitEthernet 1/0
 switch-port access vlan 1
 description link_to_DeviceA
!
interface GigabitEthernet 1/1
 switch-port access vlan 2
 description link_to_DeviceB
!
interface GigabitEthernet 1/2
 switch-port access vlan 3
 description link_to_DeviceC
!
interface VLAN 1
 ip access-apply inter_vlan_access1 in
 ip address 192.168.1.1 255.255.255.0
!
interface VLAN 2
 ip access-apply inter_vlan_access2 in
 ip address 192.168.2.1 255.255.255.0
!
interface VLAN 3
 ip access-apply inter_vlan_access3 in
 ip address 192.168.3.1 255.255.255.0
!
```

- Device A configuration file

```
hostname DeviceA
!
```

```
interface GigabitEthernet 0/1
 ip address 192.168.1.2 255.255.255.0
!
```

- Device B configuration file

```
hostname DeviceB
!
interface GigabitEthernet 0/1
 ip address 192.168.2.2 255.255.255.0
!
```

- Device C configuration file

```
hostname DeviceC
!
ip access-control-list extended access_deny
 10 deny tcp 192.168.1.0 0.0.0.255 eq telnet any eq telnet
 comment //Denies VLAN 1 access to VLAN 3 through telnet.
 20 deny tcp 192.168.1.0 0.0.0.255 eq 22 any eq 22
 comment //Deny VLAN 1 access to VLAN 3 through SSH.
!
interface GigabitEthernet 0/1
 ip address 192.168.3.2 255.255.255.0
!
line vty 0
 access-class access_deny in
 login
 password abcdef
!
```

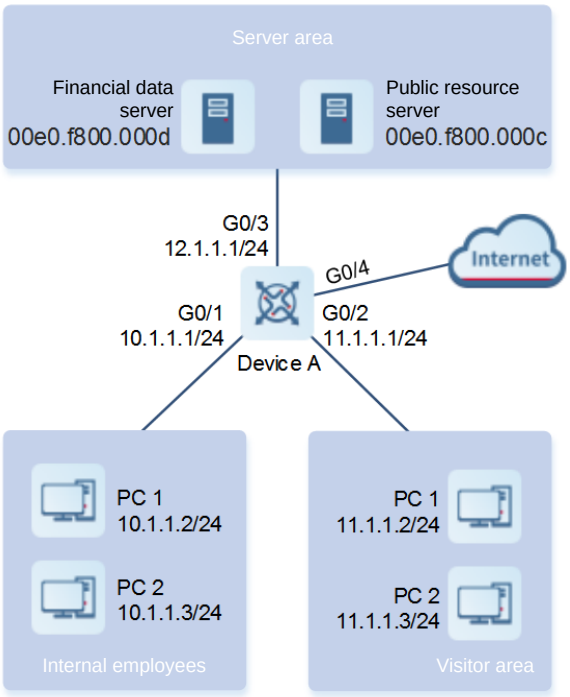
1.14.3 Configuring a MAC Extended ACL

1. Requirements

A MAC extended ACL needs to be configured to restrict the resources accessible to visitors.

2. Topology

Figure 1-5 Configuring an Extended MAC ACL



3. Notes

- On device A, configure a MAC extended ACL and add access rules. The PCs in the visitor area are allowed to access the Internet and the company's internal public server, but are not allowed to access the company's financial data server, that is, the PCs are prohibited from accessing the server with MAC address 00e0.f800.000d.
- On device A, apply the MAC extended ACL to the inbound direction of a port connected to the visitor area.

4. Procedure

(1) Configure a MAC extended ACL and add access rules.

On device A, configure a MAC extended ACL and add access rules.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# mac access-control-list extended 700
DeviceA(cfg-macacl)# deny any host 00e0.f800.000d
DeviceA(cfg-macacl)# permit any any
DeviceA(cfg-macacl)# exit
```

(2) Apply the MAC extended ACL to a port.

On device A, apply the ACL to the inbound direction of the port connected to the visitor area.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# mac access-apply 700 in
```

5. Verification

Check whether the ACL configuration commands are correct on device A.

```
DeviceA# display access-control-lists
mac access-control-list extended 700
10 deny any host 00e0.f800.000d etype-any
20 permit any any etype-any
DeviceA# display access-apply
mac access-apply 700 in
Applied On interface GigabitEthernet 0/2
```

Ping the financial data server from a visitor PC and confirm that the ping operation fails.

Ping the public resource server from a visitor PC and confirm that the ping operation succeeds.

Access the Internet through a visitor PC, for example, access www.baidu.com, and confirm that the home page can be opened.

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
mac access-control-list extended 700
 10 deny any host 00e0.f800.000d
 20 permit any any
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 10.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 mac access-apply 700 in
 ip address 11.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/3
 unset switch-port
 ip address 12.1.1.1 255.255.255.0
!
```

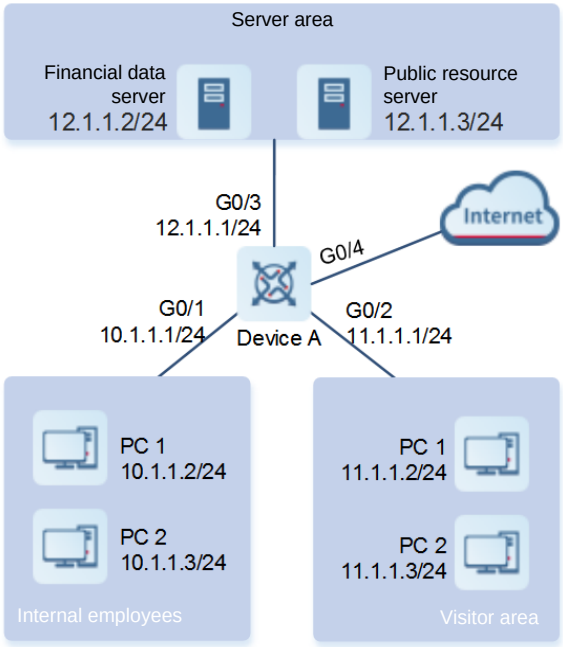
1.14.4 Configuring an Expert Extended ACL

1. Requirements

An expert extended ACL needs to be configured to restrict the resources accessible to visitors. Visitors cannot access the PCs of internal employees or the financial data server of the company, but can access the public resource server and Internet.

2. Topology

Figure 1-6 Configuring an Extended Expert ACL



3. Notes

- On device A, configure an expert extended ACL and add access rules, including:
  - Prohibiting the hosts in the visitor area from sending packets destined to the network segment of internal employees' PCs.
  - Prohibiting visitors from accessing the financial data server.
  - Permitting all the other packets to pass through.
- On device A, apply the ACL to the inbound direction of the port connected to the visitor area.

4. Procedure

(1) Configure an expert extended ACL and add access rules.

On device A, configure an expert extended ACL and add access rules.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# expert access-control-list extended 2700
DeviceA(cfg-expaclcfg-expacl)# deny ip any any 10.1.1.0 0.0.0.255 any
DeviceA(cfg-expacl)# deny ip any any host 12.1.1.2 any
DeviceA(cfg-expacl)# permit any any any any
DeviceA(cfg-expacl)# exit
```

(2) Apply the expert extended ACL to a port.

On device A, apply the ACL to the inbound direction of the port connected to the visitor area.

```
DeviceA(cfg)# interface gigabitethernet 0/2
```

```
DeviceA(cfg-if-GigabitEthernet 0/2)# expert access-apply 2700 in
```

5. Verification

Check whether the ACL configuration commands are correct on device A.

```
DeviceA(cfg)# display access-control-lists
expert access-control-list extended 2700
  10 deny ip any any 192.168.1.0 0.0.0.255 any
20 deny ip any any host 10.1.1.1 any
30 permit ip any any any any

DeviceA(cfg)# display access-apply
expert access-apply 2700in
Applied On interface GigabitEthernet 0/2
```

Ping the financial data server from a visitor PC and confirm that the ping operation fails.

Ping the public resource server from a visitor PC and confirm that the ping operation fails.

Ping the internal employee gateway 192.168.1.1 from a visitor PC and confirm that the ping operation fails.

Access the Internet through a visitor PC, for example, access www.baidu.com, and confirm that the home page can be opened.

6. Configuration Files

Device A configuration file

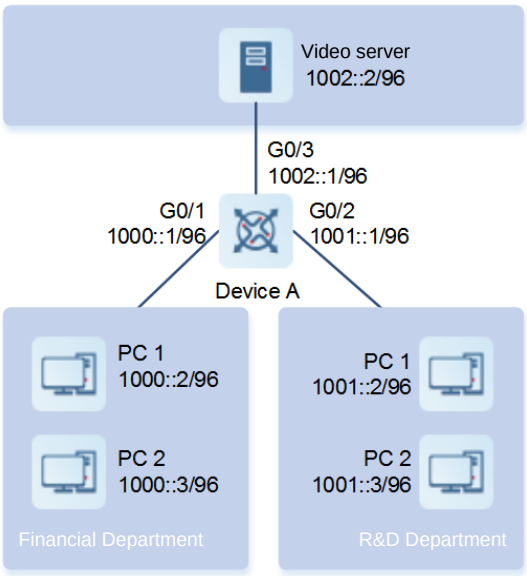
```
hostname DeviceA
!
expert access-control-list extended 2700
  10 deny ip any any 10.1.1.0 0.0.0.255 any
  20 deny ip any any host 12.1.1.2 any
  30 permit ip any any any any
!
interface GigabitEthernet 0/1
  unset switch-port
  ip address 10.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
  unset switch-port
  expert access-apply 2700 in
  ip address 11.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/3
  unset switch-port
  ip address 12.1.1.1 255.255.255.0
!
```

1.14.5 Configuring an IPv6 ACL

1. Requirements

An IPv6 ACL needs to be configured to prevent hosts in the R&D department from accessing the video server.

2. Topology



Configuring an IPv6 ACL

3. Notes

- On device A, configure an IPv6 ACL and add access rules, including:
  - Prohibiting access to the IPv6 address of the video server.
  - Permitting all the IPv6 packets to pass through.
- On device A, apply the IPv6 ACL to the inbound direction of the port connected to the R&D department.

4. Procedure

(1) Configure an IPv6 ACL and add access rules.

On device A, configure an IPv6 ACL and add access rules.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ipv6 access-control-list dev_deny_ipv6video
DeviceA(cfg-ipv6acl)# deny ipv6 any host 1002::2
DeviceA(cfg-ipv6acl)# permit ipv6 any any
DeviceA(cfg-ipv6acl)# exit
```

(2) Apply the IPv6 ACL to a port.

On device A, apply the ACL to the inbound direction of the port connected to the R&D department.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# ipv6 access-apply dev_deny_ipv6video in
```

5. Verification

Check whether the ACL configuration commands are correct on device A.

```
DeviceA(cfg)# display access-control-lists

ipv6 access-control-list dev_deny_ipv6video
10 deny ipv6 any host 200::1
20 permit ipv6 any any

DeviceA(cfg)# display access-apply
ipv6 access-apply dev_deny_ipv6video in
Applied On interface GigabitEthernet 0/2
```

Ping the video server from a PC in the R&D department and confirm that the ping operation fails.

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
ipv6 access-control-list dev_deny_ipv6video
 10 deny ipv6 any host 1002::2
 20 permit ipv6 any any
!
interface GigabitEthernet 0/1
 unset switch-port
 ipv6 address 1000::1/96
!
interface GigabitEthernet 0/2
 unset switch-port
 ipv6 access-apply dev_deny_ipv6video in
 ipv6 address 1001::1/96
!
interface GigabitEthernet 0/3
 unset switch-port
 ipv6 address 1002::1/96
!
```

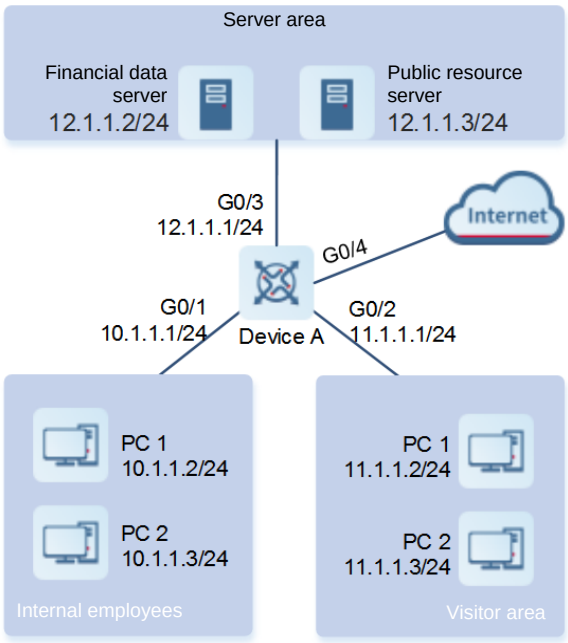
1.14.6 Configuring an ACL80

1. Requirements

An ACL80, namely, an expert advanced ACL needs to be configured to restrict the resources accessible by visitors. Visitors cannot access the PCs of internal employees or the financial data server of the company, but can access the public resource server and Internet.

2. Topology

Figure 1-7 Configuring an ACL80



3. Notes

- On device A, configure an expert advanced ACL and add access rules, including:
  - Prohibiting the hosts in the visitor area from sending packets destined to the network segment of internal employees' PCs.
  - Prohibiting visitors from accessing the financial data server.
  - Permitting all the other packets to pass through.
- On device A, apply the ACL to the inbound direction of the port connected to the visitor area.

4. Procedure

(1) Configure an expert advanced ACL and add access rules.

On device A, configure an expert advanced ACL and add access rules.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# expert access-control-list advanced acl80-guest
DeviceA(cfg-expadvacl)# deny 0800 FFFF 24 0A0101 FFFFFF 42
DeviceA(cfg-expadvacl)# deny 0800 FFFF 24 0C010102 FFFFFFFF 42
DeviceA(cfg-expadvacl)# permit 0806 FFFF 24
DeviceA(cfg-expadvacl)# permit 0800 FFFF 24
DeviceA(cfg-expadvacl)# exit
```

(2) Apply the expert advanced ACL to a port.

On device A, apply the ACL80 to the inbound direction of the port connected to the visitor area.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# expert access-apply acl80-guest in
```

5. Verification

Check whether the ACL configuration commands are correct on device A.

```
DeviceA(cfg)# display access-control-lists
expert access-control-list advanced sss
 10 deny 0800 FFFF 24 0A0101 FFFFFF 42
 20 deny 0800 FFFF 24 0C010102 FFFFFFFF 42
 30 permit 0806 FFFF 24
 40 permit 0800 FFFF 24

expert access-apply acl80-guest in
Applied On interface GigabitEthernet 0/2
```

Ping the financial data server from a visitor PC and confirm that the ping operation fails.

Ping the public resource server from a visitor PC and confirm that the ping operation succeeds.

Ping the internal employee gateway 192.168.1.1 from a visitor PC and confirm that the ping operation fails.

Access the Internet through a visitor PC, for example, access www.baidu.com, and confirm that the home page can be opened.

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
expert access-control-list advanced acl80-guest
 10 deny 0800 FFFF 24 0A0101 FFFFFF 42
 20 deny 0800 FFFF 24 0C010102 FFFFFFFF 42
 30 permit 0806 FFFF 24
 40 permit 0800 FFFF 24
!
interface GigabitEthernet 0/1
 unset switch-port
 ip address 10.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 unset switch-port
 expert access-apply 2700 in
 ip address 11.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/3
 unset switch-port
 ip address 12.1.1.1 255.255.255.0
!
```

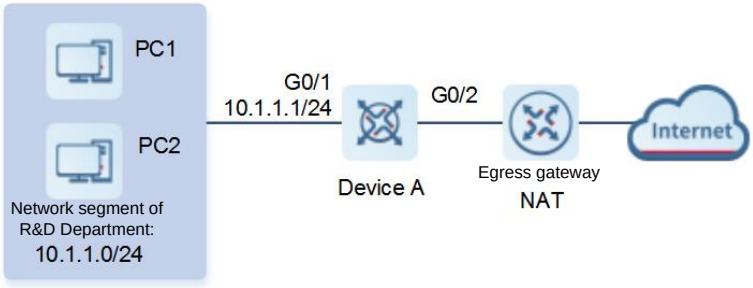
1.14.7 Configuring ACL Rules Based on Time Range

1. Requirements

ACL rules based on time range need to be configured to permit the R&D department to access the Internet only from 12:00 to 13:30 every day.

2. Topology

Figure 1-8 Configuring ACL Rules Based on Time Ranges



3. Notes

- Configure a time range on device A and add the time range entry from 12:00 to 13:30 every day.
- On device A, configure an IP standard ACL and add access rules, including:
  - Permitting packets from the source IP network segment 10.1.1.0/24 and associating the rule with the time range named access-internet.
  - Denying packets from the source IP network segment 10.1.1.0/24. This rule indicates that the Internet access is rejected beyond the time range.
  - Permitting packets from network segment addresses other than the R&D network segment addresses.
- On device A, apply the ACL to the inbound direction of the port connected to the R&D department.

4. Procedure

(1) Configure a time range.

Configure a time range on device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# time-range access-internet
DeviceA(cfg-timerange)# periodic daily 12:00 to 13:30
DeviceA(cfg-timerange)# exit
```

(2) Configure an IP standard ACL and add access rules.

On device A, configure an IP standard ACL and add access rules.

```
DeviceA(cfg)# ip access-control-list standard ip_std_internet_acl
DeviceA(cfg-stdacl)# permit 10.1.1.0 0.0.0.255 time-range access-internet
DeviceA(cfg-stdacl)# deny 10.1.1.0 0.0.0.255
DeviceA(cfg-stdacl)# permit any
```

```
DeviceA(cfg-stdacl)# exit
```

(3) Apply the IP standard ACL to a port.

On device A, apply the ACL to the inbound direction of the port connected the R&D department.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip access-apply ip_std_internet_acl in
```

5. Verification

Check whether the ACL configuration commands are correct on device A.

```
DeviceA# display time-range

time-range entry: access-internet (inactive)
   periodic Daily 12:00 to 13:30

DeviceA# display access-control-lists

ip access-control-list standard ip_std_internet_acl
 10 permit 10.1.1.0 0.0.0.255 time-range access-internet (inactive)
 20 deny 10.1.1.0 0.0.0.255
 30 permit any

DeviceA# display access-apply
ip access-apply ip_std_internet_acl in
Applied On interface GigabitEthernet 0/1
```

During the effective period of the time range (12:00 to 13:30), visit the home page of www.baidu.com from a PC in the R&D department, and confirm that it can be accessed.

During the ineffective period (time beyond 12:00 to 13:30), visit the home page of www.baidu.com from a PC in the R&D department, and confirm that it cannot be accessed.

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
ip access-control-list standard ip_std_internet_acl
 10 permit 10.1.1.0 0.0.0.255 time-range access-internet
 20 deny 10.1.1.0 0.0.0.255
 30 permit any
!
time-range access-internet
 periodic daily 12:00 to 13:30
!
interface GigabitEthernet 0/1
 unset switch-port
 ip access-apply ip_std_internet_acl in
 ip address 10.1.1.1 255.255.255.0
```

!

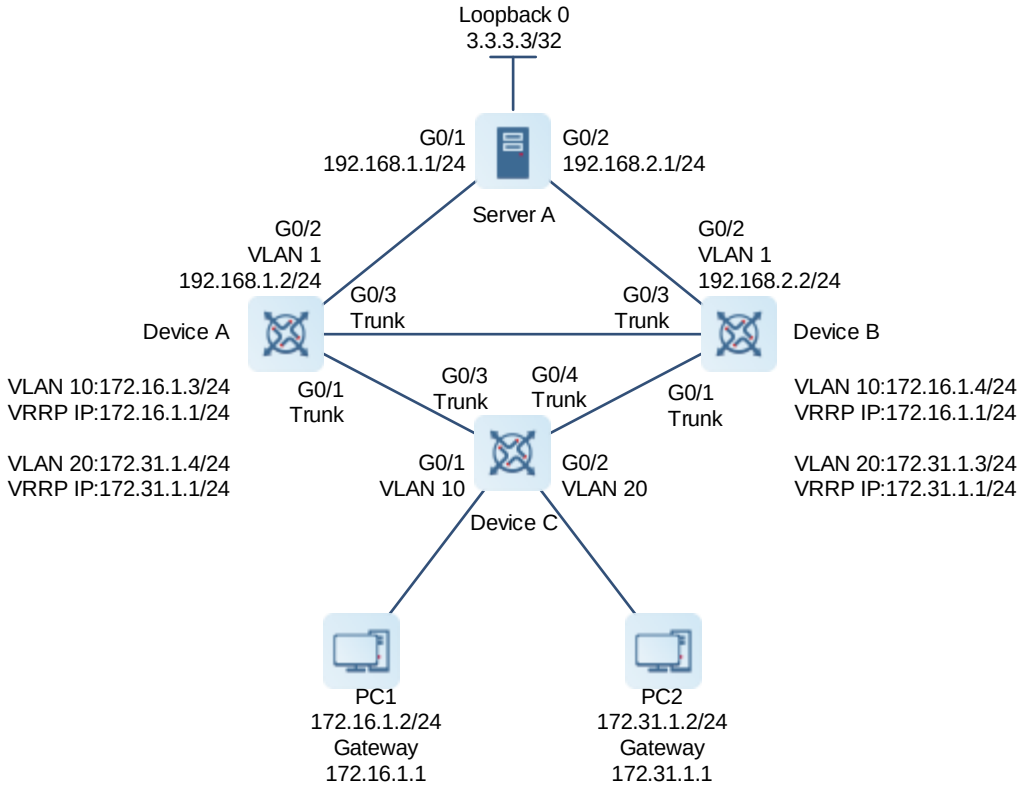
1.14.8 Configuring an SVI Router ACL

1. Requirements

Only Layer 3 communication between hosts in the scenario of a VRRP with multiple VLANs. Configure an ACL that only permits the access between hosts any denies the access request from other network segments.

2. Topology

Figure 1-9 Configuring an SVI Router ACL



3. Notes

- Device A and device B form a VRRP scenario. Both PC 1 and PC 2 are connected to device C.
- Configure STP to remove loops among device A, device B and device C.
- Use an SVI address as the gateway of PC 1 and PC 2.
- Configure an ACL that permits only the access between hosts and denies the access request from other network segments. Apply the ACL to the SVI. In this case, device A and device B are both hosts in the VRRP group.
- After the **svi router-acls enable** command is configured. Device A and device B in the VRRP group serve as a host and a backup respectively. VRRP returns to normal.

4. Procedure

(1) Configure a VLAN.

The configurations on device A, device B, device C and device D are the same. The following example configures a VLAN on device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan 10
DeviceA(cfg-vlan)# exit
DeviceA(cfg)# vlan 20
DeviceA(cfg-vlan)# exit
```

(2) Configure a VRRP group.

Configure VRRP on device A.

```
DeviceA(cfg)# interface VLAN 10
DeviceA(cfg-if-VLAN 10)# ip address 172.16.1.3 255.255.255.0
DeviceA(cfg-if-VLAN 10)# vrrp-config 10 ip 172.16.1.1
DeviceA(cfg-if-VLAN 10)# vrrp-config 10 priority 120
DeviceA(cfg-if-VLAN 10)# exit
DeviceA(cfg)# interface VLAN 20
DeviceA(cfg-if-VLAN 20)# ip address 172.31.1.4 255.255.255.0
DeviceA(cfg-if-VLAN 20)# vrrp-config 20 ip 172.31.1.1
```

Configure VRRP on device B.

```
DeviceB(cfg)# interface VLAN 10
DeviceB(cfg-if-VLAN 10)# ip address 172.16.1.4 255.255.255.0
DeviceB(cfg-if-VLAN 10)# vrrp-config 10 ip 172.16.1.1
DeviceB(cfg-if-VLAN 10)# exit
DeviceB(cfg)# interface VLAN 20
DeviceB(cfg-if-VLAN 20)# ip address 172.31.1.3 255.255.255.0
DeviceB(cfg-if-VLAN 20)# vrrp-config 20 ip 172.31.1.1
DeviceB(cfg-if-VLAN 20)# vrrp-config 20 priority 120
DeviceB(cfg-if-VLAN 20)# exit
```

(3) Configure STP to remove loops.

The configurations on device A, device B, device C and device D are the same. The following example configures STP on device A.

```
DeviceA(cfg)# stp
```

(4) Configure an ACL.

The configurations on device A and device B are the same. The following example configures an ACL on device A.

```
DeviceA(cfg)# ip access-control-list standard 10
DeviceA(cfg-stdacl)# permit host 3.3.3.3
DeviceA(cfg-stdacl)# deny any
DeviceA(cfg-stdacl)# exit
```

(5) Apply an ACL to an SVI.

The configurations on device A and device B are the same. The following example applies an ACL to an interface of device A.

```
DeviceA(cfg)# int vlan 20
DeviceA(cfg-if-VLAN 20)# ip access-apply 10 in
```

(6) Configure the **svi router-acls enable** command on device A.

The configurations on device A and device B are the same. The following example configures the **svi router-acls enable** command on device A.

```
DeviceA(cfg)# svi router-acls enable
```

5. Verification

Check the VRRP status on device A.

```
DeviceA# display-config
Interface      Grp  Pri  timer   Own  Pre  State    Master addr     Group addr
VLAN 10        10   120  3.53    -    P    Master   172.16.1.3      172.16.1.1
VLAN 20        20   100  3.60    -    P    Backup   172.31.1.3      172.31.1.1
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
vlan 1
!
vlan 10
!
vlan 20
!
stp
!
ip access-control-list standard 10
  10 permit host 3.3.3.3
  20 deny any
!
svi router-acls enable
!
interface GigabitEthernet 0/1
  switch-port mode trunk
!
interface GigabitEthernet 0/3
  switch-port mode trunk
!
interface VLAN 1
  ip address 192.168.1.2 255.255.255.0
!
interface VLAN 10
  ip address 172.16.1.3 255.255.255.0
  vrrp-config 10 priority 120
```

```
 vrrp-config 10 ip 172.16.1.1
!
interface VLAN 20
 ip access-apply 10 in
 ip address 172.31.1.4 255.255.255.0
 vrrp-config 20 ip 172.31.1.1
!
ip route-config 3.3.3.0 255.255.255.0 192.168.1.1
!
```

- Device B configuration file

```
hostname DeviceB
!
vlan 1
!
vlan 10
!
vlan 20
!
stp!
ip access-control-list standard 10
 10 permit host 3.3.3.3
 20 deny any
!
svi router-acls enable
!
interface GigabitEthernet 0/1
 switch-port mode trunk
!
interface GigabitEthernet 0/3
 switch-port mode trunk
!
interface VLAN 1
 ip address 192.168.2.2 255.255.255.0
!
interface VLAN 10
 ip access-apply 10 in
 ip address 172.16.1.4 255.255.255.0
 vrrp-config 10 ip 172.16.1.1
!
interface VLAN 20
 ip address 172.31.1.3 255.255.255.0
 vrrp-config 20 priority 120
 vrrp-config 20 ip 172.31.1.1
!
```

```
ip route-config 3.3.3.0 255.255.255.0 192.168.2.1
!
```

- Device C configuration file

```
hostname DeviceC
!
vlan 1
!
vlan 10
!
vlan 20
!
interface GigabitEthernet 0/1
 switch-port access vlan 10
!
interface GigabitEthernet 0/2
 switch-port access vlan 20
!
interface GigabitEthernet 0/3
 switch-port mode trunk
!
interface GigabitEthernet 0/4
 switch-port mode trunk
!
```

- Server A configuration file

```
hostname ServerA
!
interface GigabitEthernet 0/1
 ip address 192.168.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 ip address 192.168.2.1 255.255.255.0
!
interface Loopback 0
 ip address 3.3.3.3 255.255.255.0
!
ip route-config 172.16.1.0 255.255.255.0 192.168.1.2
ip route-config 172.31.1.0 255.255.255.0 192.168.2.2
!
```

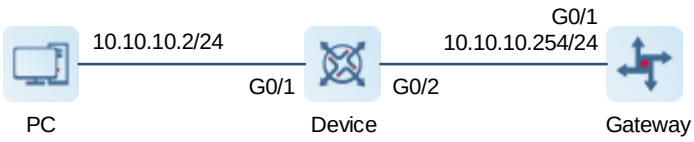
1.14.9 Configuring the ACL Packet Counting Function

1. Requirements

When you apply an ACL, you can configure the **counter-only** option to count the packets with some characteristics. The following example counts the ICMP packets discarded by the PC ping gateway and locate the discarded packets.

2. Topology

Figure 1-10 Configuring ACL Packet Counting



3. Notes

- Apply ACLs to both inbound and outbound directions of G0/1 and G0/2 on the device. You need to configure four ACLs on the device to match the ICMP packets from the PC to the gateway and from the gateway to the PC.
- Apply ACLs to both inbound and outbound directions of G0/1 on the gateway. You need to configure two ACLs on the gateway to match the ICMP packets from the PC to the gateway and from the gateway to the PC.
- Configure the **counter-only** option when you apply an ACL.
- The function only counts the packets matching the permit rule in the ACL and does not count those matching the deny rule.

4. Procedure

(1) Configure an ACL.

Configure four ACLs on the device to match the ICMP packets from the PC to the gateway and from the gateway to the PC.

```
Device> privilege
Device# config-system terminal
Device(cfg)# ip access-control-list extend 100
Device(cfg-extacl)# permit icmp host 10.10.10.1 host 10.10.10.254
Device(cfg-extacl)# exit
Device(cfg)# ip access-control-list extend 101
Device(cfg-extacl)# permit icmp host 10.10.10.254 host 10.10.10.1
Device(cfg-extacl)# exit
Device(cfg)# ip access-control-list extend 102
Device(cfg-extacl)# permit icmp host 10.10.10.1 host 10.10.10.254
Device(cfg-extacl)# exit
Device(cfg)# ip access-control-list extend 103
Device(cfg-extacl)# permit icmp host 10.10.10.254 host 10.10.10.1
Device(cfg-extacl)# exit
```

Configure two ACLs on the gateway to match the ICMP packets from the PC to the gateway and from the gateway to the PC.

```
Gateway> privilege
Gateway #config-system terminal
Gateway(cfg)# ip access-control-list extend 100
Gateway(cfg-extacl)# permit icmp host 10.10.10.1 host 10.10.10.254
Gateway(cfg-extacl)# exit
Gateway(cfg)# ip access-control-list extend 101
Gateway(cfg-extacl)# permit icmp host 10.10.10.254 host 10.10.10.1
Gateway(cfg-extacl)# exit
```

(2) Apply ACLs.

Apply ACLs to both inbound and outbound directions of the interconnected interface, G0/1, on the gateway and the device.

```
DeviceA(cfg)# interface VLAN 10
DeviceA(cfg-if-VLAN 10)# ip address 172.16.1.3 255.255.255.0
DeviceA(cfg-if-VLAN 10)# vrrp-config 10 ip 172.16.1.1
DeviceA(cfg-if-VLAN 10)# vrrp-config 10 priority 120
DeviceA(cfg-if-VLAN 10)# exit
DeviceA(cfg)# interface VLAN 20
DeviceA(cfg-if-VLAN 20)# ip address 172.31.1.4 255.255.255.0
DeviceA(cfg-if-VLAN 20)# vrrp-config 20 ip 172.31.1.1
```

Apply ACLs to both inbound and outbound directions of the interconnected interface, G0/2, on the gateway and the device.

```
Device# config-system terminal
Device(cfg)# interface gigabitEthernet 0/2
Device(cfg-if-GigabitEthernet 0/2)# ip access-apply 103 in counter-only
Device(cfg-if-GigabitEthernet 0/2)# ip access-apply 102 out counter-only
Device(cfg-if-GigabitEthernet 0/2)# exit
```

Apply ACLs to both inbound and outbound directions of the interconnected interface, G0/1, on the device and the PC.

```
Device# config-system terminal
Device(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# ip access-apply 100 in counter-only
Device(cfg-if-GigabitEthernet 0/1)# ip access-apply 101 out counter-only
Device(cfg-if-GigabitEthernet 0/1)# exit
```

5. Verification

Send a total of 15 ICMP packets three times from the PC ping gateway address, 10.10.10.254. Display the count of ICMP packets on the device and the gateway respectively. Display the count of ICMP packets on the device.

```
Device# display access-control-list
ip access-control-list extended 100
    10 permit ip host 10.10.10.1 host 10.10.10.254 (55 packets matches, 98008 bytes matches, 0:00:23 ago )
```

```
ip access-control-list extended 101
    10 permit ip host 10.10.10.254 host 10.10.10.1 (55 packets matches, 98008 bytes matches, 0:00:23 ago )
ip access-control-list extended 102
    10 permit ip host 10.10.10.1 host 10.10.10.254 (55 packets matches, 98008 bytes matches, 0:00:23 ago )
ip access-control-list extended 103
    10 permit ip host 10.10.10.254 host 10.10.10.1 (55 packets matches, 98008 bytes matches, 0:00:23 ago )
```

Display the count of ICMP packets on the gateway.

```
Gateway# display access-control-list
ip access-control-list extended 100
    10 permit ip host 10.10.10.1 host 10.10.10.254 (55 packets matches, 98008 bytes matches, 0:00:23 ago )
ip access-control-list extended 101
    10 permit ip host 10.10.10.254 host 10.10.10.1 (55 packets matches, 98008 bytes matches, 0:00:23 ago )
```

Analyze the count of ICMP packets and locate the discarded packets

15 packets are received in the inbound direction of the interconnected interface, G0/1, of the device and the PC. (ACL 100 is applied to the device.)

15 packets are sent from the outbound direction of the interconnected interface, G0/2, of the device and the gateway. (ACL 102 is applied to the device.)

15 packets are received in the inbound direction of the interconnected interface, G0/1, of the gateway and the device. (ACL 100 is applied to the gateway.)

15 packets are sent from the outbound direction of the interconnected interface, G0/1, of the gateway and the device. (ACL 101 is applied to the gateway.)

10 packets are received in the inbound direction of the interconnected interface, G0/2, of the device and the gateway. (ACL 103 is applied to the device.)

The packets were discarded in the link between the device and the gateway.

6. Configuration Files

- Device configuration file

```
hostname Device
!
ip access-control-list extended 100
 10 permit icmp host 10.10.10.1 host 10.10.10.254
!
ip access-control-list extended 101
 10 permit icmp host 10.10.10.254 host 10.10.10.1
!
ip access-control-list extended 102
 10 permit icmp host 10.10.10.1 host 10.10.10.254
!
ip access-control-list extended 103
 10 permit icmp host 10.10.10.254 host 10.10.10.1
!
interface GigabitEthernet 0/1
 ip access-apply 100 in counter-only
```

```
 ip access-apply 101 out counter-only
!
interface GigabitEthernet 0/2
 ip access-apply 103 in counter-only
 ip access-apply 104 out counter-only
!
```

- Gateway configuration file

```
hostname Gateway
!
ip access-control-list extended 100
 10 permit icmp host 10.10.10.1 host 10.10.10.254
!
ip access-control-list extended 101
 10 permit icmp host 10.10.10.254 host 10.10.10.1
!
interface GigabitEthernet 0/1
 ip access-apply 100 in counter-only
 ip access-apply 101 out counter-only
 ip address 10.10.10.254 255.255.255.0
!
```

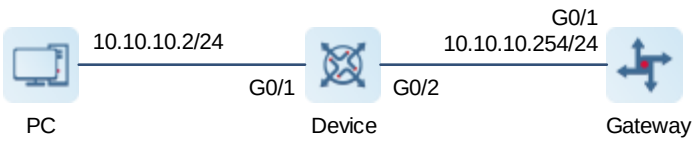
1.14.10 Configuring Hit-Log Output

1. Requirements

If the ACL contains the **hit-log** option, the number of packets with the specified feature is collected and output in the form of a hig-log. The following example counts the ICMP packets permitted by the standard ACL in a Ping operation and identifies the packet content.

2. Topology

Figure 1-11 Configuring Hit-Log Output



3. Notes

- Create a standard ACL named **std_acl**.
- Apply **ACL std_acl** in the ingress direction of interface G0/1.
- Enable **ACL std_acl** with counter.
- The output interval of hit-logs and of miss-logs are greater than 0s and 0 min respectively.

4. Procedure

(2) Configure the ACL.

Configure an ACL to permit ICMP packets from the PC to the gateway.

```
Device> privilege
Device# config-system terminal
Device(cfg)# ip access-control-list standard std_acl
Device(cfg-extacl)# permit any hit-log
Device(cfg-extacl)# exit
```

(3) Apply the ACL.

Apply the ACL in the ingress direction of interface G0/1.

```
Device(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# ip access-apply std_acl in
Device(cfg-if-GigabitEthernet 0/1)# exit
```

(4) Enable ACL counter.

Enable ACL counter.

```
Device(cfg)# ip access-control-list counter std_acl
```

(5) Configure the intervals of hit-logs and of miss-logs respectively.

Set the intervals of hit-logs and of miss-logs to 5s and 1 min respectively.

```
Device(cfg)# global access-control-list hit-log interval 5
Device(cfg)# global access-control-list miss-log interval 1
```

5. Verification

Ping the gateway IP address 10.10.10.254 from a PC and send a total of 600000 ICMP packets in 70s. Enable syslog on the device and verify hit-log and miss-log output.

```
Device# terminal monitor
Device# Mar 12 11:30:26 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (493) packet
Mar 12 11:30:31 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (495) packet
Mar 12 11:30:36 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (499) packet
Mar 12 11:30:41 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (470) packet
Mar 12 11:30:46 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (489) packet
Mar 12 11:30:51 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (493) packet
Mar 12 11:30:56 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (496) packet
Mar 12 11:31:01 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (498) packet
```

```
Mar 12 11:31:06 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (496) packet

Mar 12 11:31:11 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (466) packet

Mar 12 11:31:16 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (489) packet

Mar 12 11:31:21 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (409) packet

Mar 12 11:31:26 s6510-2: %ACLD-6-MISSLOG: ACL logging rate-limited or missed 486740 packets

Mar 12 11:31:26 S6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (499) packet

Mar 12 11:31:31 s6510-2: %ACLD-6-HITLOG: acl(std_acl) ace_sn(10) Permit icmp 10.10.10.2 (GigabitEthernet 0/1)-> 10.10.10.254, (413) packet
```

6. Configuration Files

- Device configuration file

```
hostname Device

!

ip access-control-list standard std_acl

 10 permit any hit-log

!

interface gigabitEthernet 0/1

 ip access-apply std_acl in

!

ip access-control-list counter std_acl

!

global access-control-list hit-log interval 5

global access-control-list miss-log interval 1
```

Contents

1 Configuring QoS 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.2 Configuration Task Summary 8

1.3 Configuring Traffic Classification 8

1.3.1 Overview 8

1.3.2 Configuration Tasks 9

1.3.3 Configuring a Class 9

1.3.4 Configuring a Policy 10

1.3.5 Applying a Policy 11

1.3.6 Configuring a Logical Interface Group 12

1.4 Configuring Priority Mapping 13

1.4.1 Overview 13

1.4.2 Configuration Tasks 13

1.4.3 Configuring the Trust Mode 13

1.4.4 Configuring Mappings 14

1.5 Configuring the Interface Rate Limit 15

1.5.1 Overview 15

1.5.2 Procedure 15

1.6 Configuring Congestion Management 15

1.6.1 Overview 15

1.6.2 Configuration Tasks 15
1.6.3 Notes 16
1.6.4 Configuring the CoS-to-Queue Mappings 16
1.6.5 Configuring the Scheduling Policy and Round Robin Weights for Output Queues 16
1.6.6 Configuring Bandwidth for a Queue 18
1.6.7 Enabling ECN for a Queue 18
1.6.8 Configuring Cumulative ECN Counter 19
1.6.9 Configuring the ECT Mark Type 20
1.7 Configuring Congestion Avoidance 20
1.7.1 Overview 20
1.7.2 Configuration Tasks 20
1.7.3 Enabling the WRED Function 21
1.7.4 Configuring Higher and Lower Threshold Values 21
1.7.5 Configuring the Maximum Discarding Probability 23
1.7.6 Configuring the Sampling Interval 23
1.7.7 Configuring the Sampling Weight 24
1.7.8 Configuring the CoS-to-Threshold Mappings 25
1.8 Configuring the QoS Global Policy 25
1.8.1 Overview 25
1.8.2 Procedure 25
1.9 Disabling the Packet Priority Change 26
1.9.1 Overview 26
1.9.2 Procedure 26
1.10 Configuring the Queue Display and Configuration Mode 26

1.10.1 Overview 26

1.10.2 Procedure 26

1.11 Monitoring 26

1.12 Configuration Examples 28

1.12.1 Configuring the Interface Rate Limit and Priority Re-marking 28

1.12.2 Configuring Queue Scheduling and Congestion Avoidance 31

1 Configuring QoS

1.1 Introduction

1.1.1 Overview

Quality of service (QoS) can meet users' requirements for different applications and different levels of service quality. It allocates and schedules resources based on users' requirements and provides different levels of service quality for different packets.

On a traditional IP network, a device treats all the packets in the same way, in which the device processes packets based on their arrival time according to the queuing strategy of first in first out (FIFO), and transmits the packets to the destination on a best-effort basis. When the network bandwidth is abundant, all the packets are properly processed; when the network is congested, all the packets may be discarded.

QoS assigns a transmission priority to the packets of a type to highlight the importance of the packets. Then, the devices provide special transmission services for these packets according to forwarding policies for different priorities, congestion avoidance, and other mechanisms. With QoS, a device processes real-time and important packets preferentially, processes non-real-time and common packets with lower priorities and even discards the packets upon network congestion.

QoS enhances the network performance predictability, effectively allocates network bandwidth, and reasonably utilizes network resources.

1.1.2 Principles

1. Basic Concepts

- DiffServ model

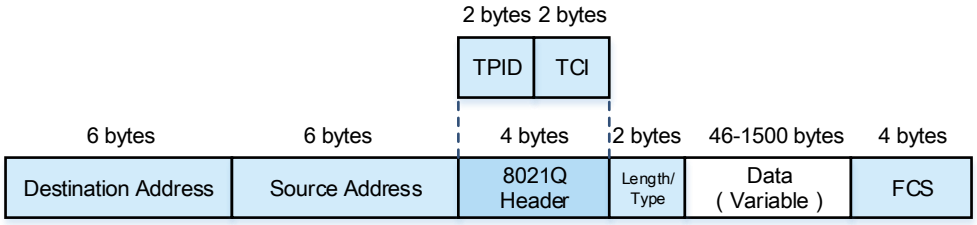
  The differentiated services (DiffServ) model classifies all packets transmitted on a network into different types. The classification information related to QoS priority marking is recorded in some fields of Layer 2 or Layer 3packets, for example, the PRI field of IEEE 802.1Q frames, type of service (ToS) field of IPv4 packets, traffic class (TC) field of IPv6 packets, and the MPLS experimental bits (EXP) field of multiprotocol label switching (MPLS) packets.

  In the network of DiffServ model, the classification information of packets can be assigned by hosts or other network devices or based on different application policies or different packet contents. A device applies the same transmission service policy to packets containing the same classification information and applies different transmission service policies to packets containing different classification information. Based on the classification information carried by packets, a device may provide different transmission priorities for different packets, reserve bandwidth for a kind of packets, discard certain packets with lower priorities, or take some other actions.

- PRI field of the IEEE 802.1q frames

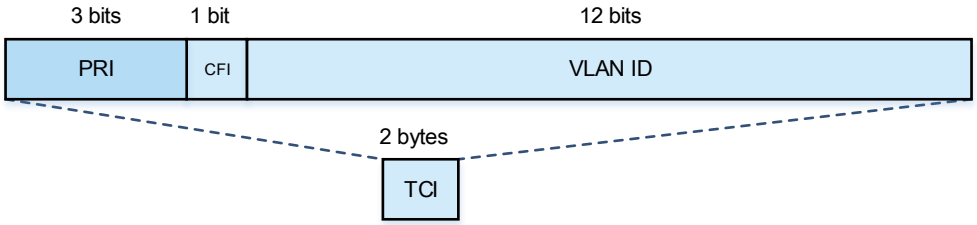
  The PRI field of IEEE 802.1Q frames (namely, the IEEE 802.1p priority) is located in the header of an Layer 2 packet containing an IEEE 802.1Q tag header, as shown in Figure 1-1.

Figure 1-1 Format of an Layer 2 Frame with an IEEE 802.1Q Tag Header



The 4-byte IEEE 802.1Q tag header contains the 2-byte tag protocol identifier (TPID) and 2-byte tag control information (TCI). TCI contains the 3-bit PRI field, as shown in Figure 1-2.

Figure 1-2 PRI Field of the IEEE 802.1q Frames

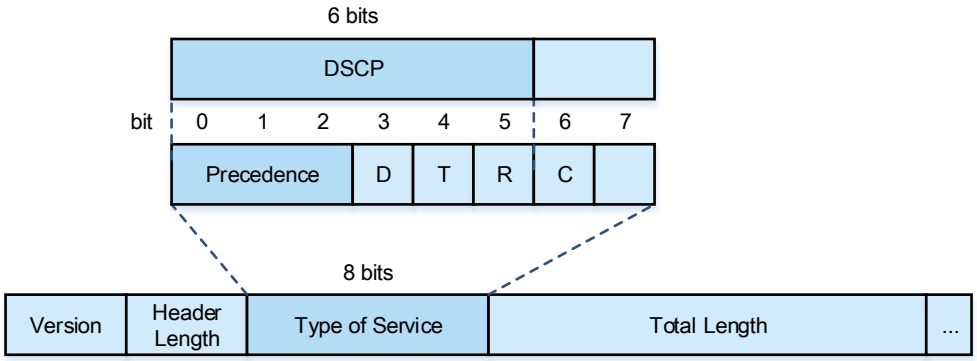


The PRI field represents eight priorities for packet transmission, and the priority values from high to low are 7, 6, ..., 1, and 0. The IEEE 802.1p priority is applicable to scenarios where Layer 3 headers do not need to be analyzed and QoS needs to be implemented only at Layer 2.

- ToS field of the IPv4 packets

  IPv4 packets use the ToS field in the IP header to indicate the priority of the packets, as shown in Figure 1-3.

Figure 1-3 ToS Field in the IP Header



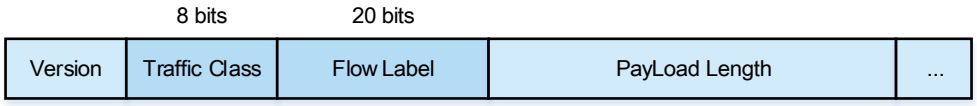
The ToS field contains eight bits, of which the first three bits are the IP PRE (precedence) field and represent eight priorities for packet transmission, with the priority values from high to low being 7, 6, ..., 1, and 0.

RFC 2474 redefines the ToS field of the IP header, in which the first 6 bits (bits 0 to 5) represent the differentiated services code point (DSCP). DSCP is used to classify packets into a maximum of 64 different categories.

- ToS field of the IPv6 packets

  IPv6 packets use the TC field in the IPv6 header to indicate the packet priority, as shown in Figure 1-4Figure 1-4.

Figure 1-4 TC Field in the IPv6 Header

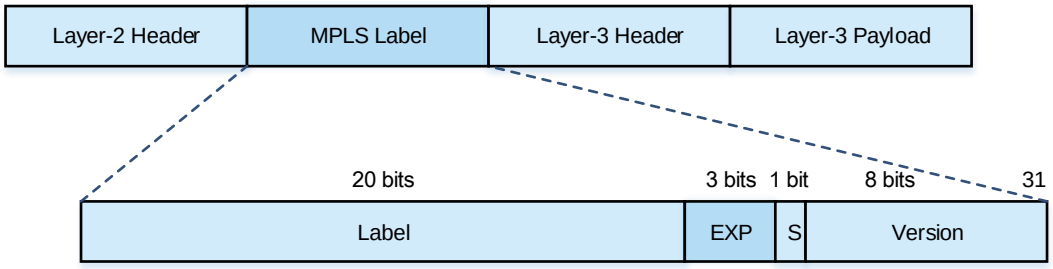


The TC field contains eight bits and provides the same function as the ToS field of IPv4 packets. The first six bits of the TC field indicate DSCP.

- EXP field of the MPLS packets

  MPLS packets use the EXP field in the label to indicate the packet priority, as shown in Figure 1-5.

Figure 1-5 EXP Field in the MPLS Label



The EXP field contains three bits, which represent eight priorities for packet transmission, and the priority values from high to low are 7, 6, ..., 1, and 0. Since the EXP field has the same length as the priority field (IP PRE) of IP packets, one-to-one mapping can be formed between them.

- CoS

  The class of service (CoS) is used to determine the packet queue number when an interface sends packets. The CoS values are 7, 6, ..., 1, and 0 from high to low.

2. Traffic Classification

Traffic classification identifies packets with certain characteristics according to rules. It is the prerequisite and basis for providing differentiated network services. Traffic classification is classified into simple traffic classification and complex traffic classification.

- Simple traffic classification

  Simple traffic classification classifies packets only based on the priorities carried in packets (for example, the PRI value of IEEE 802.1q frames, IP PRE or DSCP value of IP packets, TC value of IPv6 packets, and the EXP value of MPLS packets). Simple traffic classification implements the mapping from the priority to

the CoS of the device based on whether an interface is configured to trust the priority tag carried in uplink packets.

- Complex traffic classification

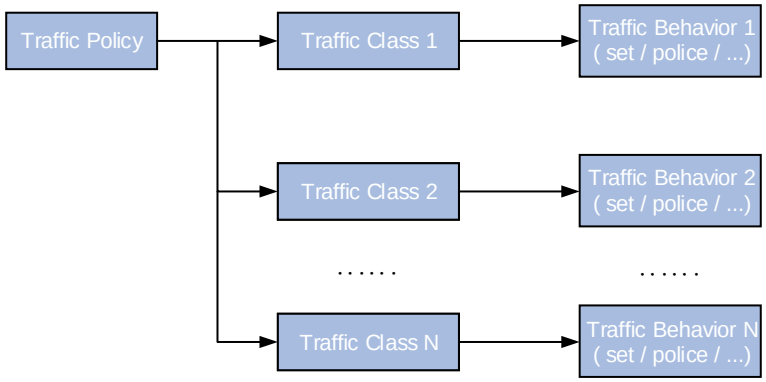
  Complex traffic classification classifies packets based on the priorities carried in packets or by identifying packets meeting certain characteristics according to access control list (ACL) rules.

  A QoS policy based on complex traffic classification associates the complex traffic classification with relevant traffic behaviors to form a policy, which is then applied to an interface to take effect. A QoS policy comprises three elements: class, traffic behavior, and policy.

  - A class comprises the class name and class rule. A class defines the matching rules of traffic classification to classify packets.
  - A traffic behavior defines the action to be taken for the classified packets, including modifying the packet priority (also known as priority re-marking) and monitoring traffic.
  - A policy comprises the policy name, and the binding relationship between a class and a traffic behavior. A policy binds a specified class and a traffic behavior and is applied to one or more interfaces to take effect.

The policy-based binding between multiple traffic classes and behaviors is as shown in Figure 1-6.

Figure 1-6 Policy-Based Binding Between Multiple Traffic Classes and Behaviors



- QoS logical interface group

  You can specify a series of interfaces (aggregation ports or Ethernet interfaces) as a QoS logical interface group, and associate a policy with the logical interface group for QoS processing. For example, the rate limiting function of a traffic behavior enables the packets complying with the rate limiting conditions to share the bandwidth limited by the policy on all the interfaces in the same logical interface group.

3. Priority Mapping

Priorities are used to identify the scheduling weights or forwarding priority of packets. Different priority types are defined for different packet types: IEEE 802.1q frames use the IEEE 802.1p priority, IP packets use the IP PRE or DSCP, and MPLS packets use the EXP priority, and so on.

After a packet enters a device interface, the packet priority is mapped to the CoS according to the trust mode configured for the interface. Table 1-1 shows the mappings between trust mode configured for an interface and the priorities.

Table 1-1 Interface Trust Mode and Priority Mapping

Interface Trust Mode	Priority Mapping
Untrusted	● No priority carried in packets is trusted. ● All packets received from an interface use the IEEE 802.1p value configured for the interface. ● The CoS values are obtained from the 802.1p-to-DSCP mapping table and DSCP-to-CoS mapping table. Packets are sent to queues based on their CoS values. ● If an IEEE 802.1Q tagged packet is sent from an egress, the packet priority is changed to the corresponding CoS value.
Trusting 802.1p	● If a packet received by an interface is an IEEE 802.1Q tagged packet, the carried IEEE 802.1p value is directly used. If a packet does not carry any tag, the IEEE 802.1p value configured for the interface is used. ● The CoS values are obtained from the 802.1p-to-DSCP mapping table and DSCP-to-CoS mapping table. Packets are sent to queues based on their CoS values. ● If an IEEE 802.1Q tagged packet is sent from an egress, the packet priority is changed to the corresponding CoS value.
Trusting DSCP	● If a packet received from an interface is a non-IP packet, it is processed in the same way of trusting 802.1p. If the received packet is an IP packet, the CoS value is obtained from the DSCP-to-CoS mapping table based on the DSCP value of the packet. Packets are sent to queues based on their CoS values. ● If an MPLS header needs to be encapsulated into a packet, the EXP value is changed according to the DSCP-to-EXP mapping table.
Trusting IP PRE	● If a packet received from an interface is a non-IP packet, it is processed in the same way of trusting 802.1p. If the received packet is an IP packet, the CoS value is obtained from the IP PRE-to-DSCP mapping table and DSCP-to-CoS mapping table based on the IP PRE value of the packet. Packets are sent to queues based on their CoS values.
Trusting EXP	● If a packet received from an interface is a non-MPLS packet, it is processed in the same way of trusting 802.1p. If the received packet is an MPLS packet, the CoS value is obtained from the EXP-to-CoS mapping table based on the EXP value of the packet. Packets are sent to queues based on their CoS values. ● After an MPLS header is decapsulated, the DSCP value is changed according to the EXP-to-DSCP mapping table.

4. Traffic Policing

Traffic policing supervises the rate of the specific traffic, limits the rate within a reasonable range, and discards the traffic out of the limit or re-marks the priority. Traffic policing adopts the committed access rate (CAR)

technology to control traffic. CAR uses the single-rate single-token bucket algorithm to judge whether the rate of each packet exceeds the specified traffic limit.

- Single-rate single token bucket

  The single-rate single token bucket contains two parameters: **CIR** and **CBS**.

  - Committed information rate (CIR): Indicates the rate at which a token is added to a token bucket.
  - Committed burst size (CBS): Indicates the capacity of a token bucket.

  The single-rate single token bucket adds tokens to a token bucket at a fixed rate (namely, CIR). When the number (that is, CBS) of tokens in a token bucket reaches the capacity of the token bucket, the excess tokens are discarded. When the device processes packets, it takes tokens out of the token bucket, with the number same as the size of packets (in bytes). When the number of tokens is sufficient, the packets can be forwarded. When the number of tokens is insufficient, the packets are discarded or the priorities of some packets are re-marked before forwarding.

- CAR processing

  CAR uses a token bucket to measure the packets passing through an interface, and implements the preset policing action based on the measurement results. These actions include:

  - Forwarding: Packets within the traffic limit are normally forwarded.
  - Discarding: Packets out of the traffic limit are discarded. The packets out of the total traffic limit of an interface are directly discarded.
  - Changing the priority and forwarding: The priorities of the packets out of the traffic limit are changed and then the packets are forwarded.

5. Congestion Management

When the receiving rate of packets exceeds the sending rate, congestion occurs on the sending interface. If no sufficient buffer is provided to store these packets, packet loss may occur. The congestion management mechanism determines the sending order of packets based on their local priorities. The congestion management function controls congestion and improves the local priorities of packets for some important data. When congestion occurs, the packets of higher priorities are sent first to ensure that key services are provided in time.

Congestion management adopts the queue scheduling mechanism. The processing is as follows:

(1) After each packet undergoes the QoS processing in a device, it obtains a CoS value finally.

(2) The device allocates the packets to the corresponding sending queues according to the CoS values.

(3) The outbound interface selects the packets in a queue for sending according to various queue scheduling policies (such as SP, RR, WRR, DRR, WFQ, SP+WRR, SP+DRR, and SP+WFQ).

- SP scheduling policy

  In strict-priority (SP) scheduling, packets are scheduled strictly based on their queue priorities from high to low. Before sending a packet, check whether there is a packet to be sent in a high-priority queue. If there is, send it. If not, check whether there is a packet to be sent in the next-level queue, and so on.

  The weakness of SP scheduling is that, when congestion occurs, if the packets in a higher priority queue exist for a long time, the packets in a lower priority queue have no opportunity of being scheduled.

- RR scheduling policy

  Round robin (RR) scheduling uses the round robin method to schedule multiple queues. Only one packet in a queue is processed each time.

- WRR scheduling policy

  The weighted round robin (WRR) scheduling solves the problem that weight cannot be set for RR scheduling. WRR scheduling also adopts the round robin method to schedule multiple queues. The number of packets in a queue processed each time is proportional to the weight of the queue. RR scheduling is equivalent to WRR scheduling with the weight 1.

  However, WRR scheduling fails to schedule the services with low delay requirements in time.

- DRR scheduling policy

  The deficit round robin (DRR) scheduling is similar to WRR scheduling, but implements scheduling based on the number of packet bytes, instead of the number of packets.

  However, DRR scheduling fails to schedule the services with low delay requirements in time.

- WFQ scheduling policy

  Weighted fair queuing (WFQ) scheduling fixes the problem that the queues using WRR scheduling have no fixed egress bandwidth. WFQ scheduling allocates an egress bandwidth to queues based on the queue weight, and different queues can have the opportunity of fair scheduling.

- SP+WRR scheduling policy

  SP scheduling is configured for one or more sending queues, and the other queues are scheduled in the WRR mode. Among SP queues, only after all the packets in an SP queue with a higher priority are sent, can the packets in an SP queue with a next higher priority be sent. Among SP and WRR queues, only after the packets in all SP queues are sent, can the packets in WRR queues be sent.

- SP+DRR scheduling policy

  SP scheduling is configured for one or more sending queues, and the other queues are scheduled in the DRR mode. Among SP queues, only after all the packets in an SP queue with a higher priority are sent, can the packets in an SP queue with a next higher priority be sent. Among SP and DRR queues, only after the packets in all SP queues are sent, can the packets in DRR queues be sent.

- SP+WFQ scheduling policy

  SP scheduling is configured for one or more sending queues, and the other queues are scheduled in the WFQ mode. Among SP queues, only after all packets in an SP queue with a higher priority are sent, can the packets in an SP queue with a next higher priority be sent. Among SP and WFQ queues, only after the packets in all SP queues are sent, can the packets in WFQ queues be sent.

6. Congestion Avoidance

Congestion avoidance monitors the usage of the outbound interface queues, and eliminates the network overload by actively discarding packets and adjusting the network traffic when network congestion occurs. Congestion avoidance avoids congestion by effectively monitoring the network traffic load, forecasting the occurrence of congestion, and discarding packets. Discarding policies include tail-drop, discarding based on random early detection (RED) results, and discarding based on weighted random early detection (WRED) results.

- Tail-drop

  Traditional packet loss policies use the tail-drop method. Tail-drop is effective for all traffic and cannot distinguish service levels. When congestion occurs, data packets at the tail of a queue are discarded until the congestion is removed.

- RED and WRED

The hosts running Transmission Control Protocol (TCP) decreases the rate of sending packets to respond to massive packet loss. After congestion is removed, the hosts increase the rate of sending packets. Tail-drop may cause TCP global synchronization. Namely, when a queue discards multiple TCP packets simultaneously, multiple TCP connections enter the congestion avoidance and slow startup state simultaneously. In this case, the tail is dropped, and the traffic is reduced and adjusted. When congestion is removed, traffic peaks appear. The process repeats constantly, the network traffic goes up and down suddenly, and the line traffic always fluctuates between the lowest traffic and the full traffic. When TCP global synchronization occurs, the connection bandwidth cannot be fully used, which causes a bandwidth waste.

To avoid this circumstance, you can use the RED or WRED packet discarding policy. The policy provides a mechanism for discarding packets at random and avoiding TCP global synchronization. When packets of a TCP connection are discarded and the remaining packets of the TCP connection are sent at a lower rate, packets of other TCP connections are still sent at higher rates. There are always some TCP connections whose packets are sent at higher rates, which increases the utilization of line bandwidth.

When WRED is used, you can set the lower threshold and maximum discarding probability for a queue. When the length of a queue is smaller than the lower threshold, no packets are discarded. When the queue length is between the lower threshold and higher threshold, WRED starts to discard packets randomly and sets a maximum discarding probability. A longer queue leads to a higher discarding probability. When the length of a queue is greater the higher threshold, packets are discarded at the maximum discarding probability.

Different from RED, WRED uses priorities to distinguish discarding policies. RED is a special case of WRED. When all the CoS values of an interface are mapped to the same lower and higher thresholds, WRED becomes RED.

1.2 Configuration Task Summary

QoS configuration includes the following tasks:

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring Traffic Classification
- Configuring Priority Mapping
- Configuring the Interface Rate Limit
- Configuring Congestion Management
- Configuring Congestion Avoidance
- Configuring the QoS Global Policy
- Disabling the Packet Priority Change
- Configuring the Queue Display and Configuration Mode

1.3 Configuring Traffic Classification

1.3.1 Overview

Traffic classification uses certain rules to identify the packets with certain characteristics. You can define the binding between multiple traffic classes and traffic behaviors to form policies and apply the policies to interfaces to realize traffic classification and processing.

1.3.2 Configuration Tasks

Traffic classification configuration includes the following tasks:

(1) Configuring a Class

(2) Configuring a Policy

(3) Applying a Policy

(4) (Optional) Configuring a Logical Interface Group

1.3.3 Configuring a Class

1. Overview

This section describes how to create a class, and defines class matching rules in class configuration mode.

2. Restrictions and Guidelines

- A class name comprises a maximum of 31 characters.
- Run the **match-rule** command to define a class matching rule.
- When setting the matching rule of a class to ACL matching, you need to create an ACL and ACL rules first.
- The new rule will overwrite the previous one and the device will only keep the latest rule. A notification will be prompted when a new rule is configured

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a class and enter the class configuration mode.

cmap-create *class-map-name*

No class is configured by default.

(4) Define a matching rule for the class. Configure at least one of the tasks.

- Set the matching rule of the class to ACL matching.

  match-rule access-apply { *acl-name* | *acl-number* }

  Numerically indexed ACLs or named ACLs are supported.

- Set the matching rule of the class to matching the PRE priorities of IP packets.

  match-rule ip precedence *precedence-value-list*

  Multiple IP PRE values can be matched at the same time. The value range of IP PRE is from 0 to 7.

- Set the matching rule of the class to matching the DSCP priorities of IP packets.

  match-rule ip dscp *dscp-value-list*

  Multiple DSCP values can be matched at the same time. The value range of DSCP priority is from 0 to 63.

No class matching rules are defined by default.

1.3.4 Configuring a Policy

1. Overview

This section describes how to create a policy, and bind classes and traffic behaviors in the policy configuration mode.

2. Restrictions and Guidelines

- A policy name comprises a maximum of 31 characters.
- When multiple classes are associated with the same policy, you are not advised to match the same flow with multiple classes; otherwise the traffic behavior bound to a class is performed on the flow randomly, and, when the device restarts, the traffic behavior bound to a class is also performed on the flow randomly.
- The **set** command is used to modify the priority tag of a specified packet, such as IEEE 802.1p and DSCP.
- The **police** command is used to limit the bandwidth of the specified traffic and configure the processing action on traffic beyond the limit, for example, discarding the traffic out of the limit, or modifying the CoS or DSCP value.
- The **drop** command is used to drop all the identified streams. To configure other traffic behaviors, you must delete this traffic behavior.
- The **unset police** or **unset apply acl** command may fail to take effect for the applied QoS flow policy configuration due to insufficient FP resources. The configuration will not rolled back automatically. In this case, you need to restore the **no** action manually. You are advised to cancel the application before changing the action and then apply the application again after.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a policy and enter the policy configuration mode.

policy-create *policy-map-name*

No policy is configured by default.

(4) Associate a class and enter the policy class configuration mode.

cmap *class-map-name*

No class is associated with a policy by default.

(5) Bind the class to a traffic behavior. Configure at least one of the tasks.

- Bind the class to the traffic behavior of modifying the IEEE 802.1p,and DSCP values.

  set { **cos** *new-cos* [**priority**] | **ip dscp** *new-dscp* [**with cos** *new-cos* **priority**] }

- Bind the class to the traffic behavior of limiting the bandwidth and processing packets beyond the limit.

  police *rate-bps burst-byte* [**exceed**-**action** { **cos** *new-cos* [**none**-**tos**] | **drop** | **dscp** *new-dscp* }]

- Bind the class to the traffic behavior of discarding packets.

  drop

A class is not bound to any traffic behavior by default.

This command can be configured only when a class associated with a policy is not bound to any traffic behavior. Namely, after a class associated with a policy is bound to the traffic behavior of discarding packets, you need to first delete the binding and then bind the class to other traffic behaviors.

(6) Enable ACL filtering to filter out packets matching the deny rule.

apply acl

(7) Enable policy statistics.

statistics enable

1.3.5 Applying a Policy

1. Overview

This section describes how to apply a configured policy to a specified interface to make the policy take effect.

2. Restrictions and Guidelines

- A policy can be applied globally or to an interface.
- If a policy is applied globally, the policy is applied to all interfaces which can be configured with policies.
- Configuring the **input** option applies a policy to the input direction of an interface.
- Configuring the **output** option applies a policy to the output direction of an interface.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

- Apply a policy in global configuration mode.

  apply-policy { **input** | **output** } *policy-map-name*

  No policy is applied globally by default.

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

(4) Apply a policy on the interface.

apply-policy { **input** | **output** } *policy-map-name*

No policy is applied to an interface by default.

1.3.6 Configuring a Logical Interface Group

1. Overview

This section describes how to associate a configured policy with a logical interface group.

2. Restrictions and Guidelines

- A maximum of 128 logical interface groups can be created.
- In global configuration mode, run the **virtual-group** command to create a logical interface group and enter the logical interface group configuration mode.
- In interface configuration mode, run the **virtual-group** command to add the interface to a specified logical interface group. If the logical interface group is not created, this command creates the logical interface group and adds the interface to the group.
- The following types of interfaces cannot be added to a logical interface group:
  - Member interface of an aggregation port
  - VSL member interface

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

- Create a logical interface group

  virtual-group *virtual-group-number*

  No logical interface group is defined by default.

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport interface *-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface ethernet-type interface-number

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

(4) Add the interface to a logical interface group.

virtual-group *virtual-group-number*

An interface is not added to any logical interface group by default.

(5) Return to the global configuration mode.

exit

(6) Enter the logical interface group configuration mode.

virtual-group *virtual-group-number*

(7) Apply a policy to the logical interface group.

apply-policy { **input** | **output** } *policy-map-name*

1.4 Configuring Priority Mapping

1.4.1 Overview

Priority mapping can be implemented based on the trust mode configured for an interface.

1.4.2 Configuration Tasks

Priority marking and mapping configuration includes the following tasks:

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the Trust Mode
- Configuring Mappings

1.4.3 Configuring the Trust Mode

1. Overview

- You can set various priorities for packets according to the trust mode configured for an interface.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Ethernet interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface *ethernet-type interface-number*

(4) Configure the trust mode for the interface. Please configure only one task.

- Configure trusting IEEE 802.1p.

  mls qos trust cos

- Configure trusting DSCP.

  mls qos trust dscp

- Configure trusting IP PRE.

mls qos trust ip-precedence

An interface is in untrusted mode by default.

(5) (Optional) Configure the IEEE 802.1p value for the interface.

mls qos cos *cos-value*

The default IEEE 802.1p value of an interface is **0**.

1.4.4 Configuring Mappings

1. Overview

This section describes how to configure mappings.

2. Restrictions and Guidelines

- Run the **mls qos map cos-dscp** command to configure the mappings from IEEE 802.1p values to DSCP values.
- Run the **mls qos map dscp-cos** command to configure the mappings from DSCP values to CoS values.
- Run the **mls qos map ip-precedence-dscp** command to configure the mappings from IP PRE values to DSCP values.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure mappings. Configure at least one of the tasks.

- Configure the IEEE 802.1p-to-DSCP mappings.

  mls qos map cos-dscp *dscp-value-list*

  The IEEE 802.1p values 0 to 7 are mapped to the DSCP values 0, 8, 16, 24, 32, 40, 48, and 56 respectively by default.

- Configure the DSCP-to-CoS mappings.

  mls qos map dscp-cos *dscp-value*&<1-8> **to** *cos-value*

  DSCP values 0 to 7 are mapped to CoS 0, DSCP values 8 to 15 are mapped to CoS 1, DSCP values 16 to 23 are mapped to CoS 2, DSCP values 24 to 31 are mapped to CoS 3, DSCP values 32 to 39 are mapped to CoS 4, DSCP values 40 to 47 are mapped to CoS 5, DSCP values 48 to 55 are mapped to CoS 6, and DSCP values 56 to 63 are mapped to CoS 7 by default.

- Configure the IP PRE-to-DSCP mappings.

  mls qos map ip-precedence-dscp *dscp-value-list*

  The IP PRE values 0 to 7 are mapped to the DSCP values 0, 8, 16, 24, 32, 40, 48, and 56 respectively by default.

1.5 Configuring the Interface Rate Limit

1.5.1 Overview

This section describes how to configure traffic limit for an interface.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Ethernet interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface aggregateport *interface-number*

(4) Configure traffic limit for the interface. Configure at least one of the tasks.

- Configure the rate limit for the traffic in the input direction of an interface.

  speed-limit input *rate-value burst-value*

- Configure the rate limit for the traffic in the output direction of an interface.

  speed-limit output *rate-value burst-value*

No traffic limit is configured for an interface by default.

1.6 Configuring Congestion Management

1.6.1 Overview

This section describes how to control congestion. The priorities of important data packets can be raised so that they are sent first when congestion occurs, thereby ensuring that key services are provided in time.

1.6.2 Configuration Tasks

The congestion management configuration includes the following tasks:

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the CoS-to-Queue Mappings
- Configuring the Scheduling Policy and Round Robin Weights for Output Queues
- Configuring Bandwidth for a Queue

- Enabling ECN for a Queue
- Configuring Cumulative ECN Counter
- Configuring the ECT Mark Type

1.6.3 Notes

Table 1-2 Products Supporting Congestion Management Functions

Function	Product
Configuring Cumulative ECN Counter	Supported by the S6580-48CQ8QC and S6580-24DC8QC
Configuring the ECT Mark Type	Supported by the S6580-48CQ8QC and S6580-24DC8QC

1.6.4 Configuring the CoS-to-Queue Mappings

1. Overview

This section describes how to configure the CoS-to-queue mappings so that packets can enter various output queues based on their CoS values.

2. Restrictions and Guidelines

- Run the **priority-queue cos-map** command to configure the CoS-to-queue mappings.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the CoS-to-queue mappings.

priority-queue cos-map *qid cos-value* &<1-8>

The CoS values 0 to 7 are mapped to the queues 1 to 8 respectively by default.

1.6.5 Configuring the Scheduling Policy and Round Robin Weights for Output Queues

1. Overview

This section describes how to configure the scheduling policy and round robin weights for output queues.

2. Restrictions and Guidelines

- Run the **priority-queue** command to set the scheduling policy to SP scheduling.
- Run the **mls qos scheduler sp** command to set the scheduling policy to SP scheduling for output queues.
- Run the **mls qos scheduler rr** command to set the scheduling policy to RR scheduling for output queues.

- Run the **mls qos scheduler wrr** command to set the scheduling policy to WRR scheduling based on the packet quantity for output queues.
- Run the **mls qos scheduler drr** command to set the scheduling policy to WRR scheduling based on the packet size for output queues.
- Run the **mls qos scheduler wfq** command to set the scheduling policy to WFQ scheduling for output queues.
- Run the **drr-queue bandwidth** command to configure the round robin weight for the DRR scheduling policy of output queues. A higher weight means that more packet bytes can be sent.
- Run the **wrr-qos-queue bandwidth** command to configure the round robin weight for the WRR scheduling policy of output queues. A higher weight means longer output time. The configurable weight range depends on the product.
- Run the **wfq-queue bandwidth** command to configure the round robin weight for the WFQ scheduling policy of output queues. A higher weight means that more packet bytes can be sent.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the scheduling policy for output queues. Please configure only one task.

- Set the scheduling policy to SP scheduling.

  priority-queue

- Configure the scheduling policy for global output queues.

  mls qos scheduler { **drr** | **rr** | **sp** | **wfq** | **wrr** }

  The default scheduling policy of global output queues is **WRR**.

- Configure the scheduling policy for interface output queues. Run the following commands in turn to configure the scheduling policy for interface output queues.

  interface *interface-type interface-number*

  mls qos scheduler { **drr** | **rr** | **sp** | **wfq** | **wrr** }

  No scheduling policy is configured on an interface by default.

(4) Configure the round robin weight for the scheduling policy of output queues. Please configure only one task.

- Configure the round robin weight for the scheduling policy of global output queues.

  { **drr-queue** | **wfq-queue** | **wrr-qos-queue** } **bandwidth** *weight-value-list*

  The default round robin weight ratio of global queues is **1:1:1:1:1:1:1:1**.

- Configure the round robin weight for the scheduling policy of interface output queues. Run the following commands in turn to configure the round robin weight for the scheduling policy of interface output queues.

  interface *interface-type interface-number*

  { **drr-queue** | **wfq-queue** | **wrr-qos-queue** } **bandwidth** *weight-value-list*

  The default round robin weight ratio of interface queues is **1:1:1:1:1:1:1:1**.

1.6.6 Configuring Bandwidth for a Queue

1. Overview

This section describes how to configure the minimum guaranteed bandwidth and maximum limited bandwidth for a queue.

2. Restrictions and Guidelines

- This command can be configured globally or on an interface, depending on the product.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the L2 or L3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

(4) Configure the minimum guaranteed bandwidth and maximum limited bandwidth for a queue.

qos queue *queue-id* **bandwidth** { **maximum** *bandwidth* | **minimum** *bandwidth* }

No minimum guaranteed bandwidth or maximum limited bandwidth is configured for a queue by default.

1.6.7 Enabling ECN for a Queue

1. Overview

With combination of the WRED technology and explicit congestion notification (ECN) technology, a terminal can perceive network congestion and notify the source device of the congestion. Upon receiving the notification, the source device reduces the packet sending rate to prevent congestion aggravation.

2. Restrictions and Guidelines

- Both the ECN and WRED functions must be configured before the ECN function is enabled for a queue.
- After the ECN function is enabled for a queue (namely, the ECN field in packets is set to **01** or **10**), when queue congestion occurs, the device changes the ECN field to **CE** (**11**), instead of discarding packets randomly according to the WRED template.
- After the ECN function is enabled for a queue, the device sets the ECN field to **CE** only when the packet rate reaches the lower discarding threshold specified in the WRED template.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the global WRED function.

queueing wred

The WRED function is disabled by default.

(4) (Optional) Configure a sampling interval for collecting statistics on the packets discarded by WRED and packets marked by ECN.

qos wred-ecn statistics sample-period *period-value*

A sampling interval for collecting statistics is configured by default.

(5) Enter the Layer 2 or Layer 3 Ethernet interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface *ethernet-type interface-number*

(6) Enable the ECN function for a queue.

qos queue *queue-id* **ecn**

The ECN function is disabled for an interface queue by default.

1.6.8 Configuring Cumulative ECN Counter

Note

This function is supported by only the S6580-48CQ8QC, and S6580-24DC8QC.

1. Overview

The device counts new ECN-marked packets by default. Some products support cumulative ECN-marked packet counting.

2. Restrictions and Guidelines

- ECN statistics will be cleared when the counter mode is changed.
- If you run the **unset qos queue** *queue-id* **ecn** command to disable the ECN function for a queue, the interface still counts the cumulative number of ECN-marked packets. Run the **display qos wred-ecn statistics** command to display the cumulative number.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Configure cumulative ECN counter.
qos wred-ecn statistics mode ecn-total

Cumulative ECN counter is not configured by default.

1.6.9 Configuring the ECT Mark Type

Note

This function is supported by only the S6580-48CQ8QC and S6580-24DC8QC.

1. Overview

The chips need to be specified with the ECN Capable Transport (ECT) mark value for packet counting. Parameters **ect0** and **ect1** correspond to value **10** and **01** respectively. If not specified, the parameter is subject to the chip.

2. Restrictions and Guidelines

N/A

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Configure ECN Capable Transport (ECT) mark type.
qos ecn remark type { **ect0** | **ect1** }

No ECT mark type is not configured by default.

1.7 Configuring Congestion Avoidance

1.7.1 Overview

Congestion avoidance monitors the usage of queues in an outbound interface and eliminates the network overload by actively discarding packets and adjusting the network traffic in the case of network congestion.

1.7.2 Configuration Tasks

Congestion avoidance configuration includes the following tasks:

(1) Enabling the WRED Function

(2) (Optional) Configuring Higher and Lower Threshold Values

(3) (Optional) Configuring the Sampling Interval

1. Overview

You can configure the sampling interval for the WRED discarding function to affect the calculation result of the average queue length.

2. Restrictions and Guidelines

- When packets are forwarded, WRED determines the discarding probability based on the average queue length of the egress queue. When the queue length is between the lower threshold and the higher threshold, WRED starts to discard packets randomly. A longer queue leads to a higher discarding probability but the probability cannot be higher than the maximum discarding probability. When the queue length is greater than the higher threshold, packets are discarded at the maximum discarding probability.

- To make the sampling interval effective, you need to first configure the WRED discarding function on an interface. Namely, the sampling interval can be configured only after the higher threshold and lower threshold for WRED to discard packets are configured on an interface.
- The sampling interval ranges from 8 μs to 512 μs. The sampling interval can affect the calculation result of the average queue length.
- Use this command together with the sampling weight configuration command to precisely adjust the WRED discarding function.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Configure the sampling interval.
wrr-queue random-detect sample-period *period-value*

No sampling interval is configured by default.

(4) Configuring the Sampling Weight

(5) (Optional) Configuring the CoS-to-Threshold Mappings

1.7.3 Enabling the WRED Function

1. Overview

This section describes how to configure WRED as the packet discarding policy.

2. Restrictions and Guidelines

The default packet discarding policy is tail-drop.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the WRED function.

queueing wred

The WRED function is disabled by default.

1.7.4 Configuring Higher and Lower Threshold Values

1. Overview

This section describes how to configure the higher and lower threshold values for WRED.

2. Restrictions and Guidelines

- When the packet length of a queue is smaller than the lower threshold, no packets are discarded. When the queue length is between the lower threshold and the higher threshold, WRED starts to discard packets randomly.
- Because the maximum value of the configuration range is equal to the current higher threshold, you need to pay attention to the configured higher threshold when configuring the lower threshold.
- The threshold for WRED to discard packets can be configured in the unit of percentage or cell. The final configuration prevails.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Ethernet interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface *ethernet-type interface-number*

(4) Configure a lower threshold value. Please configure only one task.

- Configure a lower threshold value in percentage.

  wrr-qos-queue random-detect min-threshold *queue-id threshold*&<1-2>

  The queue value range is from 1 to 8. The value range of the lower threshold value is from 1 to the set value of the higher threshold.

- Configure a lower threshold value in the unit of cell.

  wrr-qos-queue random-detect min-threshold cell *queue-id cell-value*&<1-2>

  The queue value range is from 1 to 8. The value range of the lower threshold is from 1 to the set value of the higher threshold. The default threshold value depends on the product.

(5) The threshold value is configured in percentage by default. Configure a higher threshold value. Please configure only one task.

- Configure a higher threshold value in percentage.

  wrr-qos-queue random-detect max-threshold *queue-id threshold*&<1-2>

  The queue value range is from 1 to 8. The value range of the higher threshold is from the set value of the lower threshold to the maximum value allowed for the higher threshold value.

- Configure a higher threshold value in the unit of cell.

wrr-qos-queue random-detect max-threshold cell *queue-id cell-value*&<1-2>

The queue value range is from 1 to 8. The value range of the higher threshold is from the set value of the lower threshold to the maximum value allowed for the higher threshold value.

The threshold value is configured in percentage by default.

1.7.5 Configuring the Maximum Discarding Probability

1. Overview

When the packet length in a queue is between the lower threshold and the higher threshold, you can configure the maximum probability for discarding packets at random.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 or Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

(4) Configure the maximum discarding probability.

wrr-qos-queue random-detect probability *queue-id probability*&<1-2>

The maximum discarding probability is not configured by default.

1.7.6 Configuring the Sampling Interval

4. Overview

You can configure the sampling interval for the WRED discarding function to affect the calculation result of the average queue length.

5. Restrictions and Guidelines

- When packets are forwarded, WRED determines the discarding probability based on the average queue length of the egress queue. When the queue length is between the lower threshold and the higher threshold, WRED starts to discard packets randomly. A longer queue leads to a higher discarding probability but the probability cannot be higher than the maximum discarding probability. When the queue length is greater than the higher threshold, packets are discarded at the maximum discarding probability.
- To make the sampling interval effective, you need to first configure the WRED discarding function on an interface. Namely, the sampling interval can be configured only after the higher threshold and lower threshold for WRED to discard packets are configured on an interface.
- The sampling interval ranges from 8 μs to 512 μs. The sampling interval can affect the calculation result of the average queue length.
- Use this command together with the sampling weight configuration command to precisely adjust the WRED discarding function.

6. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Configure the sampling interval.
wrr-queue random-detect sample-period *period-value*

No sampling interval is configured by default.

1.7.7 Configuring the Sampling Weight

1. Overview

You can configure the sampling weight for the WRED discarding function to affect the calculation result of the average queue length.

2. Restrictions and Guidelines

- When packets are forwarded, WRED determines the discarding probability based on the average queue length of the egress queue. When the queue length is between the lower threshold and the higher threshold, WRED starts to discard packets randomly. A longer queue leads to a higher discarding probability but the probability cannot be higher than the maximum discarding probability. When the queue length is greater than the higher threshold, packets are discarded at the maximum discarding probability.
- The sampling weight indicates the weight factor of sampled data updates. A larger sampling weight indicates a longer update interval of the average queue length and a larger average queue length.
- To make the sampling weight effective, you need to first configure the WRED discarding function on an interface. Namely, the sampling weight must be configured after the higher threshold and lower threshold are configured for the WRED discarding function on an interface.
- Use this command together with the sampling interval configuration command to precisely adjust the WRED discarding function.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Ethernet interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

  interface *ethernet-type interface-number*

(4) Configure the sampling weight.

wrr-qos-queue random-detect sample-weight *queue-id weight*-value

No sampling weight is configured by default.

1.7.8 Configuring the CoS-to-Threshold Mappings

1. Overview

This section describes how to configure the CoS-to-threshold mappings.

2. Restrictions and Guidelines

- Multiple threshold groups can be configured for the lower threshold and maximum discarding probability.
- By configuring the mappings from CoS values and threshold groups, you can select the effective threshold group mapped to a CoS, for example, you can map CoS 0 to threshold group 1, and CoS 1 to threshold group 2.
- If the packets of CoS values 0 and 1 are all added to queue 1 for scheduling, the packets of CoS 0 are processed based on the lower threshold and maximum discarding probability in group 1, and the packets of CoS 1 are processed based on the lower threshold and maximum discarding probability in group 2.
- When all the CoS values are mapped to the same threshold group, the enabled WRED becomes RED.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the CoS-to-threshold mappings.

wrr-qos-queue cos-map *threshold-id cos*&<1-8>

All the CoS values are mapped to the threshold in group 1 by default.

1.8 Configuring the QoS Global Policy

Note

This function is supported by only theS6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1.8.1 Overview

Enabling the QoS global policy enables some QoS policies by default, for example, changing the DSCP value of L2 flows.

1.8.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the global QoS policy.

mls qos enable

1.9 Disabling the Packet Priority Change

1.9.1 Overview

When a packet is sent out of a device, the priority of the packet is changed to the priority configured on the device. You can disable packet priority change, that is, the priority of a packet sent out of a device is the same as that when the packet is sent to the device.

1.9.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable packet priority change.

mls qos remark disable

1.10 Configuring the Queue Display and Configuration Mode

1.10.1 Overview

After the queue display and configuration mode is configured, the display range of all queues on the device becomes 0-7 or 1-8.

1.10.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the queue display and configuration mode.

qos-queue compatible { enable | disable }

1.11 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may interrupt services due to loss of important information.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-3 QoS Monitoring

Command	Purpose
display cmap-create [*class-map-name*]	Displays the traffic classification information.
display policy-create [*policy-map-name* [**cmap** *class-map-name*]]	Displays the QoS policy information.
display policy-create interface *interface-type interface-number*	Displays information about a policy applied to an interface.
display policy-create statistics { **interfaces** *interface-type interface-number* **| global** } [**cmap** *class-map-name*] { **input** | **output** }	Displays policy statistics.
display virtual-group { *virtual-group-number* | **summary** }	Displays information about a logical interface group.
display mls qos virtual-group { *virtual-group-number* | **policers** }	Displays information about a policy applied to a logical interface group.
display mls qos maps { **cos-dscp** | **dscp-cos** | **ip-prec-dscp** }	Displays various mappings.
display mls qos speed-limit [**interface** *interface-type interface-number*]	Displays the rate limit information of an interface.
display mls qos queueing [**interface** *interface-type interface-number*]	Displays information about the QoS queues, scheduling policies, and round robin weights.
display mls qos scheduler [**interface** *interface-type interface-number*]	Displays the scheduling policy information of an output queue.
display mls qos develop { status | **capacity** | **mstat** | **features | istat | intf_status }**	Displays QoS-related internal information.
display queueing wred interface *interface-type interface-number*	Displays the WRED configuration.
display mls qos interface { *interface-type interface-number* | **policers** }	Displays the QoS information of an interface.
display qos bandwidth [**interface** *interface-type interface-number*]	Displays information about the queue bandwidth.

Command	Purpose
display qos res [**dev** *dev-number* [**slot** *slot-number*]]	Displays QoS resource usage of the current device.
display qos wred-ecn statistics [**interface** *interface-type interface-number*]	Displays statistics on the packets discarded by WRED and the packets marked by ECN.
cleanup qos res	Clears historical peak values of QoS resource usage.
cleanup qos wred-ecn statistics	Clears statistics on packets discarded by WRED and packets marked by explicit congestion notification (ECN).
cleanup policy-create statistics [{ **interfaces** *interface-type interface-number* **| global** } [**cmap** *class-map-name*]] { **input** | **output** }	Clears policy statistics.
debugging qos lib [**event** | **message**]	Debugs the QoS library.
debugging qos server [**event** | **message**]	Debugs the QoS communication server.
debugging qos mls	Debugs QoS user command processing.
debugging qos vmsup	Debugs VMSUP configurations.

1.12 Configuration Examples

1.12.1 Configuring the Interface Rate Limit and Priority Re-marking

1. Requirements

To meet the service requirements of normal teaching, a school puts forwards the following requirements:

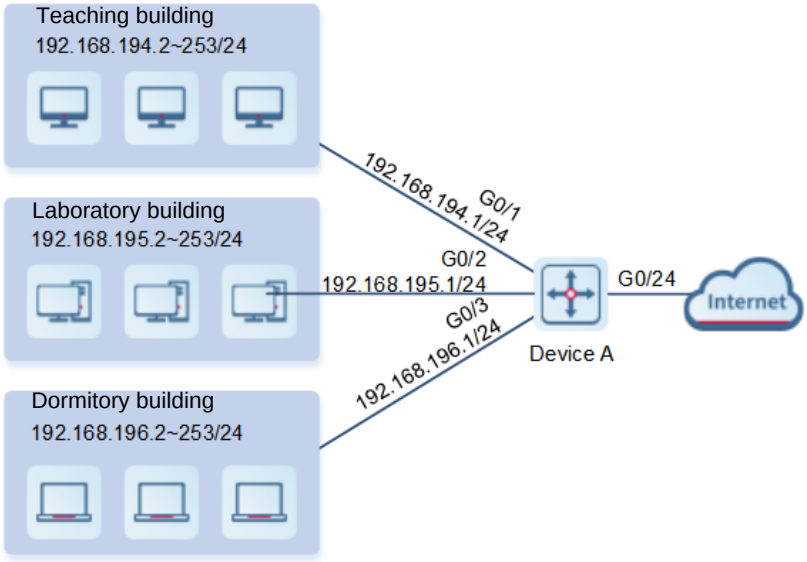
- Controlling the Internet access traffic of the school within 100 Mbps and discarding the packets out of the limit.
- Controlling the egress traffic of the dormitory building within 50 Mbps and discarding the packets out of limit.
- Controlling the rate of packets with DSCP priority 7 sent from laboratory building within 20 Mbps and discarding the packets out of limit..
- Controlling the egress traffic of the teaching building within 30 Mbps and discarding the packets out of limit.

Note

The school connects to the Internet through G0/24 of device A, and G0/1, G0/2, and G0/3 of device A connect to the teaching building (192.168.194.2~253/24), laboratory building (192.168.195.2~253/24), and dormitory building (192.168.196.2~253/24) respectively.

2. Topology

Figure 1-7 Configuring the Interface Rate Limit and Priority Re-marking



3. Notes

(1) For Internet access traffic on the outbound interface, configure the egress traffic limit on interface G0/24 of device A, and set the bandwidth limit to 100 Mbps and burst traffic limit to 256 Kbps.

(2) For the dormitory building, configure the ingress traffic limit on interface G0/3 of device A, and set the bandwidth limit to 50 Mbps and burst traffic limit to 256 Kbps.

(3) For the teaching building, configure the ingress traffic limit on interface G0/1 of device A, and set the bandwidth limit to 50 Mbps and burst traffic limit to 256 Kbps.

(4) For the laboratory building, create the class **cmap_dscp7** to match DSCP priority 7, create the policy **pmap_shiyan**, associate **cmap_dscp7** with the policy, bind the traffic behavior and set the bandwidth limit to 20 Mbps, apply **pmap_shiyan** to interface G0/2, and configure the interface to trust DSCP.

4. Procedure

(1) Configure IP addresses for interfaces (omitted).

(2) Configure rate limits for the interfaces.

For Internet access traffic on the outbound interface, configure the egress traffic limit on interface G0/24.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/24
DeviceA(cfg-if-GigabitEthernet 0/24# speed-limit output 102400 256
DeviceA(cfg-if-GigabitEthernet 0/24)# exit
```

For the dormitory building, configure the ingress traffic limit on interface G0/3.

```
DeviceA(cfg)# interface gigabitethernet 0/3
DeviceA(cfg-if-GigabitEthernet 0/3# speed-limit input 51200 256
DeviceA(cfg-if-GigabitEthernet 0/3)# exit
```

For the teaching building, configure the ingress traffic limit on interface G0/1.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1# speed-limit input 30720 256
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

(3) Configure traffic classification.

For the laboratory building, configure a class, traffic behavior, and policy.

```
DeviceA(cfg)# cmap-create cmap_dscp7
DeviceA(cfg-cmap)# match-rule ip dscp 7
DeviceA(cfg-cmap)# exit
DeviceA(cfg)# policy-create pmap_shiyan
DeviceA(cfg-policymap)# cmap cmap_dscp7
DeviceA(cfg-pmap-c)# police 20480 128 exceed-action drop
DeviceA(cfg-pmap-c)# exit
DeviceA(cfg-policymap)# exit
```

Apply the policy to an interface and configure the interface to trust DSCP.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2# apply-policy input pmap_shiyan
DeviceA(cfg-if-GigabitEthernet 0/2)# mls qos trust dscp
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

5. Verification

Check whether the interface rate limits are successfully configured.

```
DeviceA# display mls qos speed-limit
Interface: GigabitEthernet 0/1
   rate limit input Kbps = 30720 burst = 256
Interface: GigabitEthernet 0/3
   rate limit input Kbps = 51200 burst = 256
Interface: GigabitEthernet 0/24
   rate limit output Kbps = 102400 burst = 256
```

Check whether the class is successfully created.

```
DeviceA# display cmap-create cmap_dscp7

 Class Map cmap_dscp7
    Match ip dscp 7
```

Check whether the policy is successfully created.

```
DeviceA# display policy-create pmap_shiyan

 Policy Map pmap_shiyan
     Class cmap_dscp7
        police 20480 128 exceed-action drop
```

Check whether the policy is applied to the interface successfully.

```
DeviceA# display mls qos interface gigabitethernet 0/2
Interface: GigabitEthernet 0/2
```

```
Ratelimit input:
Ratelimit output:
Attached input    policy-map: pmap_shiyan
Attached output policy-create:
Default trust: dscp
Default cos: 0
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
cmap-create cmap_dscp7
  match-rule ip dscp 7
!
policy-create pmap_shiyan
  cmap cmap_dscp7
    police 20480 128 exceed-action drop
!
interface GigabitEthernet 0/1
  ip address 192.168.194.1 255.255.255.0
  speed-limit input 30720 256
!
interface GigabitEthernet 0/2
  ip address 192.168.195.1 255.255.255.0
  apply-policy input pmap_shiyan
  mls qos trust dscp
!
interface GigabitEthernet 0/3
  ip address 192.168.196.1 255.255.255.0
  speed-limit input 51200 256
!
interface GigabitEthernet 0/24
  speed-limit output 102400 256
!
```

1.12.2 Configuring Queue Scheduling and Congestion Avoidance

1. Requirements

Priority re-marking and queue scheduling need to be configured to meet the following requirements:

- When the R&D department and market department access the servers, the priorities of the server packets are as follows: packets destined for the mail server > packets destined for the file server > packets destined for the salary query server.
- No matter when the HR management department accesses the Internet or servers, device A processes the corresponding packets with the highest priority.
- Network congestion often occurs during operation of device A. WRR queue scheduling must be used to schedule IP packets from the R&D and market departments to access the mail database, file database, and

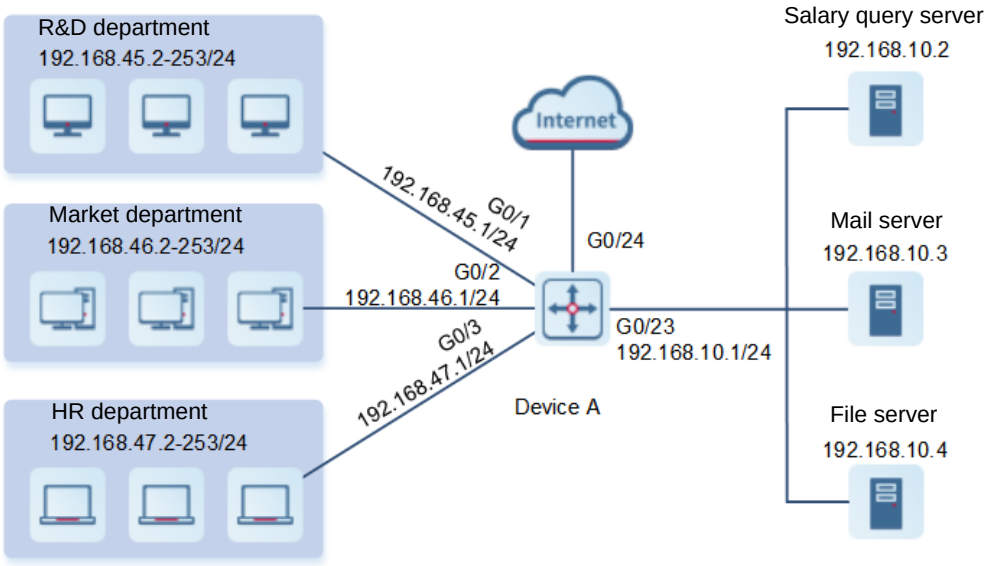
salary query database based on the ratio of 6:2:1.

Note

The R&D department (192.168.45.2~253/24), market department (192.168.46.2~253/24), and HR department (192.168.47.2~253/24) are connected to interfaces G0/1, G0/2, and G0/3 of device A respectively; the salary query server (192.168.10.2/24), mail server (192.168.10.3/24), and file server (192.168.10.4/24) are connected to interface G0/23 of device A.

2. Topology

Figure 1-8 Configuring Queue Scheduling and Congestion Avoidance



3. Notes

(1) On device A, create ACLs to filter packets used for accessing servers, create classes, and associate the classes with these ACLs.

(2) Create policies to associate with the classes and specify the IEEE 802.1p values for the packets accessing various servers in the policies: Set the IEEE 802.1p value to 4 for the packets accessing the mail server, to 3 for the packets accessing the file server, and to 2 for the packets accessing the salary query server. Apply the policies to the inbound interfaces connected to the R&D and market departments and configure the interfaces to trust IEEE 802.1p.

(3) Assign the highest priority 7 to the IEEE 802.1p value of the interface connected to the HR department, to ensure that packets from the HR department are sent with the highest priority.

(4) Set the scheduling policy to WRR for output queues and set the round robin weight to 1:1:1:2:6:1:1:0 for the queues. This means that SP scheduling is used for packets of the HR department, and the packets for accessing the mail database, file database, and salary query database from the R&D and market departments are scheduled based on the ratio of 6:2:1.

(5) Enable the WRED function.

(6) Set the lower thresholds to 10 and 20 for queue 2 of interface G0/2.

(7) Set the higher thresholds to 60 and 90 for queue 2 of interface G0/2.

(8) Set the maximum discarding probabilities to 60 and 80 for queue 2 of interface G0/2.

4. Procedure

(1) On device A, configure ACLs to filter packets used for accessing the servers and add ACL rules.

Configure the ACL for matching packets used for accessing the salary query server as well as ACL rules.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip access-control-list extended salary
DeviceA(cfg-extacl)# permit ip any host 192.168.10.2
DeviceA(cfg-extacl)# exit
```

Configure the ACL for matching packets used for accessing the mail server as well as ACL rules.

```
DeviceA(cfg)# ip access-control-list extended mail
DeviceA(cfg-extacl)# permit ip any host 192.168.10.3
DeviceA(cfg-extacl)# exit
```

Configure the ACL for matching packets used for accessing the file server as well as ACL rules.

```
DeviceA(cfg)# ip access-control-list extended file
DeviceA(cfg-extacl)# permit ip any host 192.168.10.4
DeviceA(cfg-extacl)# exit
```

(2) Configure classes on device A.

Associate a class named salary with the ACL rule for matching packets used for accessing the salary query server.

```
DeviceA(cfg)# cmap-create salary
DeviceA(cfg-cmap)# match-rule access-apply salary
DeviceA(cfg-cmap)# exit
```

Associate a class named mail with the ACL rule for matching packets used for accessing the mail server.

```
DeviceA(cfg)# cmap-create mail
DeviceA(cfg-cmap)# match-rule access-apply mail
DeviceA(cfg-cmap)# exit
```

Associate a class named file with the ACL rule for matching packets used for accessing the file server.

```
DeviceA(cfg)# cmap-create file
DeviceA(cfg-cmap)# match-rule access-apply file
DeviceA(cfg-cmap)# exit
```

(3) Configure policies on device A.

Configure the policy named toserver to access the server.

```
DeviceA(cfg)# policy-create toserver
DeviceA(cfg-policymap)#
```

Associate the class named salary with a traffic behavior.

```
DeviceA(cfg-policymap)# cmap salary
DeviceA(cfg-pmap-c)# set cos 2
DeviceA(cfg-pmap-c)# exit
```

Associate the class named mail with a traffic behavior.

```
DeviceA(cfg-policymap)# cmap mail
DeviceA(cfg-pmap-c)# set cos 4
DeviceA(cfg-pmap-c)# exit
```

Associate the class named file with a traffic behavior.

```
DeviceA(cfg-policymap)# cmap file
DeviceA(cfg-pmap-c)# set cos 3
DeviceA(cfg-pmap-c)# exit
DeviceA(cfg-map)# exit
```

(4) Apply policies on device A.

Apply a policy to the interface connected to the R&D department.

```
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# apply-policy input toserver
DeviceA(cfg-if-GigabitEthernet 0/1)# mls qos trust cos
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Apply a policy to the interface connected to the market department.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# apply-policy input toserver
DeviceA(cfg-if-GigabitEthernet 0/2)# mls qos trust cos
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

(5) On device A, configure the 802.1p field for an interface.

Configure the highest priority for the 802.1p field for the interface connected to the HR department.

```
DeviceA(cfg)# interface gigabitethernet 0/3
DeviceA(cfg-if-GigabitEthernet 0/3)# mls qos cos 7
DeviceA(cfg-if-GigabitEthernet 0/3)# exit
```

(6) On device A, configure a scheduling policy for output queues.

Set the scheduling policy to WRR for output queues.

```
DeviceA(cfg)# wrr-qos-queue bandwidth 1 1 1 2 6 1 1 0
DeviceA(cfg)# mls qos scheduler wrr
```

(7) Configure congestion avoidance on device A.

Enable the WRED function.

```
DeviceA(cfg)# queueing wred
```

Configure the higher and lower threshold values and maximum discarding probability.

```
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# wrr-qos-queue random-detect min-threshold 2 10 20
DeviceA(cfg-if-GigabitEthernet 0/2)# wrr-qos-queue random-detect max-threshold 2 60 90
DeviceA(cfg-if-GigabitEthernet 0/2)# wrr-qos-queue random-detect probability 2 60 80
```

Configure the CoS-to-threshold mappings.

```
DeviceA(cfg)# wrr-qos-queue cos-map 2 0 1 2 3
DeviceA(cfg)# exit
```

5. Verification

Check whether ACLs are successfully created.

```
DeviceA# display access-lists

ip access-control-list extended file
 10 permit ip any host 192.168.10.4

ip access-control-list extended mail
 10 permit ip any host 192.168.10.3

ip access-control-list extended salary
 10 permit ip any host 192.168.10.2
```

Check whether the classes are successfully associated with the ACLs.

```
DeviceA# display cmap-create

 Class Map salary
    Match access-apply salary
 Class Map mail
    Match access-apply mail
 Class Map file
    Match access-apply file
```

Check whether policies are successfully created, and whether the classes and traffic behaviors are successfully bound.

```
DeviceA# display policy-create

 Policy Map toserver
     Class mail
        set cos 4
     Class file
        set cos 3
     Class salary
        set cos 2
```

Check whether a policy is successfully applied to interface GigabitEthernet 0/1.

```
DeviceA# display mls qos interface gigabitethernet 0/1
Interface: GigabitEthernet 0/1
Ratelimit input:
Ratelimit output:
Attached input    policy-map: toserver
Attached output policy-create:
Default trust: cos
Default cos: 0
```

Check whether a policy is successfully applied to interface GigabitEthernet 0/2.

```
DeviceA# display mls qos interface gigabitethernet 0/2
```

```
Interface: GigabitEthernet 0/2
Ratelimit input:
Ratelimit output:
Attached input    policy-map: toserver
Attached output policy-create:
Default trust: cos
Default cos: 0
```

Check whether a policy is successfully applied to interface GigabitEthernet 0/3.

```
DeviceA# display mls qos interface gigabitethernet 0/3
Interface: GigabitEthernet 0/3
Ratelimit input:
Ratelimit output:
Attached input    policy-create:
Attached output policy-create:
Default trust: none
Default cos: 7
```

Check whether the 802.1p value is successfully configured for the interface and whether the scheduling policy and the round robin weight are successfully configured.

```
DeviceA# display mls qos queueing
Cos-queue map:
cos qid
--- ---
0     1
1     2
2     3
3     4
4     5
5     6
6     7
7     8

wrr bandwidth weights:
qid weights
--- -------
1     1
2     1
3     1
4     2
5     6
6     1
7     1
8     0

drr bandwidth weights:
qid weights
```

```
--- -------
1    1
2    1
3    1
4    1
5    1
6    1
7    1
8    1

wfq bandwidth weights:
qid weights
--- -------
1    1
2    1
3    1
4    1
5    1
6    1
7    1
8    1
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
wrr-qos-queue bandwidth 1 1 1 2 6 1 1 0
queueing wred
wrr-qos-queue cos-map 2 0 1 2 3
!
cmap-create salary
 match-rule access-apply salary
cmap-create mail
 match-rule access-apply mail
cmap-create file
 match-rule access-apply file
!
policy-create toserver
 cmap salary
  set cos 2
 cmap mail
  set cos 4
 cmap file
  set cos 3
!
interface GigabitEthernet 0/1
```

```
 ip address 192.168.45.1 255.255.255.0
 apply-policy input toserver
 mls qos trust cos
!
interface GigabitEthernet 0/2
 ip address 192.168.46.1 255.255.255.0
 wrr-qos-queue random-detect min-threshold 2 10 20
 apply-policy input toserver
 mls qos trust cos
!
interface GigabitEthernet 0/3
 ip address 192.168.47.1 255.255.255.0
 mls qos cos 7
!
```

Contents

1 Configuring MMU 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 5
1.3 Configuring Buffer Adjustment 5
1.3.1 Overview 5
1.3.2 Restrictions and Guidelines 5
1.3.3 Configuration Tasks 5
1.3.4 Notes 6
1.3.5 Configuring a Guaranteed Buffer for a Queue 6
1.3.6 Configuring a Shared Buffer for a Queue 7
1.3.7 Configuring a Guaranteed Buffer for an Ingress PG 7
1.3.8 Configuring a Shared Buffer for an Ingress PG 8
1.3.9 Configuring a Flow Control Threshold for an Ingress PG 8
1.3.10 Configuring a Flow Control Threshold Offset for an Ingress PG 9
1.3.11 Configuring Headroom Buffer Space for an Ingress PG 9
1.3.12 Configuring the Global Headroom Buffer Threshold 10
1.3.13 Configuring Cut-Through 10
1.4 Configuring Status Monitoring 11
1.4.1 Overview 11
1.4.2 Restrictions and Guidelines 11

1.4.3 Configuration Tasks 11
1.4.4 Configuring the Monitoring Data Sampling Period 11
1.4.5 Configuring the Slice ID Numbering Mode 12
1.5 Configuring Exception Detection 12
1.5.1 Overview 12
1.5.2 Configuration Tasks 12
1.5.3 Configuring a Buffer Usage Alarm Threshold for a PG 13
1.5.4 Configuring an Alarm Threshold for the Buffer Utilization of a Queue 13
1.5.5 Configuring a Buffer Utilization Alarm Threshold for a Port 14
1.5.6 Enabling or Disabling the Packet Loss Alarm Function for Queues Globally 14
1.5.7 Enabling or Disabling the Packet Loss Alarm Function for Queues in an Interface 15
1.5.8 Configuring the Packet Loss Alarm Frequency 15
1.5.9 Disabling the MMU-HA Function 16
1.5.10 Disabling the MMU-HA Restart Function 16
1.6 Monitoring 16
1.7 Configuration Examples 18
1.7.1 Configuring Ingress PG-based Buffer Application Based on Ingress PGs 18
1.7.2 Configuring the Application of a Large Buffer Based on Egress Queues 20

1 Configuring MMU

1.1 Introduction

1.1.1 Overview

The memory management unit (MMU) is a technology for allocating the chip buffers reasonably. This technology makes devices better deal with all kinds of burst flows.

Flows are not always steady and various burst flows exist in the network. When the network flow is steady and the bandwidth is sufficient, all the data flows are processed properly. When burst flows exist in the network, data flows may be discarded even if the average flow rate does not exceed the bandwidth. Data packets are stored in the buffers of a device before being forwarded. Normally, data packets in the buffers will be forwarded in microseconds. When a burst flow arises, if the instantaneous rate of the burst flow is out of the processing capacity of the device, the data packets that cannot be processed in time are piled up in the buffers and packet loss takes place once the buffers are insufficient. In this case, the MMU technology can used to reasonably configure buffers and allocate different buffer sizes to services, with a view to optimizing the network.

1.1.2 Principles

1. Basic Concepts

- Cell

  Cell is a buffer unit, namely, the minimum unit for the device to store packets. The size of each cell varies with the device. One packet can use multiple cells, while one cell can be used by only one packet.

- Port group

  All the ports that physically belong to the same switching chip and the same core (some chips are multi-core chips) are called a port group. All devices implement buffer management in port groups. For example, card M18000_40XS_CB has two single-core switching chips, so it has two port groups. The first 20 ports of the card belong to port group 1 and the last 20 ports belong to port group 2.

- Buffer slice

  The buffer of a port group may be divided into multiple buffer blocks, and each buffer block is called a buffer slice. Slice is the minimum unit for independent buffer management. Each buffer slice is managed independently, and a port queue may use one or more buffer slices.

- Guaranteed buffer

  A guaranteed buffer is also called an exclusive buffer. This part of buffer is distributed based on each port queue. The guaranteed buffer of a queue can be used by this queue only. A fixed number of guaranteed buffers are allocated to each queue by default. This part of buffer enables a queue to normally forward packets at the line rate under the stable flow.

- Shared buffer

  In the total buffer of a port group, the remaining part obtained after the guaranteed buffer of each port queue is deducted is the total shared buffer. The shared buffer is available to all the queues. A shared buffer threshold can be set for each queue. This threshold restricts the maximum number of shared buffers

that can be used by a queue. When the sum of shared buffers configured for queues in a port group exceeds the total shared buffers of the port group, the "first-come first-served" buffer occupancy mechanism is adopted.

- Service pool

  In addition to the guaranteed buffer, each port queue can use shared buffers. The shared buffers are uniformly stored in the service pool for management. One buffer slice may have multiple service pools. After using its own guaranteed buffer, each queue can request a shared buffer from the service pool within the shared buffer threshold, and the mechanism of "first-request, first-served" is adopted for the service pool. One queue can use the shared buffer of only one service pool according to the mapping configuration.

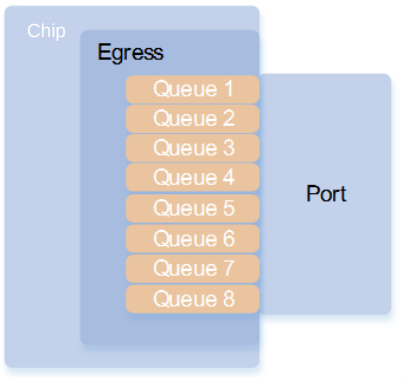
- Priority group (PG)

  PG is the abbreviation of priority group. For the sake of better and flexible buffer management, a chip offers eight PG objects in the ingress direction to implement independent buffer management. Packets are configured with different priorities (for example, 0 to 7) and can be mapped to different PG objects or one PG so that different buffer allocation schemes are provided for packets with different priorities.

- Egress queues

  Egress queues of a port are classified into unicast queues and multicast queues (the number of queues depends on the product). Logically, a switching chip is divided into the ingress direction and egress direction. The egress queues are in the egress direction of a switching chip. Before packets are sent out of a device, they need to be queued at the egress queues, as shown in Figure 1-1.

Figure 1-1 Egress Queues



Currently, there are three types of egress queue models:

- There are eight unicast queues and eight multicast queues at the egress. The known unicast packets enter the unicast queues, and all the other packets enter the multicast queues.
- There are eight unicast queues and four multicast queues at the egress. The known unicast packets enter the unicast queues, and all the other packets enter the multicast queues.
- There are only eight queues at the egress, with unicast and multicast queues not differentiated. The queues are managed as unicast queues in the device.

Caution

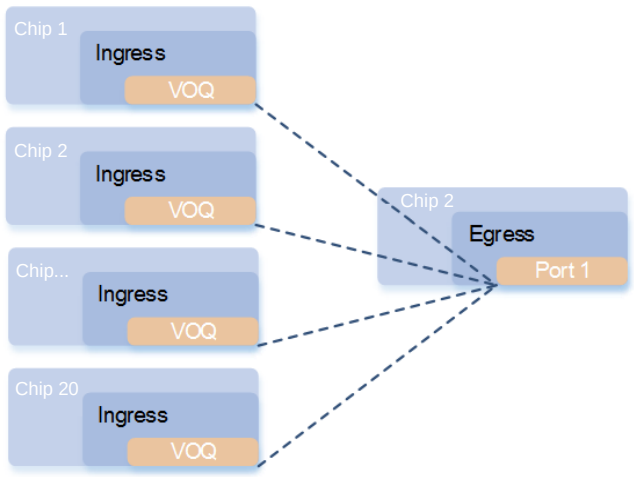
A device that supports egress queue buffer management does not support VOQ queue buffer management.

- VOQ

  VOQ is the abbreviation of virtual output queue. Logically, a switching chip is divided into the ingress direction and egress direction. A VOQ is in the ingress direction of a switching chip. Before packets are sent out of a device, they need to be queued at the VOQs. Each egress is mapped to one VOQ group (containing eight prioritized queues) on each core in all the chips where input ports are located.

  Suppose that a chassis device is capable of housing 10 line cards and each line card has 2 single-core chips, there are 20 switching chips in total. Taking port 1 on chip 2 as an example, in such a switching system, each of the 20 switching chips has one VOQ group in the ingress direction that maps to port 1 on chip 2. For example, a packet enters the device through chip 20 and is destined for port 1 on chip 2. This packet must be queued in the VOQ (on chip 20) mapped to port 1 on chip 2, as shown in Figure 1-2.

Figure 1-2 VOQs



Caution

A device that supports VOQ buffer management does not support egress queue buffer management.

- Global multicast queue

  A global multicast queue refers to the multicast queue shared by all the ports in a port group, and is different from the multicast queue owned by a single physical port.

Caution

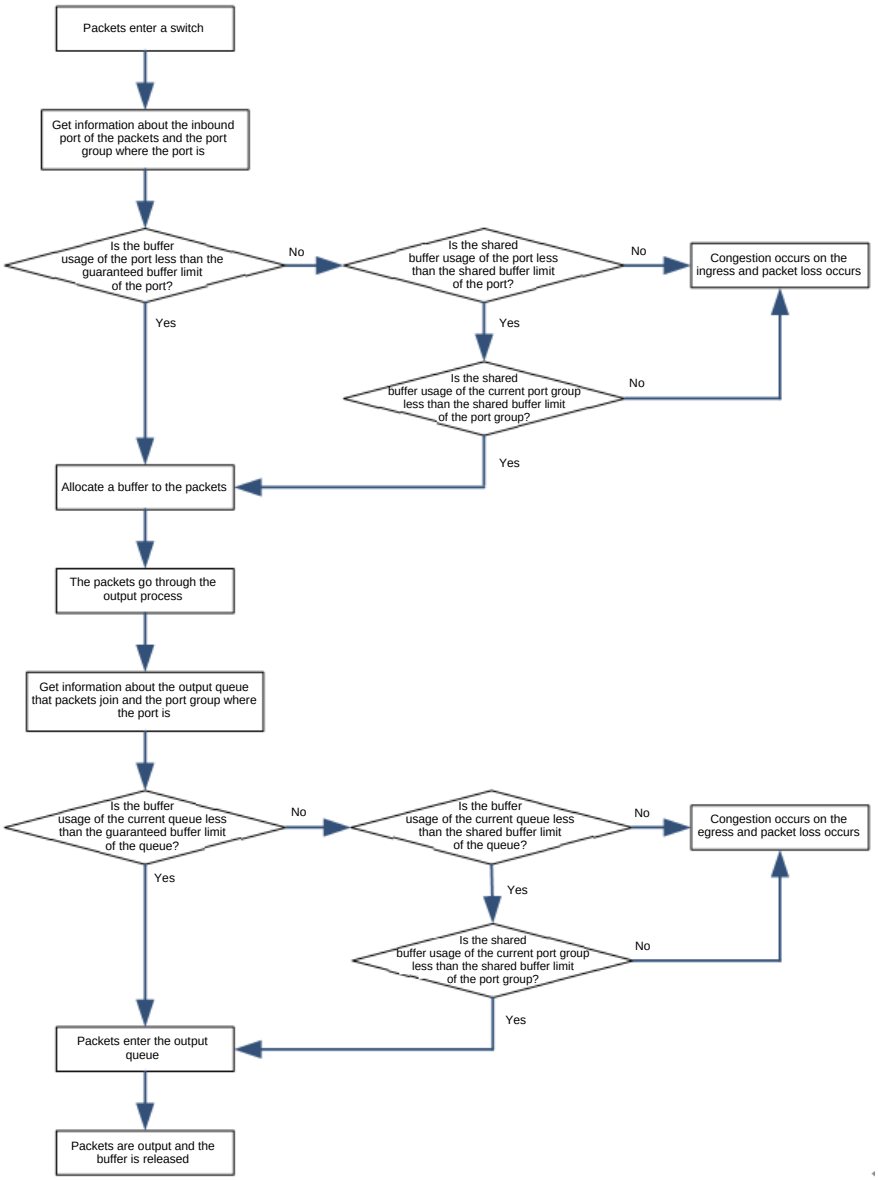
Only the device that supports VOQ buffer management supports global queue buffer management.

2. Buffer Adjustment

Users can adjust the buffer allocation of a chip so that the queue of each service has a different buffer size available, services are treated differently, and services of different priorities are provided.

In terms of hardware, the buffer is managed in the input direction and output direction. Figure 1-3 shows the processing mechanism.

Figure 1-3 Buffer Processing Mechanism



During buffer management, the buffer size is adjusted to the maximum value in the input direction to prevent packet loss in the input direction, and packet loss takes place in the output direction. Therefore, buffer adjustment is not allowed in the input direction, and a CLI command is available for buffer adjustment in the output direction only, including the guaranteed buffer and shared buffer of queues. Buffer adjustment allocates different buffer sizes to queues based on the guaranteed buffer threshold and shared buffer threshold configured for the queues.

3. Status Monitoring

Status monitoring monitors the buffer usage and the queue packet forwarding and discarding by using statistical methods, with a view to providing data support for network optimization and reasonable buffer configuration.

Status monitoring adopts the polling mode to periodically read, from a chip, the buffer usage information and information about packets forwarded/discarded by queues, and analyzes the information so as to display the device operating status.

- Monitoring buffer occupancy in the egress direction

  The information about the guaranteed buffer, shared buffer, and service pool buffer occupancy of queues in the egress direction is regularly collected. The data can be displayed via a CLI command, in MIB mode, or other methods.

- Monitoring packet forwarding and discarding of the egress queues

  The information about the quantities and bytes of packets forwarded and discarded by queues in the egress direction is regularly collected, and the forwarding rate is also calculated. The data can be displayed via a CLI command, in MIB mode, or other methods.

- Sampling interval for status monitoring data

  Status monitoring data is sampled at a fixed interval, which can be configured based on user requirements.

4. Exception Detection

The exception detection function monitors the packet forwarding and discarding of each queue, and displays alarms when the buffer occupancy exceeds the threshold or queue packets are lost.

Exception detection adopts the polling mode to periodically read device buffer occupancy information and the number of packets forwarded by queues, and displays alarms through system logs based on the current configuration. When the buffer utilization of a queue exceeds the alarm threshold, a system log alarm is printed to alert users.

1.2 Configuration Task Summary

MMU configuration includes the following tasks:

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring Buffer Adjustment
- Configuring Status Monitoring
- Configuring Exception Detection

1.3 Configuring Buffer Adjustment

1.3.1 Overview

The buffer adjustment function aims to adjust the objects (such as queues) managed based on chip buffers.

1.3.2 Restrictions and Guidelines

- Configuration on an interface can be made on a physical port only.
- You are advised to configure buffer adjustment when there is no traffic on the device; otherwise the configuration may cause temporary interruption of service traffic.

1.3.3 Configuration Tasks

Buffer adjustment configuration includes the following tasks:

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Guaranteed Buffer for a Queue
- Configuring a Shared Buffer for a Queue
- Configuring a Guaranteed Buffer for an Ingress PG
- Configuring a Shared Buffer for an Ingress PG
- Configuring a Flow Control Threshold for an Ingress PG
- Configuring a Flow Control Threshold Offset for an Ingress PG
- Configuring Headroom Buffer Space for an Ingress PG
- Configuring the Global Headroom Buffer Threshold
- Configuring Cut-

1.3.4 Notes

Table 1-1 Products Supporting Buffer Adjustment Functions

Function	Product Supporting the Function
Configuring Cut-	Supported by the S6520-64CQ, S6510-48VS8CQ, S6510-32CQ, S6510-4C, and S6250-48XS8CQ

1.3.5 Configuring a Guaranteed Buffer for a Queue

1. Overview

This section describes how to configure a guaranteed buffer for a queue so that the queue uses this part of buffer exclusively.

2. Restrictions and Guidelines

- The *queue-id* &<1-n> parameter means that up to *n* queue numbers can be configured. Whether this parameter is supported and the value range of this parameter depend on the actual product version.
- The configuration is displayed when the **display current-config** command is executed even if the default value is used.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Ethernet interface configuration mode.

interface *ethernet-type interface-number*

(4) Configure a guaranteed buffer for queues.

mmu queue-guarantee { **output** { **multicast** | **unicast** } } [*queue-id*&<1-n>] **set** *value*

No guaranteed buffer is configured for a queue by default.

1.3.6 Configuring a Shared Buffer for a Queue

1. Overview

This section describes how to configure a shared buffer for a queue so as to control the size of the shared buffer that can be used by the queue.

2. Restrictions and Guidelines

- The *queue-id*&<1-n> parameter means that up to *n* queue numbers can be configured.
- The configuration is displayed when the **display current-config** command is executed even if the default value is used.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Ethernet interface configuration mode.

interface *ethernet-type interface-number*

(4) Configure a shared buffer for queues.

mmu queue-threshold { **output** { **multicast** | **unicast** } } [*queue-id*&<1-n>] **set** { *threshold* | **cell** *value* }

No shared buffer is configured for a queue by default.

1.3.7 Configuring a Guaranteed Buffer for an Ingress PG

1. Overview

This section describes how to configure a guaranteed buffer for an ingress priority group (PG) so that the PG uses the buffer exclusively.

2. Restrictions and Guidelines

- The *pg-id* &<1-n> parameter indicates that up to *n* queues can be configured.
- The configuration is displayed when the **display current-config** command is executed even if the default value is used.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure a guaranteed buffer for an ingress PG.

mmu **pg-guarantee pg** [*pg*-id&<1-n>] set *value*

No guaranteed buffer is configured for an ingress PG by default.

1.3.8 Configuring a Shared Buffer for an Ingress PG

1. Overview

This section describes how to configure a shared buffer for an ingress PG so as to control the size of the shared buffer that can be used by the ingress PG.

2. Restrictions and Guidelines

- The *pg-id*&<1-n> parameter indicates up to *n* queues can be configured.
- The configuration is displayed when the **display current-config** command is executed even if the default value is used.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure a shared buffer for an ingress PG.

mmu ingress-threshold pg [*pg-id*&<1-n>] **set** *threshold*

No shared buffer is configured for an ingress PG by default.

1.3.9 Configuring a Flow Control Threshold for an Ingress PG

1. Overview

This section describes how to configure a flow control threshold for an ingress PG and adjust the flow control threshold of the ingress based on the PG.

2. Restrictions and Guidelines

- Whether this function is supported depends on the actual product version.
- The configuration takes effect only when flow control or PFC is enabled on a port.
- The configuration is displayed when the **display current-config** command is executed even if the default value is used.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 or Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

(4) Configure a flow control threshold for an ingress PG.

mmu xoff-threshold pg [*pg-id&<1-n>*] **set** *threshold*

No flow control threshold is configured for an ingress PG by default.

1.3.10 Configuring a Flow Control Threshold Offset for an Ingress PG

1. Overview

This section describes how to configure the flow control threshold offset for an ingress PG to adjust the interval for restoring the PG from the XOFF state to the XON state.

2. Restrictions and Guidelines

- The configuration takes effect only when flow control/PFC is enabled.
- The configuration is displayed when the **display current-config** command is executed even if the default value is used.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure a flow control threshold offset for an ingress PG.

mmu xon-threshold-offset pg [*pg-id&<1-n>*] **set** *threshold*

No flow control threshold offset is configured for an ingress PG by default.

1.3.11 Configuring Headroom Buffer Space for an Ingress PG

1. Overview

This section describes how to configure headroom buffer space for an ingress PG to adjust the headroom buffer space that can be used by the PG.

2. Restrictions and Guidelines

- The configuration takes effect only when flow control/PFC is enabled.
- The configuration is displayed when the **display current-config** command is executed even if the default value is used.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure a flow control headroom buffer threshold for a specified ingress PG.

mmu headroom-threshold pg [*pg-id&<1-n>*] **set** *threshold*

No headroom buffer space is configured for an ingress PG by default.

1.3.12 Configuring the Global Headroom Buffer Threshold

1. Overview

This section describes how to configure the global headroom buffer threshold to control the headroom buffer space reserved for a PG.

2. Restrictions and Guidelines

The configuration is displayed when the **display current-config** command is executed even if the default value is used.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a flow control headroom buffer threshold for ingress PGs.

mmu pg-headroom set *value*

No global headroom buffer threshold is configured by default.

1.3.13 Configuring Cut-Through

Note

This function is supported by only the S6520-64CQ, S6510-48VS8CQ, S6510-32CQ, S6510-4C, and S6250-48XS8CQ.

1. Overview

This section describes how to enable the cut-through function to reduce the packet forwarding latency.

2. Restrictions and Guidelines

After the function is enabled, chips determine whether to forward packets in cut-through mode based on the actual condition. For example, the cut-through function is unavailable upon congestion.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure the cut-through mode.

mmu cut-through

The cut-through function is disabled by default.

1.4 Configuring Status Monitoring

1.4.1 Overview

This section describes how to collect statistics on and monitor the buffer usage and queue packet forwarding and discarding so as to provide guidance for buffer adjustment configuration.

1.4.2 Restrictions and Guidelines

Configuration on an interface can be made on a physical port only.

1.4.3 Configuration Tasks

The status monitoring configuration includes the following tasks:

All the following configuration tasks are optional. Select the configuration tasks as required.

- Configuring the Monitoring Data Sampling Period
- Configuring the Slice ID Numbering Mode
- Configuring Buffer Statistics Tracking

1.4.4 Configuring the Monitoring Data Sampling Period

1. Overview

This section describes how to configure the monitoring data sampling period. The device samples and reports data based on the configured interval.

2. Restrictions and Guidelines

- The separate buffer statistics or packet statistics sampling period has a higher priority than the monitoring data sampling period.
- The configuration is displayed when the **display current-config** command is executed even if the default value is used.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the monitoring data sampling period.

mmu sample-period { { **buffer-counter | queue-counter** } *counter* **| set value** }

No monitoring data sampling period is configured by default.

1.4.5 Configuring the Slice ID Numbering Mode

1. Overview

This section describes how to configure the slice ID numbering mode to control slice IDs to start from 1 or 0.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the slice ID numbering mode.

mmu slice-id compatible

No slice ID numbering mode is configured by default.

1.4.6 Configuring Buffer Statistics Tracking

1. Overview

Configure buffer statistics tracking to record the buffer usage peak value.

2. Restrictions and Guidelines

- The device works in the current-use-count mode by default. After the max-use-count mode is configured, features in the current-use-count will not be active, such as some MMU-gRPC triggered events.

3. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Configure buffer statistics tracking.
mmu buffer-statistics-tracking max-use-count

The device works in the current-use-count mode without recording the buffer usage peak value by default.

1.5 Configuring Exception Detection

1.5.1 Overview

The exception detection function is used to analyze the data about device buffer and forwarding status, and generate an alarm when an exception occurs.

1.5.2 Configuration Tasks

Exception detection configuration includes the following tasks:

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Buffer Usage Alarm Threshold for a PG
- Configuring an Alarm Threshold for the Buffer Utilization of a Queue

- Configuring a Buffer Utilization Alarm Threshold for a Port
- Enabling or Disabling the Packet Loss Alarm Function for Queues Globally
- Enabling or Disabling the Packet Loss Alarm Function for Queues in an Interface
- Configuring the Packet Loss Alarm Frequency
- Disabling the MMU-HA Function
- Disabling the MMU-HA Restart Function

1.5.3 Configuring a Buffer Usage Alarm Threshold for a PG

1. Overview

This section describes how to configure a buffer usage alarm threshold for PGs. When the buffer usage of a PG exceeds the threshold, a log alarm is generated.

2. Restrictions and Guidelines

- To prevent frequent log refreshing, the alarm logs of the same PG are printed once in 30s at most. The maximum printing interval depends on the configured sampling period.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the buffer utilization alarm threshold for a PG.

mmu usage-warn-limit set *threshold*

The default buffer utilization alarm threshold is **0**, indicating that no alarm is reported.

This configuration takes effect on all port groups or buffer slices.

1.5.4 Configuring an Alarm Threshold for the Buffer Utilization of a Queue

1. Overview

When the buffer utilization of a queue exceeds the configured value, a system log alarm is printed.

2. Restrictions and Guidelines

- The *queue-id* &<1-n> parameter means that up to *n* queue numbers can be configured.
- To prevent frequent log refreshing, the alarm logs of the same queue are printed once in 30s at most. The maximum printing interval depends on the configured sampling period.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 or Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

(4) Configure a buffer utilization alarm threshold for a queue.

mmu usage-warn-limit { **multicast** | **unicast** } [*queue-id*&<1-n>] **set** *threshold*

The default buffer utilization alarm threshold is **0**, indicating that no alarm is reported.

1.5.5 Configuring a Buffer Utilization Alarm Threshold for a Port

1. Overview

When the buffer utilization of all the queues on a port exceeds the configured value, a system log alarm is printed.

2. Restrictions and Guidelines

- The *queue-id* &<1-n> parameter means that up to *n* queue numbers can be configured.
- To prevent frequent log refreshing, the alarm logs of the same port are printed once in 30s at most. The maximum printing interval depends on the configured sampling period.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 or Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

(4) Configuring a buffer utilization alarm threshold for a port.

mmu usage-warn-limit { **pg-headroom** | **priority-group** } [*pg-id&<1-n>*] **set** *value*

The default buffer utilization alarm threshold is **0**, indicating that no alarm is reported.

1.5.6 Enabling or Disabling the Packet Loss Alarm Function for Queues Globally

1. Overview

You can enable or disable the packet loss alarm function for queues globally to control the packet loss alarm function of all interface queues.

2. Restrictions and Guidelines

- The *queue-id* &<1-n> parameter indicates that up to *n* queues can be assigned with IDs.
- If this command is configured in both interface configuration mode and global configuration mode, the command configuration in interface configuration mode has a higher priority than that in global configuration mode.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable or disable the packet loss alarm function for queues globally.

mmu queue-loss-warn { **multicast** | **unicast** } [queue-id&<1-n>] { **on** | **off** }

The packet loss alarm function is disabled by default.

1.5.7 Enabling or Disabling the Packet Loss Alarm Function for Queues in an Interface

1. Overview

You can enable or disable the packet loss alarm function for queues in an interface to control the packet loss alarm function of the interface. The priority of this command is higher than that in global configuration mode.

2. Restrictions and Guidelines

- The *queue-id* &<1-n> parameter indicates that up to *n* queues can be assigned with IDs.
- If this command is configured in both interface configuration mode and global configuration mode, the command configuration in interface configuration mode has a higher priority than that in global configuration mode.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Enable or disable the packet loss alarm function for queues in an interface.

mmu queue-loss-warn { **multicast** | **unicast** } [*queue-id*&<1-n>] { **on** | **off** }

The packet loss alarm function of an interface is disabled by default.

1.5.8 Configuring the Packet Loss Alarm Frequency

1. Overview

You can configure the packet loss alarm frequency so that packet loss alarms are generated N times within the period (T).

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the packet loss alarm frequency.

mmu queue-loss-warn frequency cycle *value1* **times** *value2*

Packet loss alarms are generated 30 times within 60 seconds by default.

1.5.9 Disabling the MMU-HA Function

1. Overview

This section describes how to disable the MMU high availability (MMU-HA) function globally.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable the MMU-HA function.

mmu ha disable

The MMU-HA function is enabled by default.

1.5.10 Disabling the MMU-HA Restart Function

1. Overview

This section describes how to disable the MMU-HA restart function. After the function is disabled, the system still detects whether the MMU buffer is suspended.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable the MMU-HA restart function.

mmu ha reset disable

The MMU-HA restart function is enabled by default.

1.6 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-2 MMU Monitoring

Command	Purpose
display [**mmu**] **queue-counter** [**interface** *interface-type interface-number* [**slot** *slot-id*] | **multicast**]	Displays the packet statistics of a queue.
display [**mmu**] **queue-buffer** [**interface** *interface-type interface-number* [**slot** *slot-id*] | **multicast**]	Displays the buffer utilization information of a queue.
display [mmu] pg-buffer [**interface** *interface-type interface-number*]	Displays the buffer usage information of a PG.
display mmu buffer-mode	Displays the current buffer mode.
display mmu buffer-config interface *interface-type interface-number* [**slot** *slot-id*]	Displays the difference between the buffer configuration conducted by users and the actually effective buffer configuration.
display mmu buffer-config queue-threshold cell interface *interface-type interface-number* [**slot** *slot-id*]	Displays the difference between the static buffer threshold configured by users and the actually effective buffer configuration.
display [**mmu**] **priority-group buffer-statistics** [**interface** *interface-type interface-number* | **total**]	Displays the buffer usage information of a PG on a specified port.
display [**mmu**] **ingress-queue buffer-statistics { interface** *interface-type interface-number* | **total }** [**slot** *slot-id*]	Displays the ingress queue buffer statistics.
display [**mmu**] **egress-queue buffer-statistics** [**interface** *interface-type interface-number* | **total**]	Displays the buffer usage information of queues in the outbound direction.
display [**mmu**] **ingress-queue traffic-statistics interface** *interface-type interface-number* [**slot** *slot-id*]	Displays the ingress queue traffic statistics.
display [**mmu**] **egress-queue traffic-statistics interface** *interface-type interface-number*	Displays traffic statistics in the outbound direction.
display [**mmu**] **{ priority-group | egress-queue } buffer-statistics total**	Displays the global buffer usage information of PGs and queues in both inbound and outbound directions.
display [**mmu**] **ingress-queue traffic-statistic multicast** [**slot** *slot-id*]	Displays global multicast traffic statistics.
display [**mmu**] **ingress-queue buffer-statistics multicast** [**slot** *slot-id*]	Displays the global multicast buffer usage information.
cleanup [**mmu**] **queue-counter** [**interface** *interface-type interface-number*]	Clears the packet statistics of a queue.
cleanup [**mmu**] **queue-buffer peaked** [**interface** *interface-type interface-number*]	Clears the historical buffer utilization peak values of a queue.

Command	Purpose
cleanup mmu usage-warn-count	Clears the count of all buffer alarms.
cleanup queue-buffer usage-warn-count	Clears the count of buffer alarms of a queue.

1.7 Configuration Examples

1.7.1 Configuring Ingress PG-based Buffer Application Based on Ingress PGs

1. Requirements

An enterprise needs to ensure no packet loss on the storage data flow in the data center network.
A traditional data center uses three types of networks to carry different flows to meet different traffic requirements, as shown in the following topology.

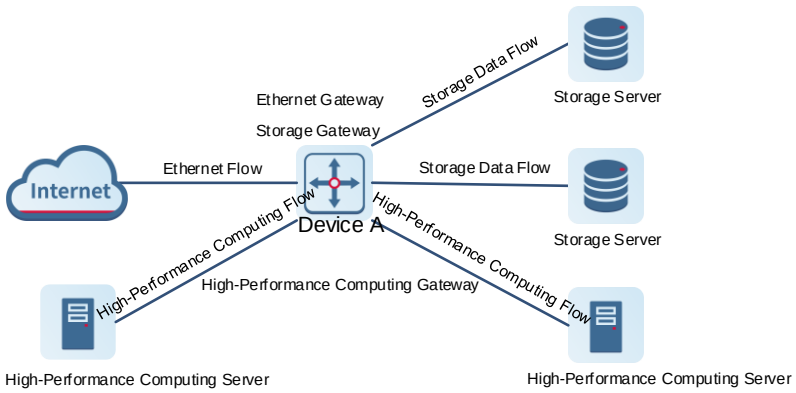
There are mainly three types of flows:

- Storage data flow: No packet loss is allowed.
- High-performance computing flow: A low latency is required.
- Ethernet flow: Certain packet loss and delay are allowed.

With the expansion of data centers, the Ethernet is used to carry all of three different types of flows, and additional mechanisms are defined to enable the Ethernet to meet the requirements of the types of network traffic. In addition, the priority-based flow control (PFC) mechanism based on ingress PGs is used to meet the requirements of no packet loss on the storage data flow and no impact on the other two types of flows when the three types of flows coexist on the Ethernet.

2. Topology

Figure 1-4 Configuring Ingress PG-based Buffer Application



3. Notes

- Map the remote direct memory access (RDMA) traffic to a PG.
- Enable the PFC function on the PG to ensure that no packet loss occurs on the traffic of this priority.

- Configure the XON and XOFF thresholds for the PG so that flow control frames can be generated.
- Configure sufficient headroom buffer space for the PG to buffer packets on the lines at the interval between effective flow control periods.
- For PFC-disabled PGs, the shared buffer threshold of the PGs should be higher than the buffer threshold of egress queues to ensure no packet loss on the ingress.
- For PFC-enabled PGs, the XOFF buffer threshold of the PGs should be lower than the buffer threshold of the egress to trigger the PFC function.
- You need to adjust values based on the network environment. Values provided in this section are merely examples for reference.

4. Procedure

(1) Configure a guaranteed buffer for an ingress PG.

Configure a guaranteed buffer for an ingress PG on the storage flow service port of device A.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list gigabitethernet 0/1-20
DeviceA(cfg-if-list)# mmu pg-guarantee pg 3 set 10
DeviceA(cfg-if-list)# exit
```

(2) Configure a flow control threshold for the ingress PG.

Configure the flow control threshold for the ingress PG on the storage flow service port of device A.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list gigabitethernet 0/1-20
DeviceA(cfg-if-list)# mmu xoff-threshold pg 3 set 20
DeviceA(cfg-if-list)# exit
```

(3) Configure a flow control threshold offset for the ingress PG.

Configure a flow control threshold offset for the ingress PG on the storage flow service port of device A.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list gigabitethernet 0/1-20
DeviceA(cfg-if-list)# mmu xon-threshold-offset pg 3 set 16
DeviceA(cfg-if-list)# exit
```

(4) Configure a headroom buffer for the ingress PG.

Configure a headroom buffer for the ingress PG on the storage flow service port of device A.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list gigabitethernet 0/1-20
DeviceA(cfg-if-list)# mmu headroom-threshold pg 3 set 300
DeviceA(cfg-if-list)# exit
```

5. Verification

On device A, verify that the command configuration for the interface GigabitEthernet 0/1 is correct.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# display this-config
```

```
Building configuration...
!
 mmu pg-guarantee pg 3 set 10
 mmu xoff-threshold pg 3 set 20
 mmu xon-threshold-offset pg 3 set 16
 mmu headroom-threshold pg 3 set 300
```

6. Configuration Files

Device A configuration file

The configurations of the interface GigabitEthernet 0/1 to the interface GigabitEthernet 0/20 are exactly the same, and therefore only the configuration of GigabitEthernet 0/1 is provided below.

```
hostname DeviceA
!
interface GigabitEthernet 0/1
 mmu pg-guarantee pg 3 set 10
 mmu xoff-threshold pg 3 set 20
 mmu xon-threshold-offset pg 3 set 16
 mmu headroom-threshold pg 3 set 300
!
```

1.7.2 Configuring Egress Queue-based Large Buffer Application

1. Requirements

An enterprise needs a buffer large enough for the net disk service to avoid packet loss occurring in the service flow.

As shown in the following figure, device A is connected to 5 clients and 35 service servers, and 15 service servers are virtualized as 15 front-end servers.

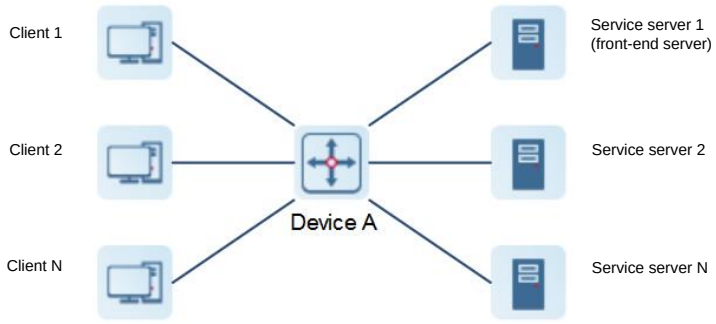
The main service flow is as follows:

- A client sends a request packet to a front-end server.
- The front-end server sends the received request packet to a service server.
- After receiving the request packet, the service server sends a response packet to the front-end server.
- After receiving the response packet, the front-end server sends it to the client.
- The client receives the response packet, indicating that a session is created successfully.
- A many-to-one flow transmission mode exists in this service model:
- The request flows of multiple clients are sent to one front-end server.
- The request flows of multiple front-end servers are sent to one service server.
- The response flows of multiple service servers are sent to one front-end server.
- The response flows of multiple front-end servers are sent to one client.

These flows are basically transmitted through device A, which can easily lead to network congestion. Such a problem can be fixed by configuring a large buffer on the device.

2. Topology

Figure 1-5 Configuring Egress Queue-based Large Buffer Application



3. Notes

- On all the service ports (namely, the ports connecting to clients and servers), set the shared buffer to 100% for the queues where the service is.
- On all the service ports, set the guaranteed buffer to the minimum size for queues not in use.
- On all the ports not in use, set the guaranteed buffer to the minimum size for all the queues.

4. Procedure

(1) Configure a shared buffer.

On device A, set a shared buffer for queues on all the service ports.

```
DeviceA> enabel
DeviceA# config-system terminal
DeviceA(cfg)# interface list gigabitethernet 0/1-40
DeviceA(cfg-if-list)# mmu queue-threshold output unicast 1 set 100
DeviceA(cfg-if-list)# mmu queue-threshold output multicast 1 set 100
DeviceA(cfg-if-list)# exit
```

(2) Configure a guaranteed buffer.

On device A, set a guaranteed buffer for queues on all the service ports.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface list gigabitethernet 0/1-40
DeviceA(cfg-if-list)# mmu queue-guarantee output unicast 2 3 4 5 6 7 8 set 1
DeviceA(cfg-if-list)# mmu queue-guarantee output multicast 2 3 4 5 6 7 8 set 1
DeviceA(cfg-if-list)# exit
```

On device A, set a guaranteed buffer for queues on all the ports not in use.

```
DeviceA(cfg)# interface list gigabitethernet 0/41-48
DeviceA(cfg-if-list)# mmu queue-guarantee output unicast set 1
DeviceA(cfg-if-list)# mmu queue-guarantee output multicast set 1
DeviceA(cfg-if-list)# exit
```

5. Verification

On device A, verify that the command configuration for interface GigabitEthernet 0/1 is correct.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# display this-config

Building configuration...
!
 mmu queue-threshold output unicast 1 set 100
 mmu queue-threshold output multicast 1 set 100
 mmu queue-guarantee output unicast 2 set 1
 mmu queue-guarantee output unicast 3 set 1
 mmu queue-guarantee output unicast 4 set 1
 mmu queue-guarantee output unicast 5 set 1
 mmu queue-guarantee output unicast 6 set 1
 mmu queue-guarantee output unicast 7 set 1
 mmu queue-guarantee output unicast 8 set 1
 mmu queue-guarantee output multicast 2 set 1
 mmu queue-guarantee output multicast 3 set 1
 mmu queue-guarantee output multicast 4 set 1
 mmu queue-guarantee output multicast 5 set 1
 mmu queue-guarantee output multicast 6 set 1
 mmu queue-guarantee output multicast 7 set 1
 mmu queue-guarantee output multicast 8 set 1
```

6. Configuration Files

Device A configuration file

The configuration of interface GigabitEthernet 0/1 to interface GigabitEthernet 0/40 is exactly the same, and therefore only the configuration of GigabitEthernet 0/1 is provided below. The configuration of interface GigabitEthernet 0/41 to interface GigabitEthernet 0/48 is exactly the same, and therefore only the configuration of GigabitEthernet 0/41 is provided below.

```
hostname DeviceA
!
interface GigabitEthernet 0/1
 mmu queue-threshold output unicast 1 set 100
 mmu queue-guarantee output unicast 2 set 1
 mmu queue-guarantee output unicast 3 set 1
 mmu queue-guarantee output unicast 4 set 1
 mmu queue-guarantee output unicast 5 set 1
 mmu queue-guarantee output unicast 6 set 1
 mmu queue-guarantee output unicast 7 set 1
 mmu queue-guarantee output unicast 8 set 1
 mmu queue-threshold output multicast 1 set 100
 mmu queue-guarantee output multicast 2 set 1
 mmu queue-guarantee output multicast 3 set 1
 mmu queue-guarantee output multicast 4 set 1
 mmu queue-guarantee output multicast 5 set 1
 mmu queue-guarantee output multicast 6 set 1
```

```
  mmu queue-guarantee output multicast 7 set 1
  mmu queue-guarantee output multicast 8 set 1
!
interface gigabitEthernet 0/41
  mmu queue-guarantee output unicast 1 set 1
  mmu queue-guarantee output unicast 2 set 1
  mmu queue-guarantee output unicast 3 set 1
  mmu queue-guarantee output unicast 4 set 1
  mmu queue-guarantee output unicast 5 set 1
  mmu queue-guarantee output unicast 6 set 1
  mmu queue-guarantee output unicast 7 set 1
  mmu queue-guarantee output unicast 8 set 1
  mmu queue-guarantee output multicast 1 set 1
  mmu queue-guarantee output multicast 2 set 1
  mmu queue-guarantee output multicast 3 set 1
  mmu queue-guarantee output multicast 4 set 1
  mmu queue-guarantee output multicast 5 set 1
  mmu queue-guarantee output multicast 6 set 1
  mmu queue-guarantee output multicast 7 set 1
  mmu queue-guarantee output multicast 8 set 1
!
```

Security Configuration

1. Configuring AAA
2. Configuring RADIUS
3. Configuring TACACS
4. Configuring SCC
5. Configuring Password Policies
6. Configuring SSH
7. Configuring CPP
8. Configuring NFPP
9. Configuring Storm Control
10. Configuring uRPF
11. Configuring Security Log Auditing

Contents

1 Configuring AAA 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.2 Restrictions and Guidelines 6

1.3 Configuration Task Summary 7

1.4 Configuring an AAA Server Group 7

1.4.1 Overview 7

1.4.2 Restrictions and Guidelines 7

1.4.3 Procedure 7

1.5 Configuring Basic Features of AAA Authentication 8

1.5.1 Overview 8

1.5.2 Restrictions and Guidelines 8

1.5.3 Configuration Tasks 9

1.5.4 Configuring Login Authentication 9

1.5.5 Configuring Enable Authentication 10

1.5.6 Configuring FTP Authentication 10

1.6 Configuring AAA Authentication Logging 11

1.6.1 Overview 11

1.6.2 Procedure 11

1.7 Configuring Basic Features of AAA Authorization 11

1.7.1 Overview 11

1.7.2 Restrictions and Guidelines 11

1.7.3 Procedure....12
1.8 Configuring the Function of Caching Command Authorization Results....13
1.8.1 Overview....13
1.8.2 Restrictions and Guidelines....13
1.8.3 Procedure....13
1.9 Configuring the Function of Allowing Users to Log in to the Slave Device....13
1.9.1 Overview....13
1.9.2 Restrictions and Guidelines....13
1.9.3 Procedure....13
1.10 Configuring Basic Features of AAA Accounting....14
1.10.1 Overview....14
1.10.2 Restrictions and Guidelines....14
1.10.3 Procedure....14
1.11 Configuring a Policy for User Accounting Start Failures....15
1.11.1 Overview....15
1.11.2 Procedure....15
1.12 Configuring Domain-Based AAA Services....16
1.12.1 Overview....16
1.12.2 Restrictions and Guidelines....16
1.12.3 Procedure....16
1.13 Monitoring....17
1.14 Configuration Examples....17
1.14.1 Configuring Local Login Authentication and Local EXEC Authorization....17
1.14.2 Configuring Enable Authentication....19

1.14.3 Configuring Command Authorization 21
1.14.4 Configuring Network Authorization 23
1.14.5 Configuring an AAA Server Group 24
1.14.6 Configuring Domain-based AAA Services 25
1.15 Common Errors 27

1 Configuring AAA

1.1 Introduction

1.1.1 Overview

Authentication, authorization and accounting (AAA) is a network security management mechanism, which provides basic authentication, authorization, and accounting (billing) services for users who access networks. The content of AAA is described as follows:

- Authentication service: Identifies users before they access networks to check whether the users have access permissions.
- Authorization service: Classifies user permissions to grants different access permissions to users.
- Accounting service: Records users' usage of network resources. The statistical data can be used for analysis and billing.

AAA boasts strong flexibility and controllability, scalability, standardized authentication, and supports multiple backup systems. Compared with simple access control functions provided by devices such as local username authentication and line password-based authentication, AAA provides a higher level of security protection and therefore becomes the overriding access control method.

1.1.2 Principles

1. Basic Concepts

- Security method

  A security method refers to the method used to implement authentication, authorization, and accounting services. Security methods can be classified into AAA authentication methods, AAA authorization methods, and AAA accounting methods by service type.

  Based on the service provider, service methods can be further classified into service methods providing no security management, service methods providing security management through a local device, and service methods providing security management through a remote server. For details, see 1.1.2 3. AAA Authentication, 1.1.2 4. AAA Authorization, and 1.1.2 5. AAA Accounting.

- Method list

  Method lists are classified into authentication method lists, authorization method lists, and accounting method lists based on the AAA service type. Each method list provides multiple security methods for this type of AAA service, to ensure that alternative methods are available if one method is unavailable.

  A method list contains one or more security methods. The sequence for the methods to take effect is subject to the sequence of security methods defined in the list, that is, a method defined earlier takes effect earlier. If a method fails to respond during AAA service provision, a device switches to the next method until all methods in the method list are tried. If a method returns a failure or all methods fail to respond, the process stops and the device cannot provide the security service for users.

Caution

A device tries the next method only when no response is received from the previous method. For example, if a method denies user access during identity authentication, the identity authentication process ends and the device does not switch to other identity authentication methods.

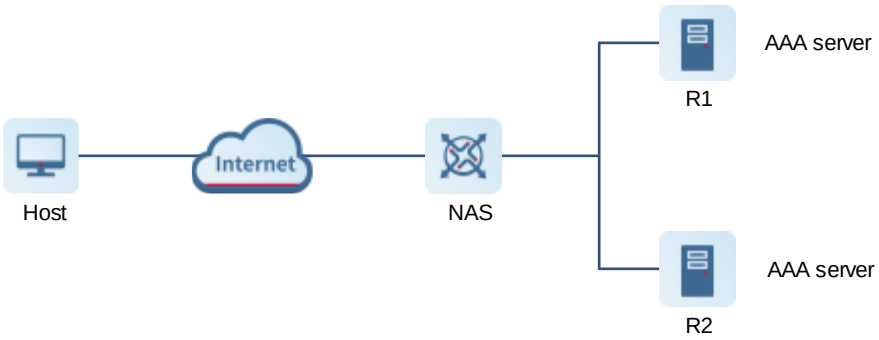
- AAA server group

  An AAA server is a device that provides remote AAA security services for users.

  An AAA server group is a collection of one or more AAA servers of the same type and provides security services for users as a security method.

2. AAA Basic Architecture

Figure 1-1 Topology of a Typical AAA Network



As shown in Figure 1-1, a typical AAA network consists of hosts (users), a network access server (NAS), and an AAA server (remote server). AAA uses the client/server (C/S) structure. A host sends an AAA request to the NAS as a client. After receiving the request, the NAS requests the AAA security services from the AAA server as a client and provides network access permissions for the host based on the returned result.

- Main roles

  Host: A host is a device that requests network access permissions and usually refers to a user.

  NAS: A NAS responds to hosts' network access requests and manages the hosts by using security methods provided by an AAA server.

  AAA server: An AAA server provides security methods for the NAS.

- Brief process of typical AAA services

  a Configure a method list and enable the AAA security services on the NAS. The method list provides three security methods: R1 server, R2 server, and local username database (with the effective sequence from left to right). Host-related identity information has been configured in the three methods.

  b After a host connects to the NAS through a network, the NAS provides the AAA security services for the host by using a method in the method list. According to the sequence that the methods take effect, the NAS requests the security services from R1 first.

  c If R1 responds to the NAS successfully, the NAS provides network access permissions for the host based on security information provided by R1. If R1 does not respond, the NAS switches the security

method and requests the security services from R2, and so on until all methods in the method list are used up.

Note

An AAA server can use a Remote Authentication Dial-In User Service (RADIUS) or Terminal Access Controller Access-Control System Plus (TACACS+) server. Unless otherwise specified in this document, the RADIUS server is used as the AAA server. For details about TACACS+, see *Configuring TACACS*.

3. AAA Authentication

AAA authentication checks information (username and password) about a user who requests the access service, to determine whether the user has the network access permission.

- AAA authentication methods

  AAA authentication methods are classified into authentication exemption, local authentication, and remote server group authentication based on the method provider.

  - Authentication exemption

    User legitimacy is not checked. This method is used only when users are very trustworthy. You are not advised to use this method.

  - Local authentication

    Local authentication uses the database on the NAS to manage user information (such as usernames, passwords, and attributes) and identify user identities.

  - Remote server group authentication

    This authentication method uses the user database on the remote server to manage user data and identify user identities. It can achieve multi-device centralized authentication with large capacity and high reliability. A remote server group supports RADIUS servers and TACACS+ servers.

- AAA authentication types

  AAA authentication is classified into login authentication, enable authentication, and File Transfer Protocol (FTP) authentication based on the user access mode.

  - Login authentication

    This authentication verifies user identities when the users log in to the command line interface (CLI) of the NAS through Secure Shell (SSH), telnet, or FTP.

  - Enable authentication

    After users log in to the NAS CLI, this authentication verifies the users' identity information before their CLI execution permissions are elevated, that is, users are authenticated before they enter the privileged EXEC mode.

  - FTP authentication

    FTP authentication is especially used for FTP users. If no FTP authentication method is configured, login authentication is used to authenticate FTP users.

4. AAA Authorization

AAA authorization is used to manage services available to users and control network access permissions of access users. After the AAA authorization service is enabled, users can use only permitted services and are granted permissions in their permitted scopes.

- AAA authorization methods

  AAA authorization methods are classified into direct authorization, local authentication, and remote server group authentication based on the method provider.

  - Direct authorization

    Direct authorization provides default permissions that the access device allows users to use. It is used only when users are very trustworthy.

  - Local authorization

    Configure attribute authorization for local users on the NAS, which completes the authorization.

  - Remote server group authorization

    Configure user attribute authorization on the remote server. The NAS completes authorization by means of a remote server group. When all servers in a remote server group fail, you can configure local authorization or direct authorization as an alternative method.

- AAA authorization types

  AAA authorization is classified into EXEC authorization, command authorization, config-commands authorization, console authorization, and network authorization based on the permission type.

  - EXEC authorization

    EXEC authorization grants different privilege levels (0–15) to users when the users log in to the NAS CLI.

    EXEC authorization is usually used in conjunction with login authentication. After login authentication and EXEC authorization are configured on the same line, if a user passes login authentication but fails in EXEC authorization, the user cannot open the CLI.

  - Command authorization

    Command authorization grants authorization to commands after users log in to the NAS CLI.

    When a user enters a command and attempts to run it, AAA sends the command to the security server. If the security server allows this command, the command is executed. Otherwise, the command is not executed and a command execution denial message is displayed.

  - Config-commands authorization

    This authorization grants authorization to commands in configuration modes (including the global configuration mode and sub-modes).

  - Console authorization

    This authorization grants authorization to commands to be executed by users who log in through the console.

    The system can differentiate users who log in through the console from those who log in through other terminals. You can enable or disable command authorization for users who log in through the console. If command authorization is disabled for users who log in through the console, the command authorization method list that has been applied to the console line no longer takes effect.

- Network authorization

  Network authorization grants available network services to users, such as traffic, bandwidth, and timeout services.

  Network authorization is based on authentication. Network authorization is available only for authenticated users. The RADIUS or TACACS+ server grants permissions to authenticated users by returning a series of attributes.

5. AAA Accounting

AAA accounting is used to record users' usage of network resources to implement billing, auditing, and tracking functions.

The accounting service starts accounting after a user passes authentication and records the user's network access activities. It stops accounting when the user logs out, and generates a network resource access report for the user.

- AAA accounting methods

  AAA accounting methods are classified into no accounting, local accounting, and remote server group accounting based on the method provider.

  - No accounting

    Accounting is not performed on users.

  - Local accounting

    Local accounting is completed by the NAS, which collects statistics on and limits the number of local user connections. The fee calculation function is unavailable.

  - Remote server group accounting

    Accounting is performed by a remote server. You can configure local accounting as an alternative method to avoid an accounting failure occurring when all servers in a remote server group fail.

- AAA accounting types

  AAA accounting can be classified into EXEC accounting, command accounting, and network accounting based on the accounting content.

  - EXEC accounting

    Accounting is both performed when users log in to and log out of the NAS CLI.

  - Command accounting

    Command accounting is used to record commands run by a user after the user logs in to the NAS CLI from a terminal.

  - Network accounting

    Network accounting is used to record information about sessions of users who access the network.

6. Multi-Domain AAA

In multi-domain AAA, users are assigned to different domains and independent security methods are defined for each domain so that users in different domains use different security methods.

In a multi-domain environment, you can configure a separate AAA method list for each domain on the NAS. Therefore, users with the same attribute can be assigned to the same domain for the ease of management.

- Basic principles of AAA service based on domain names
  a A user submits identity authentication information to the NAS.
  b The NAS resolves the domain name in the identity authentication information.
  c The NAS checks whether the domain name is defined. If no, the NAS refuses to provide security services. If yes, the NAS makes further judgment.
  d If no method list is defined for the domain name, the NAS refuses to provide security services. If a method list is defined, the NAS provides security services.
- Username forms

  A username can be expressed in four forms. In the usernames below, "domain-name" indicates a domain name.
  - userid@domain-name
  - domain-name\userid
  - userid.domain-name
  - userid

  In the preceding username forms, the username form (userid) with no domain name uses the default domain name (**default**).
- Domain name matching

  If the username of an authenticated user carries domain information but the domain is not configured on the NAS, the NAS cannot provide AAA services for the user.

  Exact match is adopted to match domain names carried in usernames with those configured on the NAS. For example, if domain.com and domain.com.cn are configured on the NAS and a username is expressed in the form of aaa@domain.com, the NAS determines that the user belongs to domain.com rather than domain.com.cn.

1.2 Restrictions and Guidelines

- Before configuring any AAA function, run the **aaa-config new-pattern** command to enable the AAA security services. Otherwise, no command can be configured.
- AAA authentication, authorization, and accounting are three independent processes. You can use only the authentication service without authorization or accounting.
- The basic methodology and process of configuring the AAA authentication, authorization, or accounting services are as follows:
  a Enable AAA security services.
  b Determine the required AAA security service type based on the scenario.
  c Configure an AAA method list for the security service type.
  d In a corresponding mode, apply an AAA method list to make it take effect (methods do not need to be applied for some security services. For details, see specific configuration tasks).

  If this step is not configured, methods configured in step C do not take effect but the default method is used.
- Configure user information before configuring security methods.

- To use the remote server group method, configure user data on remote servers. For details about how to configure user data, see the operation instructions of the remote servers.
- To use the local method, configure local user data on the local device. For details about local users, see *Configuring Basic Management*.

1.3 Configuration Task Summary

AAA configuration includes the following tasks:

(1) (Optional) Configuring an AAA Server Group

(2) Configuring AAA Authentication

- Configuring Basic Features of AAA Authentication
- (Optional) Configuring AAA Authentication Logging

(3) (Optional) Configuring AAA Authorization

- Configuring Basic Features of AAA Authorization
- (Optional) Configuring the Function of Caching Command Authorization Results
- (Optional) Configuring the Function of Allowing Users to Log in to the Slave Device

(4) (Optional) Configuring AAA Accounting

- Configuring Basic Features of AAA Accounting
- (Optional) Configuring a Policy for User Accounting Start Failures

(5) (Optional) Configuring Domain-Based AAA Services

1.4 Configuring an AAA Server Group

1.4.1 Overview

AAA usually uses the remote server group method to provide AAA services for users. This method provides more flexible and rich security services than the local method.

1.4.2 Restrictions and Guidelines

- One or more servers can be added to each server group.
- Both RADIUS and TACACS+ support server group configuration. The configuration procedure below uses RADIUS as an example. For details about a TACACS+ server group, see *Configuring TACACS*.
- When configuring the authentication, authorization, and accounting method lists, you can specify different server groups for them.
- In a custom server group, only servers in the default server group can be specified and applied.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an AAA custom server group.

aaa-config server-group server rds *group-name*

No server group is configured by default.

The name of a server group cannot be set to a predefined keyword "radius" or "tacacs+".

(4) Add an AAA server group member.

server { *ipv4-address* | *ipv6-address* } [**authentication-port** *auth-port* | **accounting-port** *acct-port*] *

No AAA server group member is added to a server group by default.

(5) (Optional) Specify a VRF instance for the AAA server group.

ip vrf binding *vrf-name*

No VRF instance is specified for an AAA server group by default.

When a server group in a specified VRF instance is used, the source IP address used by the device to communicate with a remote server needs to be acquired from the VRF instance. If you run the **ip radius-mode source-interface** command to specify the source interface of request packets, the IP address obtained from this source interface is prior to that acquired from the VRF instance.

(6) Configure the MGMT interface for the AAA server group.

ip oob [**via mgmt** *mgmt-number*]

An AAA server group uses MGMT 0 by default.

1.5 Configuring Basic Features of AAA Authentication

1.5.1 Overview

AAA authentication is used to verify user identities. After AAA authentication is configured, users can access network resources only after passing the authentication specified in security methods in the method list.

1.5.2 Restrictions and Guidelines

- After AAA authentication is configured, if no authentication method is configured and the default authentication method does not exist, users can log in through the console without authentication but other types of access users need to pass local authentication.
- If multiple methods are defined in the method list, when a method fails to respond, the device switches to the next method. The execution sequence of methods is the sequence of methods configured in the method list. If a method responds and returns a failure, the device does not switch to the next method.
- If you have passed login authentication (except the authentication exemption method) when logging in to the CLI, the device records the username, and then uses the username for authentication when Enable authentication is performed. You do not need to enter the username again but the entered password must match the username. If login authentication fails or the authentication exemption method is used when you log in to the CLI, the username will not be recorded, and you need to enter the username for authentication in the further Enable authentication.
- The authentication exemption method allows any user to access network resources without authentication. You are not advised to use this method. Only in some special cases (all access users are trustworthy and no network failure caused by authentication system problems is allowed), this method can be used as an alternative method when the device receives no response in the remote server group authentication and

local authentication.

1.5.3 Configuration Tasks

Configure basic features of AAA authentication. Select at least one of the following tasks to configure.

- Configuring Login Authentication
- Configuring Enable Authentication
- Configuring FTP Authentication

1.5.4 Configuring Login Authentication

1. Overview

After login authentication is configured, access users who log in through SSH, telnet, or FTP need to be authenticated before accessing network resources.

2. Restrictions and Guidelines

A configured login authentication method must be applied to a line that needs login authentication. Otherwise, the method does not take effect.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable AAA security services.

aaa-config new-pattern

The AAA security services are disabled by default.

(4) Configure a login authentication method list.

aaa-config authentication line-login { *list-name* | **default** } *method*&<1-4>

No login authentication method list is configured by default.

(5) (Optional) Configure the maximum number of consecutive failed login attempts.

aaa-config lock authentication times *max-attempts*

(6) The default maximum number of consecutive failed login attempts is **3**. **0** indicates that the number of consecutive failed login attempts is not limited.

(7) (Optional) Configure the user lockout duration.

aaa-config lock authentication lockout-time *lockout-time*

After the number of login attempts of a user exceeds the configured number of consecutive failed login attempts, the user is locked out for 10 minutes by default.

(8) Enter the configuration mode of a specified line.

line { **console** | **vty** } *first-line* [*last-line*]

(9) Apply the login authentication method list to the line.

line-login authentication { *list-name* | **default** }

The default login authentication method list applied to a line is the default method list.

1.5.5 Configuring Enable Authentication

1. Overview

After Enable authentication is configured and a user logs in to the NAS CLI through a terminal, the user must be authenticated before running the **enable** command to enter the privileged EXEC mode.

1. Restrictions and Guidelines

After an Enable authentication method list is configured, the Enable authentication function takes effect automatically.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable AAA security services.

aaa-config new-pattern

The AAA security services are disabled by default.

(4) Configure an Enable authentication method list.

aaa-config authentication enable reset *method*&<1-4>

No Enable authentication method list is configured by default.

1.5.6 Configuring FTP Authentication

1. Overview

After FTP authentication is configured, FTP users need to pass FTP authentication upon login.

2. Restrictions and Guidelines

If the AAA FTP security service is enabled on a device, users must perform FTP authentication through AAA.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable AAA security services.

aaa-config new-pattern

The AAA security services are disabled by default.

(4) Enable the FTP server function.

ftp-server enable

The FTP server function is disabled by default.

(5) Configure an FTP authentication method list.

aaa-config authentication ftp { *list-name* | **default** } *method*&<1-4>

No FTP authentication method list is configured by default.

(6) Configure FTP to use AAA accounts for login authentication.

ftp-server authentication { *list-name* | **default** }

FTP does not support login authentication on AAA accounts by default.

1.6 Configuring AAA Authentication Logging

1.6.1 Overview

The device generates logs when an AAA user passes authentication. If a large number of users go online concurrently, the device generates considerable logs, which may degrade the device performance or cause frequent screen refreshing. In this case, you can disable the logging function or restrict the logging rate.

1.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure AAA authentication logging. Please configure only one task.

- Disable AAA authentication logging.

  unset aaa-config log enable

  AAA authentication logging is enabled by default.

- Configure the rate of AAA authentication logging.

  aaa-config log speed-limit *rate-limit*

  Five AAA authentication logs are printed per second by default.

1.7 Configuring Basic Features of AAA Authorization

1.7.1 Overview

AAA authorization is used to manage services available to users and control network use permissions of access users. After the basic features of AAA authorization are enabled, users can use only the permitted services and are granted permissions in their permitted scopes.

1.7.2 Restrictions and Guidelines

- EXEC authorization is usually used together with login authentication. Authentication and authorization can use different methods and different servers, and the authentication and authorization results of the same user may be different. If a user passes login authentication but fails in EXEC authorization, the user cannot open the CLI.

- When configuring command authorization, specify the command level, which is used as the default level of commands. For example, if a command is visible to users above level 14, the default level of the command is 14.
- EXEC authorization and command authorization need to be applied to a line so that they take effect.
- Command authorization is unavailable when role-based access control (RBAC) is enabled. For details about RBAC, see *Configuring RBAC*.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an AAA authorization method. Configure at least one of the tasks.

- Configure an EXEC authorization method list to authorize permitted users to open the CLI.

  aaa-config aaa-authorization execute { *list-name* | **default** } *method*&<1-4>

  No EXEC authorization method list is configured by default.

- Configure a command authorization method list to authorize permitted users to run a command.

  aaa-config aaa-authorization cmd *level* { *list-name* | **default** } *method*&<1-4>

  No command authorization method list is configured by default.

- Configure a network authorization method list to authorize permitted users to access network services.

  aaa-config aaa-authorization net { *list-name* | **default** } *method*&<1-4>

  No network authorization method list is configured by default.

- Grant authorization to all commands in configuration modes.

  aaa-config aaa-authorization config-mode-commands

  The authorization for commands in configuration modes is disabled by default.

- Enable command authorization for users who log in through the console. You can enable command authorization for users who log in through the console and users who log in through other terminals separately.

  aaa-config aaa-authorization console

  Command authorization is disabled for users who log in through the console by default.

(4) (Optional) Enter the configuration mode of a specified line.

line { **console** | **vty** } *first-line* [*last-line*]

(5) (Optional) Apply the authorization method to the specific line. Please configure only one task.

- Apply the EXEC authorization method to the specific line.

  aaa-authorization execute { *list-name* | **default** }

  EXEC authorization is disabled by default.

- Apply the command authorization method to the specific line.

  aaa-authorization cmd *level* { *list-name* | **default** }

The command authorization function is disabled by default.

1.8 Configuring the Function of Caching Command Authorization Results

1.8.1 Overview

After this function is configured, the device locally caches command authorization results returned by the AAA server. For subsequent commands with the level same as the cached authorized commands, command authorization can be performed based on the locally cached results.

1.8.2 Restrictions and Guidelines

Authorization results in the local cache apply only to the current session and commands of the current level. The authorization of other sessions or commands of other levels are performed by the remote server and local authorization can be performed only after their results are cached locally.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the function of caching command authorization results.

aaa-config command-author cache

The function of caching command authorization results is disabled by default.

1.9 Configuring the Function of Allowing Users to Log in to the Slave Device

1.9.1 Overview

The function of allowing users to log in to the slave device is used to control whether a user is allowed to log in to the console of the slave device.

1.9.2 Restrictions and Guidelines

- The **aaa-config new-pattern** command must be configured first.
- The configuration remains valid unless the function is disabled explicitly on the master device.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable AAA security services.

aaa-config new-pattern

The AAA security services are disabled by default.

(4) Disable the function of allowing users to log in to the console of the slave device.

unset aaa-config slave-login allow

The function of allowing users to log in to the console of the slave device is enabled by default.

1.10 Configuring Basic Features of AAA Accounting

1.10.1 Overview

AAA accounting is used to record users' usage of network resources, login and logout processes, and command execution information.

1.10.2 Restrictions and Guidelines

- EXEC accounting is performed only on users who pass login authentication and log in to the NAS. When login authentication is not configured or the authentication exemption method is adopted, EXEC accounting is not performed.
- If accounting start is not performed on a user upon login, accounting stop will not be performed when the user logs out.
- The remote server group method used in the command accounting function supports only the TACACS+ protocol.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an AAA accounting method. Configure at least one of the tasks.

- Configure an EXEC accounting method list.

  aaa-config aaa-accounting execute { *list-name* | **default** } **start-and-stop** *method*&<1-4>

  NO EXEC accounting method list is configured by default.

  A configured EXEC accounting method list must be applied to a line so that it takes effect.

- Configure a command accounting method list.

  aaa-config aaa-accounting cmd *level* { *list-name* | **default** } **start-and-stop** *method*&<1-4>

  No command accounting method list is configured by default.

  A configured command accounting method list must be applied to a line so that it takes effect.

- Configure a network accounting method list.

  aaa-config aaa-accounting net { *list-name* | **default** } **start-and-stop** *method*&<1-4>

  No network accounting method list is configured by default.

(4) (Optional) Enable the accounting update function.

aaa-config accounting update

Accounting update is disabled by default.

After this function is configured, the device updates accounting information periodically. This function helps improve the accounting accuracy. You are advised to configure it.

(5) (Optional) Configure the accounting update interval.

aaa-config accounting update time *interval*

The default accounting update interval is **5** minutes.

It is recommended that the accounting update interval not be configured unless otherwise specified.

(6) Apply the accounting method. Please configure only one task.

- Enter the line configuration mode and apply the EXEC accounting method. Run the following commands in sequence:

  line { **console** | **vty** } *first-line* [*last-line*]

  aaa-accounting execute { *list-name* | **default** }

  A line is associated with the default method list by default.

- Enter the line configuration mode and apply the command accounting method. Run the following commands in sequence:

  line { **console** | **vty** } *first-line* [*last-line*]

  aaa-accounting cmd [*accounting-commands-level*] { *list-name* | **default** }

  A line is associated with the default method list by default.

1.11 Configuring a Policy for User Accounting Start Failures

1.11.1 Overview

After this function is configured, the user status is online or offline if the accounting start fails.

1.11.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the user online/offline status after the accounting start fails.

aaa-config accounting start-fail { **offline** | **online** }

The user online/offline status after an accounting start failure is not configured by default.

1.12 Configuring Domain-Based AAA Services

1.12.1 Overview

The domain-based AAA services classify users into different domains and the domains use different security methods to provide AAA services.

1.12.2 Restrictions and Guidelines

- The device supports a maximum of 32 domains.
- After the domain-based AAA services are enabled, if information of a user does not carry domain information, the user is assigned to the default domain **reset**. A method list for the **reset** domain must be configured on the device. Otherwise, users without domain information cannot use AAA services.

1.12.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable domain-based AAA services.

aaa-config domain enable

The domain-based AAA services are disabled by default.

(4) Create a domain and enter the domain configuration mode.

aaa-config domain { *domain-name* | **default** }

No domain is configured by default.

(5) Configure an authentication method list for the AAA domain.

authentication { **dot1x** | **enable** | **line-login** } { *list-name* | **default** }

The default authentication method list used in an AAA domain is the default method list.

(6) Configure an accounting method list for the AAA domain.

accounting { **cmd** | **execute** | **net** } { *list-name* | **default**}

The default accounting method list used in an AAA domain is the default method list.

(7) Configure an authorization method list for the AAA domain.

aaa-authorization { **cmd** | **execute** | **net** } { *list-name* | **default** }

The default authorization method list used in an AAA domain is the default method list.

(8) (Optional) Set the domain status.

state { **active** | **block** }

A configured domain is active by default.

(9) (Optional) Configure whether usernames carry domain name information.

username-format { **with-domain** | **without-domain** }

Usernames carry domain name information during interaction between the NAS and the server by default.

(10) (Optional) Configure a limit on the number of users supported in the domain.

access-limit *access-limit-number*

No limit on the number of users supported in a domain is configured by default.

1.13 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 AAA Monitoring

Command	Purpose
display aaa-info accounting update	Displays accounting update information.
display aaa-info capability	Displays display AAA capabilities.
display aaa-info domain [**reset** | *domain-name*]	Displays information about a configured domain.
display aaa-info lockout	Displays lockout parameter configuration in the current login authentication.
display aaa-info group	Displays all AAA server groups.
display aaa-info methodlists	Displays all AAA method lists.
display aaa-info user { **all** | **by-id** *session-id* | **by-name** *user-name* | **lockout** }	Displays AAA user information.
cleanup aaa-config user lock list { **all** | **user-name** *user-name-id* }	Clears the list of locked users.

1.14 Configuration Examples

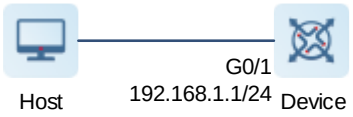
1.14.1 Configuring Local Login Authentication and Local EXEC Authorization

1. Requirements

You can use local login authentication and local EXEC authorization together to control the permissions of users who log in to the device. In this method, you need to configure accounts and their permissions only in the local user database of the device, to control user permissions with no need to use an authentication server.

2. Topology

Figure 1-2 Topology of Local Login Authentication and Local EXEC Authorization



3. Notes

- Enable AAA security services on the device.
- Add accounts to the local user database of the device.
- Create an authentication method list and an authorization method list.
- Apply the authentication method and authorization method to a specific line.

4. Procedure

Enable AAA security services.

```
Device> privilege
Device# config-system terminal
Device(cfg)# aaa-config new-pattern
```

Add accounts **host1** and **host2** to the local user database of the device and configure account permissions.

```
Device(cfg)# user-account host1 privilege 15 password password1
Device(cfg)# user-account host2 privilege 4 password password2
```

Create a login authentication method list named **login-method**, in which the authentication method is local authentication, and create an EXEC authorization method list named **exec-method**, in which the authorization method is local authorization.

```
Device(cfg)# aaa-config authentication line-login login-method native
Device(cfg)# aaa-config authorization execute exec-method native
```

Apply the authentication method list and authorization method list to a line.

```
Device(cfg)# line vty 0 35
Device(cfg-line)# line-login authentication login-method
Device(cfg-line)# aaa-authorization execute exec-method
```

5. Verification

Run the **display aaa-info methodlists** command on the device to display configured method lists.

```
Device# display aaa-info methodlists

Authentication methodlists:
aaa-config authentication line-login login-method native

Accounting methodlists:

Authorization methodlists:
aaa-config aaa-authorization execute exec-method native
```

Log in to the device through telnet by using client software (such as PuTTY or Xshell) and verify that different users have different account permissions.

- Log in to the device as user **host1** and check that the privilege level of the user is 15.

```
User Access Verification

Username:host1
Password:password1



Device# display privilege
Current privilege level is 15
```

- Log in to the device as user **host2** and check that the privilege level of the user is 4.

```
User Access Verification

Username:host2
Password:password2



Device# display privilege
Current privilege level is 4
```

6. Configuration Files

```
!
user-account host1 privilege 15 password password1
user-account host2 privilege 4 password password2
!
aaa-config new-pattern
!
aaa-config authorization execute exec-method native
aaa-config authentication line-login login-method native
!
line vty 0 35
 aaa-authorization execute exec-method
 line-login authentication login-method
!
end
```

7. Common Errors

- No authentication method list and authorization method list are applied to a specific line.
- No account is configured in the local user database.

1.14.2 Configuring Enable Authentication

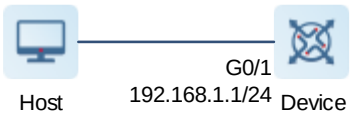
1. Requirements

Users can have execution permissions with a privilege level of 0 to 15. After Enable authentication is configured, users must be authenticated when running the **privilege** command to switch the permission to a higher level.

As shown in Figure 1-3, when both local authentication and Enable password authentication are configured on the device, local authentication is preferred. If local authentication is invalid, privilege cipher authentication is used.

2. Topology

Figure 1-3 Topology of Enable Authentication



3. Notes

- Enable AAA security services on the device.
- Configure the local user database and Enable password.
- Configure an Enable authentication method list.

Note

Only one Enable authentication method list can be defined. Therefore, you do not need to define the name of the Enable authentication method list. After the default method list is configured, the Enable authentication function takes effect automatically.

4. Procedure

Enable AAA security services.

```
Device> privilege
Device# config-system terminal
Device(cfg)# aaa-config new-pattern
```

Add account **host1** to the local user database and set the privilege level of the account to **15**.

```
Device(cfg)# user-account host1 privilege 15 password password1
```

Set the password to **enpassword1** for privilege level 15 in Enable authentication.

```
Device(cfg)# privilege safe-cipher level 15 enpassword1
```

Configure an Enable authentication method list, in which local authentication is prior to Enable password authentication.

```
Device(cfg)# aaa-config authentication enable default native enable
```

5. Verification

Log in to the device as user **host1** and check that the privilege level of the user is 15.

```
User Access Verification

Username:host1
Password:password1
```

```
Device# display privilege
Current privilege level is 15
```

Switch the privilege level to 4 and verify that no password is required for authentication.

```
Device# privilege 4
Device# display privilege
Current privilege level is 4
```

Switch the privilege level to 15 and verify that identity authentication is needed. Local authentication has a higher priority. Therefore, you can enter the password of **host1** (**password1**) for authentication to successfully switch the privilege level to 15.

```
Device# privilege 15
Password:password1

Device# display privilege
Current privilege level is 15
```

6. Configuration Files

```
!
user-account host1 privilege 15 password password1
!
aaa-config new-pattern
!
aaa-config authentication enable default native enable
!
privilege safe-cipher 5 $1$7b8Y$v6u11F23p9y3vwp6
!
end
```

1.14.3 Configuring Command Authorization

1. Requirements

Command authorization needs to be configured for login users so that users who log in to the device through terminals can only use commands in their authorized scopes.

2. Topology

Figure 1-4 Topology of Command Authorization



3. Notes

- Enable AAA security services on the device.
- Configure an IP address and a shared key for the TACACS+ server on the device.

- Configure the login authentication method list and command authorization method list on the device.
- Configure to-be-authenticated users and authorized command scopes on the TACACS+ server.
- Apply the authentication method and authorization method to a specific line.

4. Procedure

(1) On the TACACS+ server, add login users, configure authorized command scopes, and configure the server to provide AAA services for the device. (The configuration is omitted here. For the configuration of the TACACS+ server, use the operation manual of the server.)

(2) Enable AAA security services.

```
Device> privilege
Device# config-system terminal
Device(cfg)# aaa-config new-pattern
```

(3) Configure the IP address and shared key of the TACACS+ server (the IP address is set to **10.110.1.2** and shared key to **sharekey** in this example).

```
Device(cfg)# tacplus-server host 10.110.1.2 share-key sharekey
```

(4) Configure a login authentication method list named **login-method**, in which TACACS+ server authentication is prior to local authentication.

```
Device(cfg)# aaa-config authentication line-login login-method server-group tacacs+ native
```

(5) Add the account **host1** to the local user database and configure the account permissions.

```
Device(cfg)# user-account host1 privilege 15 password password1
```

(6) Configure a command authorization method list named **command-method** for level-15 commands. In this list, TACACS+ server authorization is prior to local authorization.

```
Device(cfg)# aaa-config authorization cmd 15 command-method server-group tacacs+ native
```

(7) Apply the authentication method list and authorization method list to a line.

```
Device(cfg)# line vty 0 35
Device(cfg-line)# line-login authentication login-method
Device(cfg-line)# aaa-authorization execute command-method
```

5. Configuration Files

```
!
aaa new-model
!
aaa authentication login login-method group tacacs+ local
aaa authorization commands 15 command group tacacs+ local
!
tacacs-server host 10.110.1.2 key 7 $10$062$/SJrCPEEZKM=$
!
username host1 privilege 15 password 7 $10$247$knWuVat9D+rHyNk=$
!
line vty 0 35
 login authentication login-method
 authorization exec command-method
!
```

6. Common Errors

- The shared key configured on the TACACS+ server is different from that configured on the device.
- A command authorization configuration error occurs on the TACACS+ server.

1.14.4 Configuring Network Authorization

1. Requirements

Network authorization needs to be configured to grant available network services to users, such as the traffic, bandwidth, and timeout services.

2. Topology

Figure 1-5 Topology of Network Authorization



3. Notes

- Enable AAA security services.
- Configure a network authorization method list.
- Apply the network authorization method list to a specific interface line (skip this step if you configure the default method).

4. Procedure

Enable AAA security services.

```
Device> privilege
Device# config-system terminal
Device(cfg)# aaa-config new-pattern
```

Configure the default network authorization method list.

```
Device(cfg)# aaa-config authorization net default server-group radius none
```

5. Verification

Run the **display aaa-info methodlists** command on the device to display configuration results.

```
Device# display aaa-config methodlists

Authentication methodlists:

Accounting methodlists:

Authorization methodlists:
aaa-config aaa-authorization net efault server-group radius none
```

6. Configuration Files

```
!
aaa-config new-pattern
!
aaa-config aaa-authorization net default server-group radius none
!
```

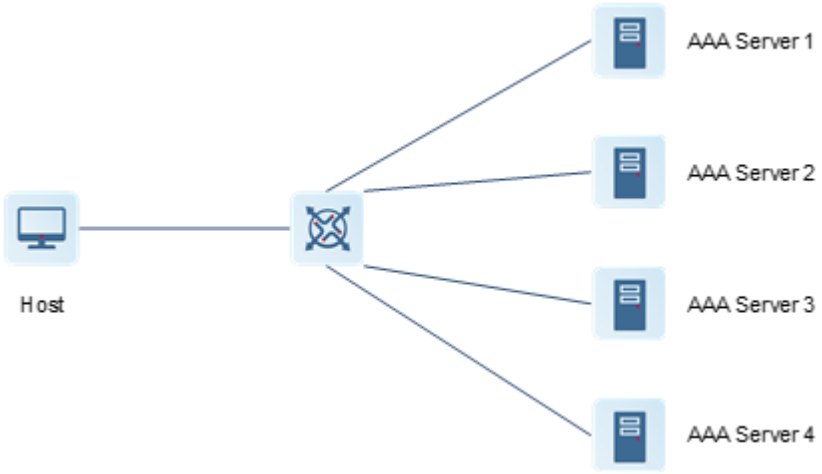
1.14.5 Configuring an AAA Server Group

1. Requirements

A custom AAA server group needs to be created to provide AAA services.

2. Topology

Figure 1-6 Topology of an AAA Server Group



3. Notes

- Configure AAA servers.
- Create an AAA custom server group.
- Add server group members to the custom server group.

4. Procedure

(1) Add AAA servers (the following uses RADIUS servers as an example).

```
Device> privilege
Device# config-system terminal
Device(cfg)# rds-server host 10.1.1.1
Device(cfg)# rds-server host 10.1.1.2
Device(cfg)# rds-server host 10.1.1.3
Device(cfg)# rds-server host 10.1.1.4
```

(2) Configure a global RADIUS server communication key.

```
Device(cfg)# rds-server share-key radiuskey
```

(3) Create a server group named **group1** and add members to the group.

```
Device(cfg)# aaa-config group server rds group1
Device(cfg-gs-radius)# server 10.1.1.1
Device(cfg-gs-radius)# server 10.1.1.2
Device(cfg-gs-radius)# exit
```

(4) Create a server group named **group2** and add members to the group.

```
Device(cfg)#aaa-config group server rds group2
Device(cfg-gs-radius)# server 10.1.1.3
Device(cfg-gs-radius)# server 10.1.1.4
Device (cfg-gs-radius)# exit
```

5. Verification

Run the **display aaa-info group** command on the device to display the configuration results.

```
Device# display aaa-info group
Type            Reference    Name
---------- ---------- ----------
radius         1              radius
tacacs+        1               tacacs+
radius         1              group1
radius         1              group2
```

6. Configuration Files

```
!
rds-server host 10.1.1.1
rds-server host 10.1.1.2
rds-server host 10.1.1.3
rds-server host 10.1.1.4
rds-server share-key radiuskey
!
aaa-config group server rds group1
 server 10.1.1.1
 server 10.1.1.2
!
aaa-config group server rds group2
 server 10.1.1.3
 server 10.1.1.4
!
```

7. Common Errors

For RADIUS servers that use non-default authentication ports and non-default accounting ports, the authentication port or accounting port is not specified when the **server** command is run to add servers.

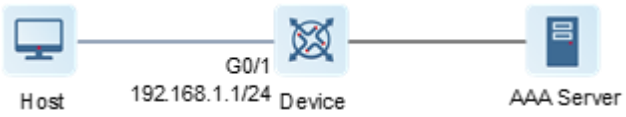
1.14.6 Configuring Domain-based AAA Services

1. Requirements

After a domain is created and a security method list is specified for the domain, users belonging to the domain use security methods of the domain for authentication, authorization, and accounting.

2. Topology

Figure 1-7 Topology of Domain-Based AAA Services



3. Notes

- Configure a RADIUS server in advance so that RADIUS server authentication and accounting can be performed.
- Enable AAA security services.
- Configure an AAA security method list.
- Enable domain-based AAA services.
- Create a domain.
- Associate the AAA security method list with the domain.
- Configure domain attributes.

4. Procedure

(1) Enable AAA security services.

```
Device> privilege
Device# config-system terminal
Device(cfg)# aaa-config new-pattern
```

(2) Configure an AAA server (the following uses the RADIUS server as an example).

```
Device(cfg)# rds-server host 192.168.1.2 share-key radiuskey
```

(3) Configure an AAA security method list.

```
Device(cfg)# aaa-config authentication line-login default server-group radius
Device(cfg)# aaa-config aaa-accounting net list3 start-and-stop server-group radius
```

(4) Enable domain-based AAA services.

```
Hostname(cfg)# aaa-config domain enable
```

(5) Create a domain and configure domain attributes.

```
Device(cfg)# aaa-config domain domain.com
Device(cfg-aaa-domain)# authentication line-login reset
Device(cfg-aaa-domain)# aaa-accounting net list3
Device(cfg-aaa-domain)# username-format without-domain
```

5. Verification

Run the **display aaa-config domain** command on the device to display domain configuration.

```
Device# display aaa-config domain domain.com

=============Domain domain.com=============
```

```
State: Active
Username format: With-domain
Access limit: No limit

Selected method list:
 authentication line-login reset
 aaa-accounting net list3
```

6. Configuration Files

```
!
aaa-config new-pattern
aaa-config domain enable
!
aaa-config domain domain.com
 authentication line-login default
 aaa-accounting net list3
 username-format without-domain
!
aaa-config accounting net list3 start-and-stop server-group radius
aaa-config authentication dot1x default server-group radius
!
radius-server host 192.168.1.2 key radiuskey
!
```

1.15 Common Errors

- When a local method is configured to provide AAA security services, no user information is configured in the local database.
- When a remote server group method is configured to provide AAA security services, no RADIUS/TACACS+ server is configured and no user information is configured on the server.

Contents

1 Configuring RADIUS 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 8
1.2 Configuration Task Summary 8
1.3 Configuring RADIUS Basic Features 8
1.3.1 Overview 8
1.3.2 Restrictions and Guidelines 8
1.3.3 Procedure 9
1.4 Configuring a RADIUS Server Group 10
1.4.1 Overview 10
1.4.2 Restrictions and Guidelines 10
1.4.3 Procedure 10
1.5 Configuring RADIUS Attributes 11
1.5.1 Overview 11
1.5.2 Configuration Tasks 11
1.5.3 Configure the Attribute Format 11
1.5.4 Configuring Private Attributes 12
1.5.5 Configuring Compatibility with Devices of Other Vendors 13
1.6 Configuring RADIUS Reachability Detection 14
1.6.1 Overview 14
1.6.2 Restrictions and Guidelines 14

1.6.3 Procedure....15
1.7 Configuring the DSCP Value for RADIUS Packets....15
1.7.1 Overview....15
1.7.2 Procedure....15
1.8 Configuring the Units of Data Flows and Packets to Be Sent to a RADIUS Server....16
1.8.1 Overview....16
1.8.2 Procedure....16
1.9 Configuring the Accounting-On Function....16
1.9.1 Overview....16
1.9.2 Procedure....16
1.10 Configuring the Accounting Packet Copying Function....17
1.10.1 Overview....17
1.10.2 Procedure....17
1.11 Monitoring....17
1.12 Configuration Examples....18
1.12.1 Configuring RADIUS Authentication, Authorization, and Accounting....18
1.12.2 Configuring RADIUS Reachability Detection....20

1 Configuring RADIUS

1.1 Introduction

1.1.1 Overview

The Remote Authentication Dial-In User Service (RADIUS) is a distributed authentication, authorization and accounting (AAA) protocol based on the client/server architecture. It works with AAA to provide AAA security services for users.

When providing AAA security services, a network device serves as a RADIUS client to provide user information to the RADIUS server to initiate security service requests. The remote RADIUS server stores user information and network service information, and provides security services for users by responding to requests from the RADIUS client.

RADIUS is usually applied in networks that have high security requirements and allow remote user access. As a completely open protocol, many systems (such as UNIX, Windows 2000, and Windows 2008) support the installation of RADIUS as a system component. Therefore, the RADIUS server becomes the most widely used security server.

1.1.2 Principles

1. Basic Concepts

- Client/server mode

  RADIUS uses the client/server mode and consists of RADIUS clients and the RADIUS server.

  - As the initiator of RADIUS requests, a RADIUS client usually runs on a network access server (NAS). It provides user information for the RADIUS server, and receives and processes responses from the RADIUS server, for example, accepting or rejecting user access, and requesting user-specific information.
  - The RADIUS server maintains the IP addresses and shared key of RADIUS clients as well as configured user information, and provides authentication, authorization, and accounting services by responding to requests from RADIUS clients. One RADIUS server usually corresponds to multiple RADIUS clients.

- Shared key

  A shared key is a string shared between a RADIUS client and the server. It is used to encrypt and decrypt user passwords during RADIUS packet transmission and generate packet verification fields, in an effort to ensure RADIUS communication security.

- RADIUS server group

  Security methods used in AAA security services are based on server groups. Likewise, RADIUS provides AAA security services based on RADIUS server groups. One security method corresponds to one RADIUS server group and one or more RADIUS severs can be configured in each RADIUS server group.

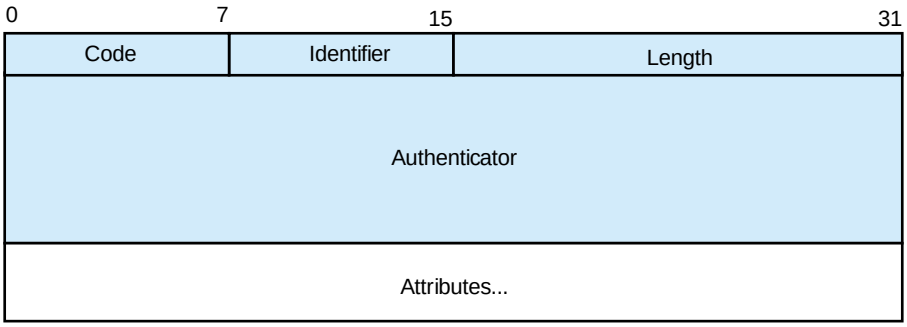
  If multiple RADIUS servers are configured in a RADIUS server group, when a device fails to communicate with the first RADIUS server or the first RADIUS server is unreachable, the device automatically attempts

to communicate with the second RADIUS server, and so on. The attempt lasts till the communication succeeds or the communication with all RADIUS servers fails.

2. RADIUS Packet

- Packet structure

Figure 1-1 Structure of a RADIUS Packet



Fields contained in the packet structure shown in Figure 1-1 are described in Table 1-1.

Table 1-1 Description of Fields Contained in a RADIUS Packet

Field	Length	Description	Remarks
Code	1 byte	Identifies the packet type.	Values corresponding to common packet types and their meanings are as follows: **1**: Access-Request **2**: Access-Accept **3**: Access-Reject **4**: Accounting-Request **5**: Accounting-Response **11**: Access-Challenge
Identifier	1 byte	Matches a request with a response packet.	Request packets and response packets of the same type have the same identifier value.
Length	2 bytes	Indicates the length of a RADIUS packet.	The value is the sum of the lengths of the **Code**, **Identifier**, **Length**, **Authenticator**, and **Attributes** fields. If the length of a received packet is greater than the value of **Length**, the content beyond the value of **Length** is ignored as filling information. If the length of a received packet is smaller than the value of **Length**, the packet is discarded.

Field	Length	Description	Remarks
Authenticator	16 bytes	Verifies packets and encrypts and decrypts user passwords.	N/A
Attributes	Variable length	Carries authentication, authorization, and accounting information.	This field usually contains multiple attributes and each attribute is represented in the type, length, value (TLV) format.

- RADIUS attributes

  RADIUS attributes carry authentication, authorization, and accounting information, and each attribute is represented in the TLV format, as described in Table 1-2.

Table 1-2 Description of RADIUS Attribute Fields

Field	Length	Description
Type	1 byte	Attribute type
Length	1 byte	Total length of an attribute (TLV)
Value	Variable length	Attribute description

Table 1-3 RADIUS Attribute List

Attribute No.	Attribute Name	Attribute No.	Attribute Name
1	User-Name	43	Acct-Output-Octets
2	User-Password	44	Acct-Session-Id
3	CHAP-Password	45	Acct-Authentic
4	NAS-IP-Address	46	Acct-Session-Time
5	NAS-Port	47	Acct-Input-Packets
6	Service-Type	48	Acct-Output-Packets
7	Framed-Protocol	49	Acct-Terminate-Cause
8	Framed-IP-Address	50	Acct-Multi-Session-Id
9	Framed-IP-Netmask	51	Acct-Link-Count
10	Framed-Routing	52	Acct-Input-Gigawords
11	Filter-ID	53	Acct-Output-Gigawords
12	Framed-MTU	55	Event-Timestamp
13	Framed-Compression	60	CHAP-Challenge

Attribute No.	Attribute Name	Attribute No.	Attribute Name
14	Login-IP-Host	61	NAS-Port-Type
15	Login-Service	62	Port-Limit
16	Login-TCP-Port	63	Login-LAT-Port
18	Reply-Message	64	Tunnel-Type
19	Callback-Number	65	Tunnel-Medium-Type
20	Callback-ID	66	Tunnel-Client-Endpoint
22	Framed-Route	67	Tunnel-Server-Endpoint
23	Framed-IPX-Network	68	Acct-Tunnel-Connection
24	State	69	Tunnel-Password
25	Class	70	ARAP-Password
26	Vendor-Specific	71	ARAP-Features
27	Session-Timeout	72	ARAP-Zone-Access
28	Idle-Timeout	73	ARAP-Security
29	Termination-Action	74	ARAP-Security-Data
30	Called-Station-Id	75	Password-Retry
31	Calling-Station-Id	76	Prompt
32	NAS-Identifier	77	Connect-Info
33	Proxy-State	78	Configuration-Token
34	Login-LAT-Service	79	EAP-Message
35	Login-LAT-Node	80	Message-Authenticator
36	Login-LAT-Group	81	Tunnel-Private-Group-id
37	Framed-AppleTalk-Link	82	Tunnel-Assignment-id
38	Framed-AppleTalk-Network	83	Tunnel-Preference
39	Framed-AppleTalk-Zone	84	ARAP-Challenge-Response
40	Acct-Status-Type	85	Acct-Interim-Interval
41	Acct-Delay-Time	86	Acct-Tunnel-Packets-Lost
42	Acct-Input-Octets	87	NAS-Port-Id

- Introduction to RADIUS attributes

  This section describes some of the important attributes.

  - Attribute 31

  When a RADIUS client sends a request packet to the RADIUS server, information is filled in attribute 31 (**Calling-Station-ID**) to identify the user to be authenticated. The MAC address of a user is often used as the attribute content. For example, in IEEE 802.1X authentication, the device uses the MAC address of the device where the IEEE 802.1X client resides as the attribute content.

Different RADIUS servers have different requirements for the format of the MAC address in attribute 31. Therefore, ensure that the MAC address format used by RADIUS clients is the same as that used by the server. Currently, the following three MAC address formats are supported.

Table 1-4 Supported Formats of the MAC Address in Attribute 31

MAC Address Format	Description
ietf	Standard format specified in the Internet Engineering Task Force (IETF) standard (RFC3580). It uses hyphens (-) as the separator, for example, 00-D0-F8-33-22-AC.
normal	Dotted hexadecimal format using dots (.) as the separator, for example, 00d0.f833.22ac.
unformatted	Default format with no separator, for example, 00d0f83322ac.

- Attribute 26

Attribute 26 (**Vendor-Specific**) is also called a private attribute, which allows device vendors to define attributes. Device vendors can define multiple attributes in the TLV format in attribute 26. Table 1-5 lists private attributes supported by our company's products.

Table 1-5 List of Private Attributes

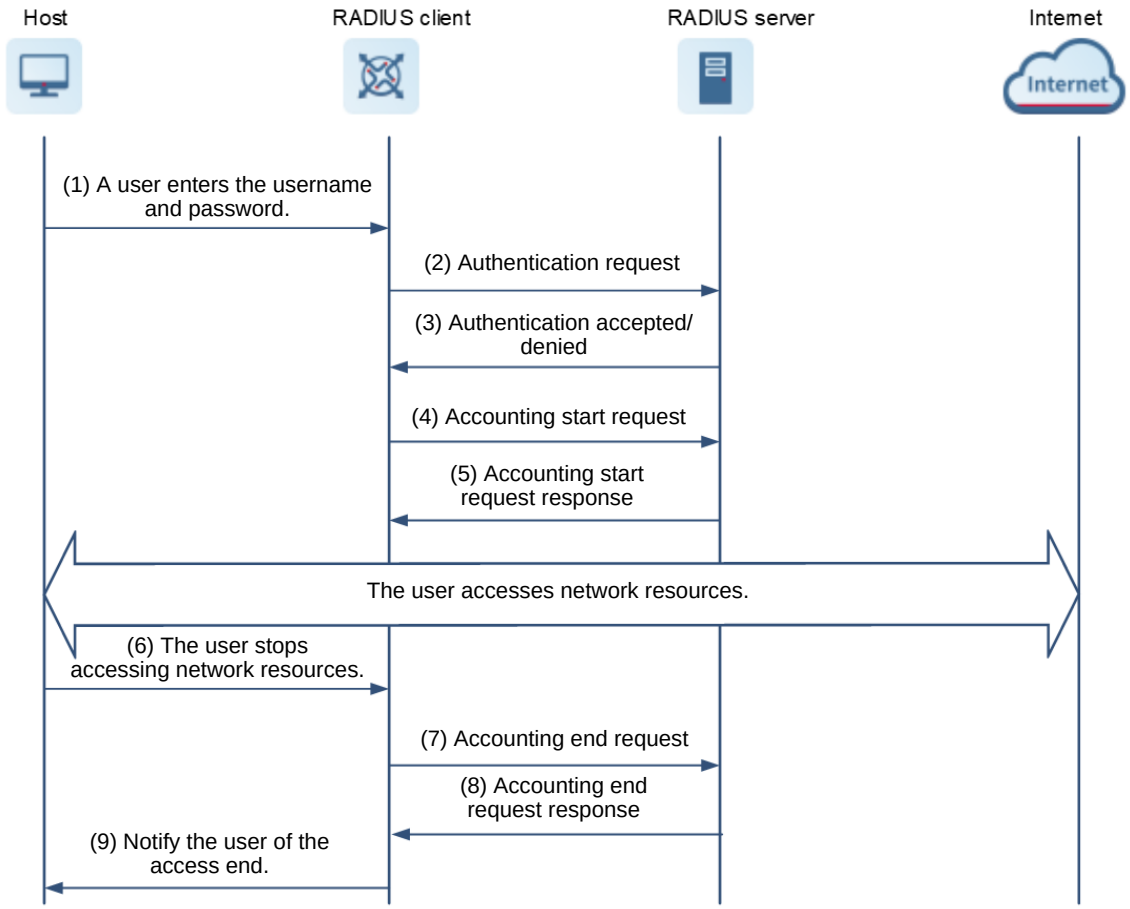
ID	Function	Type	Extended Type
1	max-down-rate	1	76
2	port-priority	2	77
3	user-ip	3	3
4	VLAN-id	4	4
5	last-supplicant-version	5	5
6	net-ip	6	6
7	user-name	7	7
8	password	8	8
9	file-directory	9	9
10	file-count	10	10
11	file-name-0	11	11
12	file-name-1	12	12
13	file-name-2	13	13
14	file-name-3	14	14
15	file-name-4	15	15
16	max-up-rate	16	16
17	current-supplicant-version	17	17
18	flux-max-high32	18	18
19	flux-max-low32	19	19
20	proxy-avoid	20	20

ID	Function	Type	Extended Type
21	dailup-avoid	21	21
22	ip-privilege	22	22
23	login-privilege	42	42
26	IPv6-multicast-address	79	79
27	IPv4-multicast-address	87	87
62	sdg-type	62	62
85	sdg-zone-name	85	85
103	sdg-group-name	103	103

3. RADIUS Authentication, Authorization, and Accounting

RADIUS provides authentication, authorization, and accounting for users. The interaction process is shown in Figure 1-2.

Figure 1-2 RADIUS Authentication, Authorization, and Accounting Process



- RADIUS authentication and authorization process

  a A user enters the username and password and sends them to a RADIUS client.

b After receiving the username and password, the RADIUS client sends an authentication request packet to the RADIUS server. The password in the authentication request packet is encrypted.

c After receiving the authentication request, the RADIUS server checks whether the username and password are legitimate. If yes, the RADIUS server accepts the authentication request and delivers authorization information of the user. If no, the RADIUS server rejects the authentication request.

- RADIUS accounting process

a If the RADIUS server returns an authentication success message in the user authentication process, the RADIUS client sends an accounting start request packet to the RADIUS server.

b The RADIUS server returns an accounting start response packet and starts accounting.

c When stopping accessing network resources, the user sends a disconnection request to the RADIUS client.

d The RADIUS client sends an accounting end request packet to the RADIUS server.

e The RADIUS server returns an accounting end response packet and stops accounting.

f The user is disconnected from the network and cannot access network resources.

4. RADIUS Packet Retransmission Upon Timeout

After sending a packet to the RADIUS server, a RADIUS client uses a timer to test whether the RADIUS server responds within the specified time. If the RADIUS server does not respond within the specified time, the RADIUS client retransmits the packet. The following timeout and retransmission parameters need to be configured:

- RADIUS server timeout duration

  The response time of a RADIUS server depends on its performance and the network environment. Configure a proper timeout duration based on the actual situation.

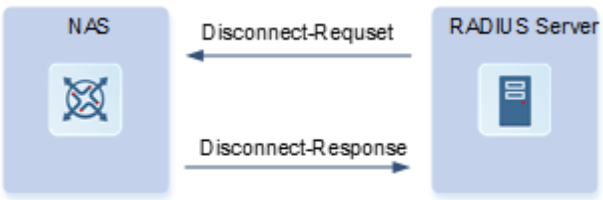
- Retransmission count
- Whether accounting update packets need to be retransmitted

5. RADIUS Forcible Logoff

Users can go offline actively. In addition, the RADIUS server can send messages to force authenticated users offline.

RADIUS forcible logoff is a user logoff management method defined in the protocol of Dynamic Authorization Extensions to RADIUS. The NAS and RADIUS server exchange Disconnect-Messages (DMs) to force authenticated users offline. The protocol of Dynamic Authorization Extensions to RADIUS enables NASs from different vendors to be compatible with RADIUS servers in user logoff processing.

Figure 1-3 DM Interaction of Dynamic Authorization Extensions to RADIUS



The DM interaction between a RADIUS server and the NAS is as follows:

(1) The RADIUS server actively sends a Disconnect-Request message to port UDP3799 of the NAS to initiate a user logoff request.

(2) The NAS searches for the user based on the user session information and username carried in the Disconnect-Request message and logs off the user.

(3) The NAS sends the logoff result to the RADIUS server through a Disconnect-Response message.

1.1.3 Protocols and Standards

- RFC 2865: Remote Authentication Dial In User Service (RADIUS)
- RFC 2866: RADIUS Accounting
- RFC 2867: RADIUS Accounting Modifications for Tunnel Protocol Support
- RFC 2868: RADIUS Attributes for Tunnel Protocol Support
- RFC 2869: RADIUS Extensions
- RFC 3576: Dynamic Authorization Extensions to Remote Authentication Dial In User Service (RADIUS)

1.2 Configuration Task Summary

RADIUS configuration includes the following tasks:

(1) Configuring RADIUS Basic Features

(2) Configuring a RADIUS Server Group

(3) (Optional) Configuring RADIUS Attributes Select at least one of the following tasks to configure.

- Configure the Attribute Format
- Configuring Private Attributes
- Configuring Compatibility with Devices of Other Vendors

(4) (Optional) Configuring RADIUS Reachability Detection

(5) (Optional) Configuring the DSCP Value for RADIUS Packets

(6) (Optional) Configuring the Units of Data Flows and Packets to Be Sent to a RADIUS Server

(7) (Optional) Configuring the Accounting-On Function

(8) (Optional) Configuring the Accounting Packet Copying Function

1.3 Configuring RADIUS Basic Features

1.3.1 Overview

This section describes how to configure parameters for the communication between a network device and a RADIUS server to ensure normal communication between them. After the parameters are configured, the device can correctly send AAA authentication, authorization, and accounting requests to the RADIUS server and give responses.

1.3.2 Restrictions and Guidelines

- Before configuring RADIUS, ensure that the network communication between the device and the RADIUS

server is in good condition.

- Before configuring RADIUS IPv6 authentication, ensure that the RADIUS server supports RADIUS IPv6 authentication.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a RADIUS remote server.

rds-server host [**oob** [**via Mgmt** *Mgmt_number*]] { *ipv4-address* | *ipv6-address* } [**authentication-port** *auth-port-number* | **accounting-port** *acct-port-number*] * [**detect user-name** *username* [**ignore-accounting-port**] [**ignore-authentication-port**] [**timeout** *idle-time*]] [**share-key** [**0** | **7**] *text-string*]

No RADIUS remote server is configured by default.

If a RADIUS server is not added to a RADIUS server group, the device uses the global routing table when communicating with the RADIUS server. Otherwise, the device uses the virtual routing and forwarding (VRF) routing table of the RADIUS server group.

(4) Configure a shared key for the communication between the device and the RADIUS server.

rds-server share-key [**0** | **7**] *key*

No shared key for the communication between the device and the RADIUS server is configured by default.

A shared key is the basis for correct communication between the device and the RADIUS server. The same shared key must be configured on them both to ensure normal communication between them.

(5) (Optional) Configure the source address for RADIUS packets.

ip radius-mode source-interface *interface-type interface-number*

The source IP address of RADIUS packets is set by the network layer by default.

After this command is configured, the device uses the first IP address of the interface specified in the command as the source address of RADIUS packets. Ensure that the communication between the configured address and the RADIUS server is normal. Specifying the source IP address for packets to be sent to the RADIUS server can reduce the NAS information maintenance workload on the RADIUS server.

(6) (Optional) Configure the source UDP port to be used by the device to send RADIUS packets.

rds-server source-port *source-port*

The source UDP port used by the device to send RADIUS packets is a random packet by default.

(7) (Optional) Configure the waiting time, after which the device retransmits a RADIUS request packet.

rds-server timeout *timeout*

The default waiting time before the retransmission of a RADIUS packet is **5** seconds.

(8) (Optional) Configure the accounting update packet retransmission function for Web-authenticated users.

rds-server account update retransmit

The accounting update packet retransmission function is enabled for Web-authenticated users by default.

The configuration does not affect users of other authentication types.

(9) (Optional) Enable the function of copying and sending RADIUS accounting packets to servers in a specified server group.

rds-server accounting-copy *group*

By default, the function of copying and sending RADIUS accounting packets is disabled.

When this command is run on an interface, the RADIUS accounting packets on the interface are copied and sent to servers in a specified server group.

(10) (Optional) Enable compatibility with username encapsulation format.

radius user-name compatible

Compatibility with username encapsulation format is disabled by default.

After this function is enabled, the username in the RADIUS packet can be encapsulated in its original format instead of UTF-8.

(11) (Optional) Configure the value of the **NAS-IP-Address** field in RADIUS packets.

rds-server nas-ip-address { *ipv4-address* | *ipv6-address* }

By default, the value of the **NAS-IP-Address** field is not configured.

1.4 Configuring a RADIUS Server Group

1.4.1 Overview

One or more RADIUS servers can be added to each RADIUS server group, which provides AAA services for users as a whole.

1.4.2 Restrictions and Guidelines

- RADIUS security services are a type of AAA services and need to be used in combination with AAA features. RADIUS provides security services for users as one method in an AAA method list.
- When configuring the authentication and accounting method lists, you can specify server groups separately for them.
- In a user-defined server group, you can only specify and apply servers in the default server group.
- RADIUS server groups support VRF instances. When a server group in a specified VRF instance is used, the source address used by the device to communicate with a remote server must be obtained from the VRF instance. If you run the **ip radius-mode source-interface** command to specify the source interface for request packets, the IP address obtained from this source interface takes priority over that found in the VRF instance.
- The name of a server group cannot be set to the predefined keyword **radius**.
- If a server group is configured with an NAS IP address, the NAS IP address is used for RADIUS packet encapsulation preferentially. If not, the IP address of the interface sending the packet is used.
- The NAS IP address configured on the server group

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a RADIUS server group.

aaa-config group server rds *group-name*

No RADIUS server group is configured by default.

You can group RADIUS servers so that authentication and accounting can be completed by different server groups.

(4) Add a server to the RADIUS server group.

server { *ipv4-address* | *ipv6-address* } [**authentication-port** *acct-port* | **accounting-port** *auth-port*] *

No server is added to a RADIUS server group by default.

(5) (Optional) Specify a VRF instance for the RADIUS server group.

ip vrf binding *vrf-name*

No VRF instance is specified for a RADIUS server group by default.

The VRF instance specified for a RADIUS server group must use a valid name configured using the **vrf definition** command in global configuration mode.

(6) (Optional) Configure an MGMT port to be used by the RADIUS server group.

ip oob [**via mgmt** *mgmt-number*]

No MGMT port to be used by a RADIUS server group is configured by default.

(7) (Optional) Configure the NAS IP address of the RADIUS server group.

nas-ip-address *ip-addr*

No NAS IP address of the RADIUS server group is configured by default.

1.5 Configuring RADIUS Attributes

1.5.1 Overview

RADIUS attributes are used to carry authentication, authorization, and accounting information to be used in packet interaction. You can configure the attribute format, private attribute processing method, and vendor compatibility.

1.5.2 Configuration Tasks

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configure the Attribute Format
- Configuring Private Attributes
- Configuring Compatibility with Devices of Other Vendors

1.5.3 Configure the Attribute Format

1. Overview

This section describes how to configure the attribute format to be used in encapsulation and parsing.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure RADIUS attributes. Configure at least one of the tasks.

- Configure the format of the MAC address used in attribute 31 (**Calling-Station-ID**).

  ds-server attribute 31 mac-addr format { **3hyphen** | **standard** | **with-dot** | **none-format** | { { **colon-split | dot-split | hyphen-split** } **{ mode1 | mode2** } [**lowercase | uppercase**] } }

  The MAC address used in attribute 31 uses the unformatted pattern.

  Some RADIUS servers (mainly used for IEEE 802.1X authentication) can identify only MAC addresses in the IETF format. In this case, set the format of MAC addresses used in attribute 31 to the IETF format.

- Configure the encapsulation format for the **NAS-Port-ID** attribute.

  rds-server attribute nas-port-id format { **mode**1 | **with-dot** | **port-vid** | **qinq** }

  The default encapsulation format of the **NAS-Port-ID** attribute is the normal format.

  Use this command to configure the encapsulation format for the **NAS-Port-ID** attribute applicable to 802.1Q in 802.1Q (QinQ) scenarios or non-QinQ scenarios.

- Configure the parsing mode for the **Class** attribute.

  rds-server attribute class user-flow-control { **format-16bytes** | **format-32bytes** | **unit bit/s** | **unit byte/s** }

  The function of parsing the rate limit configuration from the **Class** attribute of RADIUS packets is disabled by default.

  Configure this command if a server needs to deliver the rate limit value by using the **Class** attribute.

- Configure whether RADIUS authentication request packets carry a specified attribute.

  rds-server authentication attribute *type* { **package** | **unpackage** }

  According to the RFC standard, attributes that must be carried in authentication requests are carried by default, attributes that do not need to be carried are not carried by default, and other attributes are in the unset state by default.

- Configure whether RADIUS accounting request packets carry a specified attribute.

  rds-server account attribute *type* { **package | unpackage** }

  According to the RFC standard, attributes that must be carried in accounting requests are carried by default, attributes that do not need to be carried are not carried by default, and other attributes are in the unset state by default.

1.5.4 Configuring Private Attributes

1. Overview

Vendors can define private attributes to implement multiple features that are not supported by the standard RADIUS protocol. You can configure the methods used by devices to process private attributes based on the actual environment. Private attributes configured in this section are all our company's private attributes.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure private attributes. Configure at least one of the tasks.

- Configure the device to set the private attribute **port-priority** delivered by the server to the class of service (CoS) value of an interface.

  rds set qos cos

  The device sets the private attribute **port-priority** delivered by the server to the differentiated services code point (DSCP) value by default.

- Configure the device to support the **CUI** attribute.

  rds support cui

  The function of supporting the **CUI** attribute by RADIUS is disabled by default.

1.5.5 Configuring Compatibility with Devices of Other Vendors

1. Overview

In actual deployment environments, the devices may interconnect to RADIUS servers of other vendors. Private attributes defined by different vendors are different. It is necessary to enable compatibility with the configuration of devices of other vendors.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure compatibility with devices of other vendors. Configure at least one of the tasks.

- Configure whether RADIUS authentication request packets carry the private attributes of a specified vendor.

  rds-**server authentication vendor** { **cisco | cmcc** | **microsoft** } **package**

  Authentication request packets do not carry private attributes of other vendors by default.

- Configure whether RADIUS accounting request packets carry the private attributes of a specified vendor.

  rds-**server account vendor** { **cisco | cmcc** | **microsoft** } **package**

  Accounting request packets do not carry private attributes of other vendors by default.

- Configure RADIUS to parse private attributes of Cisco, Huawei, and Microsoft devices carried in packets.

  rds vendor-**specific attribute support** { **cisco** | **huawei** | **ms** }

  RADIUS parses private attributes of Cisco, Huawei, and Microsoft devices carried in packets by default.

- Enable the function of not differentiating private vendor IDs during RADIUS packet parsing.

  rds vendor-specific extend

  RADIUS identifies only private vendor ID during packet parsing by default.

- Configure the mode of parsing private attributes.

  rds vendor-specific attribute support { **cisco** | **huawei** | **ms** }

  RADIUS parses private attributes of Cisco, Huawei, and Microsoft devices carried in packets by default.

  If the server is an application server, the RADIUS private attribute type needs to be configured to achieve compatibility.

1.6 Configuring RADIUS Reachability Detection

1.6.1 Overview

A RADIUS server can be only in the reachable or unreachable state. The device will not send authentication, authorization, or accounting requests of access users to an unreachable RADIUS server unless all servers in the RADIUS server group are unreachable.

The device maintains the reachability status of each RADIUS server. The device selects a reachable RADIUS server preferentially to improve the handling performance of RADIUS services.

The device actively detects whether a specified RADIUS server is reachable. After the active detection function is configured, the device periodically sends detection requests (authentication requests or accounting requests) to the RADIUS server. The active detection interval is **60** minutes when the RADIUS server is reachable and **1** minute when the RADIUS server is unreachable.

1.6.2 Restrictions and Guidelines

- To enable active detection for a specified RADIUS server, perform the following configuration when configuring the RADIUS server:
  - Configure a test username for the RADIUS server.
  - Configure at least one tested port (authentication port or accounting port) for the RADIUS server.
- It is judged that a RADIUS server is unreachable only when both conditions below are met:
  - The device fails to receive a correct response packet from the RADIUS server within the specified timeout duration.
  - The number of times that the device sends request packets to the same RADIUS server consecutively reaches the specified timeout count.
- If any of the following conditions is met for an unreachable RADIUS server, the RADIUS server is considered reachable.
  - The device receives correct responses from the RADIUS server.
  - The duration in which the RADIUS server is unreachable exceeds the time configured using the **rds-server deadtime** command, and active detection is disabled for the RADIUS server.
  - The authentication port or accounting port of the RADIUS server is updated on the device.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the criteria for the device to judge that a RADIUS server is unreachable.

rds-server dead-condition { **time** *timeout* | **time** *timeout* **attempts** *tries-number* | **attempts** *tries-number* }

The criteria for judging that a RADIUS server is unreachable are that the timeout duration is 60 seconds and the consecutive timeout count is 10 by default.

If a RADIUS server meets both the duration condition and the consecutive request timeout count condition, it is deemed that the RADIUS server is unreachable.

(4) Configure active detection and set the IP address, authentication port, or accounting port for the remote RADIUS server.

rds-server host { *ipv4-address* | *ipv6-address* } [**accounting-port** *acct-port-number* | **authentication-port** *auth-port-number*] *

No active detection is configured by default.

(5) (Optional) Configure the number of times that the device sends requests to a RADIUS server consecutively before confirming that the RADIUS server is unreachable.

rds-server retransmit *retransmit-times*

The device sends requests to a RADIUS server three times consecutively before confirming that the RADIUS server is unreachable by default.

(6) (Optional) Configure the duration for the device to stop sending request packets to an unreachable RADIUS server.

rds-server deadtime *deadtime*

Even if a RADIUS server is unreachable, the device still sends requests to the RADIUS server by default.

If active detection is enabled for a RADIUS server, the time parameter configured by the **rds-server deadtime** command does not take effect on the RADIUS server.

1.7 Configuring the DSCP Value for RADIUS Packets

1.7.1 Overview

DSCP is in the type of service (ToS) field of the IP header and is used to identify the packet transmission priority.

A larger DSCP value indicates a higher packet priority. The default DSCP value of RADIUS packets is **0**. You can configure the DSCP value for RADIUS packets to change the transmission priority of RADIUS packets.

1.7.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the DSCP value for RADIUS packets.

radius dscp *dscp-value*

The default DSCP value of RADIUS packets is **0**.

1.8 Configuring the Units of Data Flows and Packets to Be Sent to a RADIUS Server

1.8.1 Overview

The default units of data flows and data packets to be sent to a RADIUS server are bytes and packets respectively. You can configure the units of data flows and data packets to be sent to a RADIUS server based on the actual network situation.

1.8.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the units of data flows and data packets to be sent to a RADIUS server.

rds data-flow-format { { **data byte** | **data giga-byte** | **data kilo-byte** | **data mega-byte** } | { **packet giga-packet** | **packet kilo-packet** | **packet mega-packet** | **packet one-packet** } } *

The default units of data flows and data packets to be sent to a RADIUS server are bytes and packets respectively.

1.9 Configuring the Accounting-On Function

1.9.1 Overview

The accounting-on function is used to notify a RADIUS server of the device restart. After the device is restarted, online users are forced offline. However, the RADIUS server does not perceive the device restart and does not log off the users. As a result, the users encounter an exception when initiating re-authentication. Therefore, it is necessary to enable the accounting-on function.

1.9.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the function of sending accounting-on packets upon device restart.

rds-server accounting-on enable

The function of sending accounting-on packets upon device restart is enabled by default.

1.10 Configuring the Accounting Packet Copying Function

1.10.1 Overview

RADIUS accounting packets are sent to one server specified in the accounting method list by default. After this function is enabled, the device copies accounting packets and sends the packets to all servers in a server group.

1.10.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

(4) Configure the function of copying and sending accounting packets to servers in a specified group.

rds-sever accounting-copy *group*

The function of copying and sending RADIUS accounting packets is disabled by default.

1.11 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

Run the **debugging** commands to output debugging information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-6 RADIUS Monitoring

Command	Purpose
display radius parameter	Displays global parameters of a RADIUS server.
display radius server	Displays the configuration of a RADIUS server.

Command	Purpose
display radius vendor-specific	Displays the configuration of the RADIUS private attribute type.
display radius auth statistics	Displays RADIUS authentication statistics.
display radius acct statistics	Displays RADIUS accounting statistics.
display radius group	Displays the configuration of a RADIUS server group.
display radius attribute	Displays RADIUS standard attributes.
debugging radius detail	Debugs RADIUS packets.
debugging radius extend event	Debugs the function of Dynamic Authorization Extensions to RADIUS.
debugging radius extend detail	Debugs packets of Dynamic Authorization Extensions to RADIUS.
debugging radius extend error	Prints information related to packets of Dynamic Authorization Extensions to RADIUS for debugging.

1.12 Configuration Examples

1.12.1 Configuring RADIUS Authentication, Authorization, and Accounting

1. Requirements

A RADIUS server needs to be used for authentication, authorization, and accounting of login users.

2. Topology

Figure 1-4 Topology of RADIUS Authentication, Authorization, and Accounting



3. Notes

- Configure device information and add login users on the RADIUS server.
- Enable AAA security services on the device.
- Configure RADIUS server information on the device.
- Configure the RADIUS authentication method list, authorization method list, and accounting method list.
- Apply the RADIUS authentication, authorization, and accounting methods to a specific line.

4. Procedure

(1) Configure device information and add login users on the RADIUS server. The configuration is omitted here. For details, see the RADIUS server configuration.

(2) Enable AAA security services.

```
Device> privilege
Device# config-system terminal
Device(cfg)# aaa-config new-pattern
```

(3) Configure a RADIUS server (the IP address of the RADIUS server is set to 192.168.1.3 and the shared key is set to **sharekey** here).

```
Device(cfg)# rds-server host 192.168.1.3 share-key sharekey
```

(4) Configure the authentication, authorization, and accounting method lists.

```
Device(cfg)# aaa-config authentication line-login radius-method server-group rds
Device(cfg)# aaa-config authorization execute radius-method server-group rds
Device(cfg)# aaa-config accounting execute radius-method start-stop server-group rds
```

(5) Apply the authentication, authorization, and accounting methods to a line.

```
Device(cfg)# line vty 0 4
Device(cfg-line)# line-login authentication radius-method
Device(cfg-line)# aaa-authorization execute radius-method
Device(cfg-line)# aaa-accounting execute radius-method
```

5. Verification

After entering the correct username and password, a user can log in to the device successfully.

```
User Access Verification

Username:hostname1
Password:password1



Device#
```

After login, the user has only the privilege level granted by the server and can run commands only under this privilege level.

After the user logs out, accounting information of the user can be queried on the RADIUS server. For details about how to query accounting information, see the RADIUS server configuration.

6. Configuration Files

```
!
aaa-config new-pattern
!
aaa-config accounting execute radius-method start-stop server-group radius
aaa-config authorization execute radius-method server-group radius
aaa-config authentication line-login radius-method server-group radius
!
rds-server host 192.168.1.3 share-key 7 $10$275$g8oXDDIPVeA=$
```

```
!
line console 0
line vty 0 4
 aaa-accounting execute radius-method
 aaa-authorization execute radius-method
 line-login authentication radius-method
!
```

7. Common Errors

The shared key configured on the device is inconsistent with that used by the RADIUS server.

1.12.2 Configuring RADIUS Reachability Detection

1. Requirements

RADIUS reachability detection needs to be configured to identify unreachable RADIUS servers.

2. Topology

Figure 1-5 Topology of RADIUS Reachability Detection



3. Notes

- Configure the global criteria for judging that a RADIUS server is unreachable.
- Configure an IP address for the RADIUS server and configure active detection parameters.

4. Procedure

Configure the global criteria for judging that a RADIUS server is unreachable as follows: The consecutive timeout count is **5** and the timeout duration is **120** seconds.

```
Device> privilege
Device# config-system terminal
Device(cfg)# rds-server dead-condition time 120 attempts 5
```

Set the IP address of the RADIUS server to 192.168.1.3, detection username to **detect**, and detection interval to **90** minutes, and disable the server authentication port detection.

```
Device(cfg)# rds-server host 192.168.1.3 detect user-name detect ignore-accounting-port timeout 90
```

5. Verification

Disconnect the network communication between the device and the server with the IP address 192.168.1.3. Initiate RADIUS authentication through the device. After 120 seconds, run the **display rds server** command to check that the server status is "Dead".

```
Device# display rds server
Server IP:       192.168.1.3
Accounting Port:    1813
```

```
Authen    Port:         1812
mom         flag:         1
Test Username:       detect
Test Idle Time:      1 Minutes
Test Ports:             Authen and Accounting
Server State:         Dead
      Current duration 1254s, previous duration 2918s
      Dead: total time 1257s, count 1
      Statistics:
            Authen: request 22, timeouts 81
            Author: request 22, timeouts 81
            Account: request 23, timeouts 83
```

6. Configuration Files

```
!
rds-server host 192.168.1.3 detect user-name detect ignore-accounting-port timeout 90 share-key 7
$10$275$g8oXDDIPVeA=$
rds-server dead-condition time 120 attempts 5
!
```

Contents

1 Configuring TACACS....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.1.3 Protocols and Standards....4
1.2 Configuration Task Summary....4
1.3 Configuring TACACS+ Basic Features....5
1.3.1 Overview....5
1.3.2 Restrictions and Guidelines....5
1.3.3 Procedure....5
1.4 Configuring TACACS+ Reachability Detection....6
1.4.1 Overview....6
1.4.2 Restrictions and Guidelines....6
1.4.3 Procedure....7
1.5 Configuring a TACACS+ Server Group....7
1.5.1 Overview....7
1.5.2 Restrictions and Guidelines....7
1.5.3 Procedure....8
1.6 Monitoring....8
1.7 Configuration Examples....9
1.7.1 Configuring TACACS+ for Login Authentication....9
1.7.2 Configuring Separated Authentication, Authorization and Accounting by Using TACACS+....10

1.7.3 Configuring Reachability Detection 13

1 Configuring TACACS

1.1 Introduction

1.1.1 Overview

Terminal Access Controller Access Control System (TACACS) is a distributed authentication, authorization and accounting (AAA) protocol based on the client/server architecture. TACACS+ has enhanced functions on the basis of the TACACS protocol, to provide the authentication, authorization and accounting services for users. This following describes TACACS+.

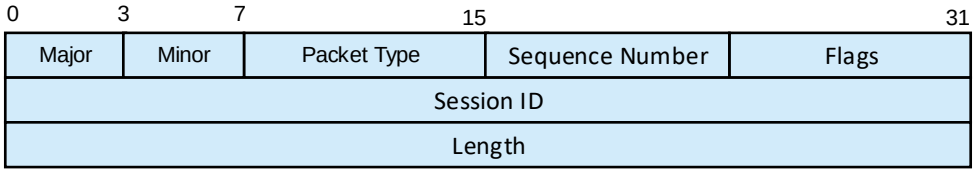
Both TACACS+ and Remote Authentication Dial-In User Service (RADIUS) are capable of providing AAA security services. Compared with RADIUS, TACACS+ has the following advantages:

- It is implemented based on the Transmission Control Protocol (TCP) and provides more reliable network transmission.
- TACACS+ encrypts both packet bodies and passwords to boost security.
- Authentication is separated from authorization. The authentication server and authorization server can be deployed separately in remote deployment.
- TACACS+ supports authorization and accounting on configuration commands.

1.1.2 Principles

1. TACACS+ Packet

Figure 1-1 TACACS+ Packet Structure



Fields contained in the TACACS+ packet structure shown in Figure 1-1 are described in Table 1-1.

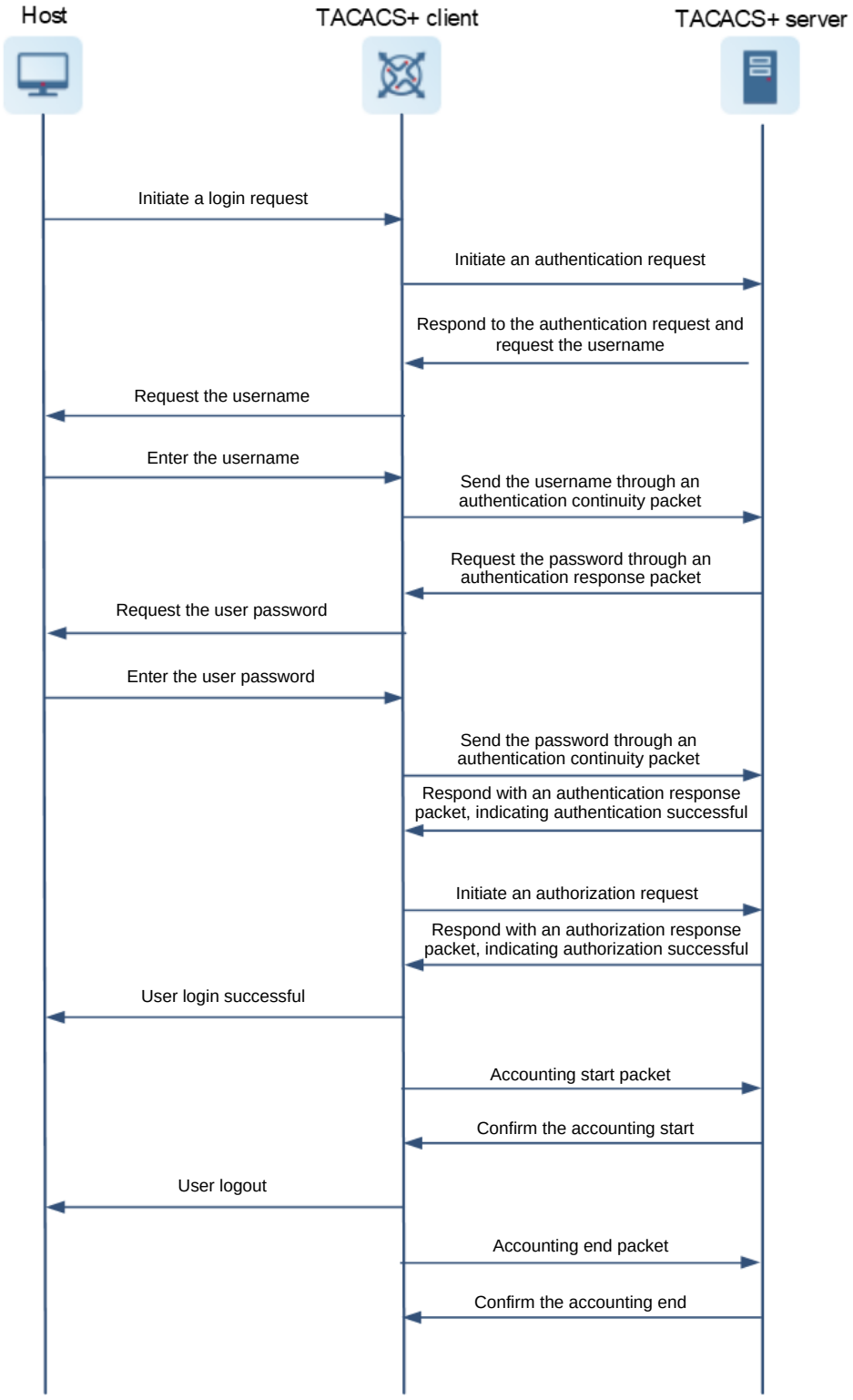
Table 1-1 Description of Fields Contained in a TACACS+ Packet

Field	Length	Description
Major Version	4 bits	Major TACACS+ version
Minor Version	4 bits	Minor TACACS+ version
Packet Type	1 byte	Packet type. The options are as follows: **0x01**: Indicates an authentication packet. **0x02**: Indicates an authorization packet. **0x03**: Indicates an accounting packet.

Field	Length	Description
Sequence Number	1 byte	Sequence number of a data packet in the current session (The sequence number of the first TACACS+ data packet in a session must be 1 and the sequence number of each subsequent data packet increases by one. The client sends data packets only with an odd sequence number and TACACS+ Daemon sends data packets only with an even sequence number.)
Flags	1 byte	Whether the data packet is encrypted
Session ID	4 bytes	TACACS+ session ID
Length	4 bytes	Body length of a TACACS+ data packet (excluding the header)

2. Packet Exchange Process

Figure 1-2 TACACS+ Packet Exchange Flowchart



As shown in Figure 1-2, the TACACS+ protocol basically involves the authentication, authorization, and accounting processes. The network access server (NAS) requests security services from the TACACS+ server as a TACACS+ client.

(1) Authentication process

a A user initiates a login request to the NAS.

b After receiving the user request, the NAS sends an authentication start packet to the TACACS+ server to initiate an authentication request.

c The TACACS+ server responds to the authentication request and asks the NAS to provide the username.

d The NAS asks the user terminal to provide the username.

e The user enters the username and sends it to the NAS.

f The NAS sends the username to the TACACS+ server through an authentication continuity packet.

g After receiving the username, the TACACS+ server continues to ask the NAS for the user password.

h The NAS requests the user terminal to provide the password.

i The user enters the password and sends it to the NAS.

j The NAS sends the password to the TACACS+ server through an authentication continuity packet.

k After receiving the password, the TACACS+ server verifies the password and then returns an authentication success message to the NAS.

(2) Authorization process

a The NAS initiates an authorization request to the TACACS+ server.

b The TACACS+ server responds to the authorization request by returning an authorization success message to the NAS.

c After receiving the authorization success message, the NAS allows the user to log in to the NAS and perform operations within the authorized scope.

(3) Accounting process

a The NAS sends an accounting start request to the TACACS+ server.

b After receiving the accounting start request, the TACACS+ server starts accounting and returns an accounting request acceptance message to the NAS.

c If the user logs out, the NAS sends an accounting end request to the TACACS+ server after receiving a user logout message.

d After receiving the accounting end request, the TACACS+ stops accounting and returns an accounting end message.

1.1.3 Protocols and Standards

- RFC 1492: Terminal Access Controller Access Control System

1.2 Configuration Task Summary

TACACS+ configuration includes the following tasks:

- Configuring TACACS+ Basic Features
- (Optional) Configuring TACACS+ Reachability Detection
- (Optional) Configuring a TACACS+ Server Group

1.3 Configuring TACACS+ Basic Features

1.3.1 Overview

This section describes how to configure parameters for the communication between the device and a TACACS+ server to ensure normal communication between them. After the parameters are configured, the device can correctly send AAA authentication, authorization, and accounting requests to the TACACS+ server and give responses.

1.3.2 Restrictions and Guidelines

Configuring TACACS+ basic features only ensures proper communication between the device and a TACACS+ server. To provide AAA security services for users, you still need to reference the TACACS+ server in an AAA method list.

When the single-connection mode is disabled, the multi-connection mode is used by default. That is, a TCP connection is established for each session and released after the session ends. When the single-connection mode is enabled, a TCP connection is established during the first interaction with the server. Subsequent sessions with the server use the TCP connection. A TCP connection does not need to be established for each session.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable AAA.

aaa-config new-pattern

(4) Configure a TACACS+ server.

tacplus-server host [**oob** [**via Mgmt** *Mgmt-number*]] { *ipv4-address* | *ipv6-address* } [**protocol-port** *port-number*] [**detect user-name** *username*] [**verify-time** *idle-time*] [**timeout** *timeout*] [**single-connection**] [**share-key** [**0** | **7**] *key*]

No TACACS+ server is configured by default.

You can specify a shared key for a TACACS+ server when configuring the server. If it is not specified, the global shared key will be used.

(5) (Optional) Configure the global shared key for the communication between the device and TACACS+ server.

tacplus-server share-key [**0** | **7**] *key*

No global shared key is configured for the device and TACACS+ server by default.

(6) (Optional) Configure the source address for TACACS+ packets.

ip tacacs-mode source-interface *interface-type interface-number*

No source address is configured for TACACS+ packets and the address is set by the network layer by default.

You can run this command to specify an outbound interface for packets. After this command is configured, the first IP address of the interface is used as the source address of TACACS+ packets.

(7) (Optional) Configure the timeout duration for the TACACS+ server.

tacplus-server timeout *timeout*

The default timeout duration of a TACACS+ server is **5** seconds in the communication between the device and the TACACS+ server.

(8) (Optional) Configure the single connection timeout duration for a TACACS+ server.

tacacs-server single-connect-timeout *timeout*

The default single connection timeout duration is 30 minutes, and the range is from 0 to 480. After the timeout duration times out, a connection is reestablished when the single-connection mode is used again.

1.4 Configuring TACACS+ Reachability Detection

1.4.1 Overview

A TACACS+ server can be only in the reachable or unreachable state. The device will not send authentication, authorization, or accounting requests of access users to an unreachable TACACS+ server unless all servers in the TACACS+ server group are unreachable.

The device maintains the reachability status of each TACACS+ server. The device selects a reachable TACACS+ server preferentially to improve the handling performance of TACACS+ services.

The device actively detects whether a specified TACACS+ server is reachable. After the active detection function is configured for a TACACS+ server, the device periodically sends detection requests (authentication requests or accounting requests) to the TACACS+ server. The active detection interval is **60** minutes when the TACACS+ server is reachable and **1** minute when the TACACS+ server is unreachable.

1.4.2 Restrictions and Guidelines

- To enable active detection for a specified TACACS+ server, perform the following configuration when configuring the TACACS+ server:
  - Configure a test username for the TACACS+ server.
  - Configure at least one tested port (authentication port or accounting port) for the TACACS+ server.
- It is judged that a TACACS+ server is unreachable only when both conditions below are met:
  - The device fails to receive a correct response packet from the TACACS+ server within the specified timeout period.
  - The consecutive transmission count of a request packet sent by the device to the same TACACS+ server reaches the specified timeout count.
- If any of the following conditions is met for an unreachable TACACS+ server, the TACACS+ server is considered reachable:
  - The device receives a correct response from the TACACS+ server.
  - The duration in which the TACACS+ server is unreachable exceeds the time configured using the **tacplus-server deadtime** command and active detection is disabled for the TACACS+ server.
  - The authentication port or accounting port of the TACACS+ server is updated on the device.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the criteria for the device to judge that a TACACS+ server is unreachable.

tacplus-server dead-condition { **time** *timeout* | **time** *timeout* **attempts** *tries-number* | **attempts** *tries-number* }

The criteria for the device to judge that a TACACS+ server is unreachable are that the server timeout duration is **60** seconds and the consecutive timeout count is **10** by default.

(4) (Optional) Configure the duration for the device to stop sending request packets to an unreachable TACACS+ server.

tacplus-server deadtime *deadtime*

Even if a TACACS+ server is unreachable, the device still sends requests to the TACACS+ server by default.

If active detection is enabled for a TACACS+ server, the time parameter configured by the **tacplus-server deadtime** command does not take effect on the TACACS+ server.

1.5 Configuring a TACACS+ Server Group

1.5.1 Overview

One or more TACACS+ servers can be added to each TACACS+ server group, which provides AAA services for users as a whole.

When the single-connection mode is disabled, the multi-connection mode is used by default. That is, a TCP connection is established for each session and released after the session ends. When the single-connection mode is enabled, a TCP connection is established during the first interaction with the server. Subsequent sessions with the server use the TCP connection. A TCP connection does not need to be established for each session.

1.5.2 Restrictions and Guidelines

- TACACS+ security services are a type of AAA services and need to be used in combination with the AAA features. TACACS+ provides security services for users as one method in an AAA method list.
- When configuring the authentication, authorization, and accounting method lists, you can specify server groups separately for them.
- In a user-defined server group, you can only specify and apply servers in the default server group.
- TACACS+ server groups support virtual routing and forwarding (VRF) instances. When a server group in a specified VRF instance is used, the source address used by the device to communicate with a remote server must be obtained from the VRF instance. If you run the **ip tacplus-mode source-interface** command to specify the source interface for request packets, the IP address obtained from this source interface takes priority over that found in the VRF instance.

- The name of a server group cannot be set to the predefined keyword **tacacs+**.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a TACACS+ server group.

aaa-config group server tacplus *group-name*

No TACACS+ server group is configured by default.

You can group TACACS+ servers so that authentication, authorization, and accounting can be completed by different server groups.

(4) Add a server to the TACACS+ server group.

server { *ipv4-address* | *ipv6-address* } [**single-connection**]

No server is added to a TACACS+ server group by default.

(5) (Optional) Specify a VRF instance for the TACACS+ server group.

ip vrf binding *vrf-name*

No VRF instance is specified for a TACACS+ server group by default.

The VRF instance configured for a TACACS+ server group must use a valid name configured using the **vrf definition** command in global configuration mode.

(6) Configure an MGMT port to be used by the TACACS+ server group.

ip oob [**via mgmt** *mgmt-number*]

No MGMT port to be used by a TACACS+ server group is configured by default.

(7) (Optional) Configure the function of carrying the **nas-ip** attribute in packets.

nas-ip { *ipv4-address* | *ipv6-address* }

Packets do not carry the **nas-ip** attribute by default.

1.6 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** commands to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-2 TACACS+ Monitoring

Command	Purpose
display tacacs	Displays the interaction between the device and each TACACS+ server.
debugging tacacs+	Debugs TACACS+.

1.7 Configuration Examples

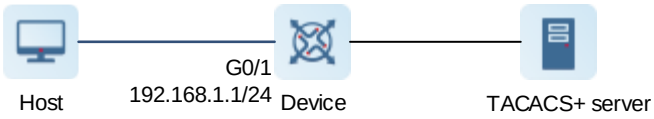
1.7.1 Configuring TACACS+ for Login Authentication

1. Requirements

A TACACS+ server needs to be used for authentication of login users.

2. Topology

Figure 1-3 Topology of Login Authentication Using TACACS+



3. Notes

- Configure device information and add login users on the TACACS+ server.
- Enable AAA security services on the device.
- Configure TACACS+ server information on the device.
- Configure a TACACS+ authentication method list.
- Apply the TACACS+ authentication methods to a specific line.

4. Procedure

(1) Configure device information and add login users on the TACACS+ server. The configuration is omitted here. For details, see the TACACS+ server configuration.

(2) Enable AAA security services.

```
Device> privilege
Device# config-system terminal
Device(cfg)# aaa-config new-pattern
```

(3) Configure a TACACS+ server (the IP address of the TACACS+ server is set to 192.168.1.3 and the shared key is set to **sharekey** here).

```
Device(cfg)# tacplus-server host 192.168.1.3 share-key sharekey
```

(4) Configures an authentication method list.

```
Device(cfg)# aaa-config authentication line-login tacacs-method server-group tacacs+
```

(5) Apply the authentication methods to a line.

```
Device(cfg)# line vty 0 4
Device(cfg-line)# line-login authentication tacacs-method
```

5. Verification

After entering the correct username and password, a user can log in to the device successfully.

```
User Access Verification

Username:hostname1
Password:password1



Device#
```

6. Configuration Files

```
!
aaa-config new-pattern
!
aaa-config authentication line-login tacacs-method server-group tacplus
!
tacplus-server host 192.168.1.3 share-key 7 $10$275$g8oXDDIPVeA=$
!
line console 0
line vty 0 4
 line-login authentication tacacs-method
!
```

7. Common Errors

The shared key configured on the device is inconsistent with that used by the TACACS+ server.

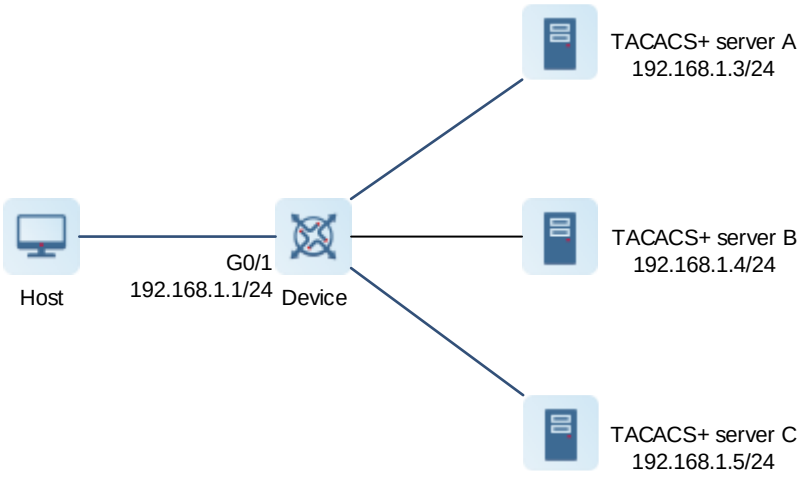
1.7.2 Configuring Separated Authentication, Authorization and Accounting by Using TACACS+

1. Requirements

The user authentication, authorization, and accounting services need to be separately completed by different TACACS+ server groups.

2. Topology

Figure 1-4 Topology of Separated Authentication, Authorization, and Accounting of TACACS+



3. Notes

- Configure device information and add login users on the TACACS+ server.
- Enable AAA security services on the device.
- Configure TACACS+ servers on the device and add them to different TACACS+ server groups.
- Configure TACACS+ authentication, authorization, and accounting method lists.
- Apply the TACACS+ authentication, authorization, and accounting methods to a specific line.

4. Procedure

(1) Configure device information and add login users on the TACACS+ server. The configuration is omitted here. For details, see the TACACS+ server configuration.

(2) Enable AAA security services.

```
Device> privilege
Device# config-system terminal
Device(cfg)# aaa-config new-pattern
```

(3) Configure TACACS+ server A and add it to TACACS+ server group **tacacs1**. (The IP address of TACACS+ server A is set to 192.168.1.3 and the shared key is set to **sharekey** here.)

```
Device(cfg)# tacplus-server host 192.168.1.3 share-key sharekey
Device(cfg)# aaa-config group server tacplus tacacs1
Device(cfg-gs-tacacs+)# server 192.168.1.3
Device(cfg-gs-tacacs+)# exit
```

(4) Configure server groups **tacacs2** and **tacacs3** by referring to the steps of configuring server group **tacacs1**.

```
Device(cfg)# tacplus-server host 192.168.1.4 share-key sharekey
Device(cfg)# aaa-config group server tacplus tacacs2
Device(cfg-gs-tacacs+)# server 192.168.1.4
Device(cfg-gs-tacacs+)# exit
Device(cfg)# tacplus-server host 192.168.1.5 share-key sharekey
Device(cfg)# aaa-config group server tacplus tacacs3
```

```
Device(cfg-gs-tacacs+)# server 192.168.1.5
Device(cfg-gs-tacacs+)# exit
```

(5) Configure an authentication method list to use server group **tacacs1**, an authorization method list to use server group **tacacs2**, and an accounting method list to use server group **tacacs3**.

```
Device(cfg)# aaa-config authentication line-login tacacs-method group tacacs1
Device(cfg)# aaa-config authorization execute radius-method server-group tacacs2
Device(cfg)# aaa-config accounting execute radius-method start-and-stop server-group tacacs3
```

(6) Apply the authentication methods, authorization methods, and accounting methods to a line.

```
Device(cfg)# line vty 0 4
Device(cfg-line)# line-login authentication tacacs1
Device(cfg-line)# aaa-authorization execute tacacs2
Device(cfg-line)# aaa-accounting execute tacacs3
```

5. Verification

After entering the correct username and password configured on TACACS+ server A, a user can log in to the device successfully.

```
User Access Verification

Username:hostname1
Password:password1



Device#
```

After login, the user has only the privilege level granted by TACACS+ server B and can only run commands under this privilege level.

After the user logs out, accounting information of the user can be queried on TACACS+ server C. For details about how to query accounting information, see the TACACS+ server configuration.

6. Configuration Files

```
!
aaa-config new-pattern
!
aaa-config accounting execute radius-method start-and-stop server-group tacacs3
aaa-config authorization exec radius-method server-group tacacs2
aaa-config authentication line-login radius-method server-group tacacs1
!
tacplus-server host 192.168.1.3 share-key 7 $10$275$g8oXDDIPVeA=$
tacplus-server host 192.168.1.4 share-key 7 $10$275$g8oXDDIPVeA=$
tacplus-server host 192.168.1.5 share-key 7 $10$275$g8oXDDIPVeA=$
!
!
aaa-config group server tacplus tacacs1
 server 192.168.1.3
!
```

```
aaa-config group server tacplus tacacs2
 server 192.168.1.4
!
aaa-config group server tacplus tacacs3
 server 192.168.1.5
!
line console 0
line vty 0 4
 aaa-accounting execute tacacs3
 aa-authorization execute tacacs2
 line-login authentication tacacs1
!
```

7. Common Errors

- The shared key configured on the device is inconsistent with that used by the TACACS+ server.
- Undefined servers are added to a server group.

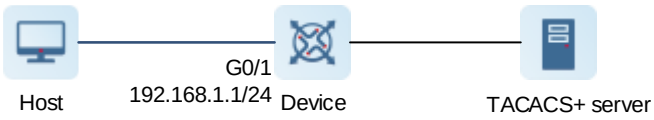
1.7.3 Configuring Reachability Detection

1. Requirements

TACACS+ reachability detection needs to be configured to identify unreachable TACACS+ servers.

2. Topology

Figure 1-5 Topology of TACACS+ Reachability Detection



3. Notes

- Configure the global criteria for judging that a TACACS+ server is unreachable.
- Configure an IP address for the TACACS+ server and configure active detection parameters.

4. Procedure

Configure the global criteria for judging that a TACACS+ server is unreachable as follows: The consecutive timeout count is **5** and the timeout duration is **120** seconds.

```
Device> privilege
Device# config-system terminal
Device(cfg)# racplus-server dead-condition time 120 attempts 5
```

Set the IP address of the TACACS+ server to 192.168.1.3, detection username to **detect**, detection interval to **90** minutes.

```
Device(cfg)# tacplus-server host 192.168.1.3 detect user-name detect verify-time 90
```

5. Verification

Disconnect the network communication between the device and the server with the IP address 192.168.1.3. Initiate TACACS+ authentication through the device. After 120 seconds, run the **display tacacs-mode** command to check that the server status is "Dead".

```
Device# display tacacs-mode server
Tacacs+ Server : 192.168.1.3/49
state: Dead
Socket Opens: 0
Socket Closes: 0
Total Packets Sent: 0
Total Packets Recv: 0
Reference Count: 1

Tacacs capability:
    mgmt enable: true
    vrf enable: true
```

6. Configuration Files

```
!
tacplus-server host 192.168.1.3 detect user-name detect verify-time 90
tacplus-server dead-condition time 120 attempts 5
!
```

Contents

1 Configuring SCC 1

1.1 Overview 1

1.2 Configuration Task Summary 1

1.3 Configuring the IPv4 User Capacity 1

1.3.1 Overview 1

1.3.2 Restrictions and Guidelines 1

1.3.3 Procedure 1

1.4 Monitoring 2

1 Configuring SCC

1.1 Overview

The Security Control Center (SCC) provides common configuration methods and policy integration for various access control and network security services, so that these services can coexist on one device to meet diversified access and security control requirements in various scenarios.

Typical access control services include 802.1x, web authentication, Address Resolution Protocol (ARP) check, and IP source guard, and typical network security services include access control list (ACL), NFPP, and gateway-targeted ARP spoofing prevention. When two or more access control or network security services are enabled on the same device, the SCC coordinates coexistence based on related policies.

1.2 Configuration Task Summary

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the IPv4 User Capacity

1.3 Configuring the IPv4 User Capacity

1.3.1 Overview

The IPv4 user capacity is used to limit the maximum number of IPv4 access users on a port of the device. When the total number of IPv4 access users on a port reaches the limit, new users cannot access the network over this port. You need to select a proper user capacity based on the actual environment to guard against brutal force impacts from unauthorized users and improve the operational stability of the device.

1.3.2 Restrictions and Guidelines

IPv4 users include those generated through 802.1x authentication, web authentication, and other binding functions. IPv4 users on a port may be generated over the port or globally. For example, when a global IPv4 user is bound to a port by running the corresponding command, the user is also calculated as a user on the port.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

interface aggregateport *interface-number*

(4) Configure the IPv4 user capacity on a port.

nac-author-user maximum *max-user-number*

The IPv4 user capacity on a port is not limited by default.

1.4 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** commands to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging switch immediately after use.

Table 1-1 Monitoring

Command	Purpose
display nac-author-user [**interface** *interface-type interface-number*]	Displays IPv4 user entries on a port.
display access-control packet statistics [**interface** *interface-type interface-number* | **vlan** *vlan-id*]	Displays statistics about packets filtered due to access control.
debugging scc event	Debugs the SCC running process.
debugging scc user	Debugs SCC user entries.
debugging scc acl-display summary	Debugs ACL summary stored in the SCC delivered by various services.
debugging scc acl-display all	Debugs all ACLs stored in the SCC.

Contents

1 Configuring Password Policies 1
1.1 Introduction 1
1.1.1 Overview 1
1.2 Restrictions and Guidelines 1
1.3 Configuration Task Summary 1
1.4 Configuring the Minimum Password Length 1
1.4.1 Overview 1
1.4.2 Procedure 1
1.5 Enabling Strong Password Detection 2
1.5.1 Overview 2
1.5.2 Procedure 2
1.6 Enabling Forcible Weak Password Change 2
1.6.1 Overview 2
1.6.2 Procedure 2
1.7 Configuring a Password Lifecycle 3
1.7.1 Overview 3
1.7.2 Procedure 3
1.8 Configuring Repeated Password Use Prevention 3
1.8.1 Overview 3
1.8.2 Procedure 3
1.9 Configuring Encrypted Password Storage 4
1.9.1 Overview 4
1.9.2 Procedure 4

1.10 Enabling Special Character Detection....4
1.10.1 Overview....4
1.10.2 Procedure....4
1.11 Monitoring....4
1.12 Configuration Examples....5
1.12.1 Requirements....5
1.12.2 Topology....5
1.12.3 Notes....5
1.12.4 Procedure....5
1.12.5 Verification....6
1.12.6 Configuration Files....6

1 Configuring Password Policies

1.1 Introduction

1.1.1 Overview

A password policy provides the password detection function for applications that need to configure accounts and passwords (for example, creating local users) on a device to ensure that users use passwords that meet security specifications and prevent password crackdown due to low password complexity.

1.2 Restrictions and Guidelines

The password lifecycle, repeated password use prevention, and forcible weak password change functions take effect only to global passwords (**privilege cipher** and **privilege safe-cipher**) and local user passwords (**user-account** *username* **password** *password*) and are invalid to passwords in line configuration mode.

1.3 Configuration Task Summary

All the configuration tasks below are optional. Select the configuration tasks as required.

(1) Configuring the Minimum Password Length

(2) Enabling Strong Password Detection

(3) Enabling Forcible Weak Password Change

(4) Configuring a Password Lifecycle

(5) Configuring Repeated Password Use Prevention

(6) Configuring Encrypted Password Storage

(7) Enabling Special Character Detection

1.4 Configuring the Minimum Password Length

1.4.1 Overview

The minimum password length is used to limit the length of user passwords.

If the password entered by a user is shorter than the minimum password length, the system displays an error prompt, asking the user to specify another password of an appropriate length.

1.4.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the minimum password length.

password policy min-size *min-size*

No minimum length is configured for passwords by default.

1.5 Enabling Strong Password Detection

1.5.1 Overview

Strong password detection is used to detect the complexity of a password and prevent password crackdown due to low complexity.

The strong password detection function will send an alarm and disallow the user to configure the password in the following scenarios:

- The password is the same as or similar to the account.
- The password contains only one or two types of digits, uppercase letters, lowercase letters, and special characters.
- The password length is shorter than the specified length (no less than eight characters by default).

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the strong password detection function.

password policy strong

The strong password detection function is enabled by default.

1.6 Enabling Forcible Weak Password Change

1.6.1 Overview

The forcible weak password change function is used together with another password policy (such as the minimum password length or strong password detection).

After the forcible weak password change function is enabled, a warning prompt will be displayed if a user uses a weak password (containing less than eight bytes, or containing only one or two types of digits, uppercase letters, and lowercase letters) during login or configuration.

If both the forcible weak password change function and another password policy are enabled, the password will continue to be checked according to the other password policy during user login. If the password does not meet requirements of the other password policy, a prompt for changing the password will be displayed. The user is allowed to log in only after the new password meets requirements of the password policy.

1.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable forcible weak password change.

password policy forced-password-modify

Forcible weak password change is disabled by default.

1.7 Configuring a Password Lifecycle

1.7.1 Overview

The password lifecycle defines the validity time of a password.

If a user enters a password that has already expired during login, the system gives a prompt, indicating that the password has expired, and asks the user to reset the password.

1.7.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the password lifecycle.

password policy life-cycle *life-cycle*

No password lifecycle is configured by default.

1.8 Configuring Repeated Password Use Prevention

1.8.1 Overview

The repeated password use prevention function prevents users from using historically configured passwords as new passwords.

The system records passwords used by a user to a historical password list. If a newly configured password is within the list, the system gives a prompt and asks the user to specify another password. The maximum number of records in the password list can be manually configured. When the number of records in the password list reaches the limit, a new password record will overwrite the earliest password record.

1.8.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-terminal terminal

(3) Prevent repeated use of passwords configured in the latest specified number of times.

password policy no-repeat-times *no-repeat-times*

Repeated password use is allowed by default.

1.9 Configuring Encrypted Password Storage

1.9.1 Overview

Encrypted password storage is used to store passwords after encryption.

When administrators run the **display current-config** command to display configurations or run the **save terminal** command to save configuration files, passwords configured by users are displayed in ciphertext format if encrypted password storage is enabled and displayed in plaintext format if encrypted password storage is disabled.

1.9.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable encrypted password storage.

service password-encryption-storage

Encrypted password storage is disabled by default.

1.10 Enabling Special Character Detection

1.10.1 Overview

After strong password detection and special character detection are configured, passwords that contain only special characters are invalid and cannot be configured successfully. Special characters include space, tilde (~), backtick (`), exclamation mark (!), at sign (@), number sign (#), dollar sign ($), percent sign (%), caret (^), ampersand (&), asterisk (*), brackets (()), underscore (_), plus sign (+), minus sign (–), equal sign (=), braces ({}), vertical bar (|), square brackets ([]), backslash (\), colon (:), quotation mark ("), semicolon (;), apostrophe ('), angle brackets (<>), comma (,), period (.), and slash (/).

1.10.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable special character detection.

password policy printable-character-check

Special character detection is disabled by default.

1.11 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-1 Monitoring

Command	Purpose
display password policy	Displays configured password security policies.

1.12 Configuration Examples

1.12.1 Requirements

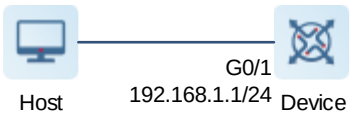
To ensure password complexity, you can configure password policies on a device to limit users' passwords.

Password security requirements are as follows:

- A password must be a string of at least eight characters.
- A password cannot be the same as the account.
- A password contains only one or two types of digits, uppercase letters, and lowercase letters.
- A password is valid for 90 days.
- Passwords must be stored in encrypted mode.
- A password cannot be the same as historical passwords configured in the latest three times.

1.12.2 Topology

Figure 1-1 Configuring Password Policies



1.12.3 Notes

- Set the minimum password length to 8.
- Enable strong password detection.
- Set the password lifecycle to 90 days.
- Enable encrypted password storage.
- Configure repeated password use prevention to disable the use of historical passwords configured in the latest three times.

1.12.4 Procedure

Enter the global configuration mode.

```
Device> privilege
Device# config-system terminal
```

Set the minimum password length to 8.

```
Device(cfg)# password policy min-size 8
```

Enable strong password detection.

```
Device(cfg)# password policy strong
```

Set the password lifecycle to 90 days.

```
Device(cfg)# password policy life-cycle 90
```

Enable encrypted password storage.

```
Device(cfg)# service password-encryption-storage
```

Configure repeated password use prevention to disable the use of historical passwords configured in the latest three times.

```
Device(cfg)# password policy no-repeat-times 3
```

1.12.5 Verification

Run the **display password policy** command to display configured password policies.

```
Device# display password policy
Global password policy configurations:
  Password encryption:                  Enabled
  Password strong-check:                Enabled
  Password forced-password-modify:      Disabled
  Password min-size:                    Enabled (8 characters)
  Password life-cycle:                  Enabled (90 days)
  Password no-repeat-times:             Enabled (max history record: 3)
  Password printable-character-check:   Disabled
```

1.12.6 Configuration Files

```
!
password policy life-cycle 90
password policy no-repeat-times 3
password policy min-size 8
password policy strong
service password-encryption-storage
```

Contents

1 Configuring SSH....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
Protocols and Standards....3
1.2 Configuration Task Summary....4
1.3 Configuring SSH Server....4
1.3.1 Overview....4
1.3.2 Configuration Tasks....4
1.3.3 Enabling SSH Server....5
1.3.4 Configuring the SSH Server Version....5
1.3.5 Configuring the Listening Port....5
1.3.6 Configuring the Key and Key Exchange Algorithm of the SSH Server....6
1.3.7 Configuring the Encryption Algorithm....7
1.3.8 Configuring Authentication Parameters....7
1.3.9 Configuring an ACL....8
1.3.10 Configuring IP Address Blocking....8
1.3.11 Disconnecting an SSH Client Connection....9
1.4 Configuring SCP Server....9
1.4.1 Overview....9
1.4.2 Restrictions and Guidelines....9
1.4.3 Procedure....10
1.5 Configuring SSH Client....10

1.5.1 Overview 10
1.5.2 Configuration Tasks 10
1.5.3 Establishing a Connection with the SSH Server 10
1.5.4 Restoring an Established SSH Client Session 11
1.5.5 Disconnecting a Suspended SSH Client Session 11
1.6 Configuring SCP Client 12
1.6.1 Overview 12
1.6.2 Procedure 12
1.7 Configuring SFTP Server 12
1.7.1 Overview 12
1.7.2 Procedure 12
1.8 Configuring SFTP Client 13
1.8.1 Overview 13
1.8.2 Procedure 13
1.9 Monitoring 14
1.10 Configuration Examples 15
1.10.1 Configuring Line Password Authentication for SSH Users 15
1.10.2 Configuring Local User Authentication for SSH Clients 17
1.10.3 Configuring AAA Authentication for SSH Clients 19
1.10.4 Configuring Public Key Authentication for SSH Clients 21
1.10.5 Configuring SCP Server 22
1.10.6 Configuring SSH Client 24
1.10.7 Configuring SCP Client 25
1.10.8 Configuring SFTP Server 26

1.10.9 Configuring SFTP Client 27

1 Configuring SSH

1.1 Introduction

1.1.1 Overview

Secure Shell (SSH) is a network security protocol that requires encryption and authentication and is used for remote access and file transfer.

SSH is based on the server/client structure. The device supports both the SSH server and client functions. As an SSH server, the device can connect to multiple SSH clients. As an SSH client, the device allows users to establish SSH connections with devices that support the SSH server function.

The SSH function is similar to the Telnet service. However, the encryption and authentication features of SSH can ensure stronger security guarantee for users. When users log in to the device through an insecure network environment, SSH can effectively protect the device against attacks, such as IP address spoofing and plaintext password interception.

1.1.2 Principles

1. SSH Version

SSHv1 and SSHv2 are available, and the two are mutually exclusive. Compared with SSHv1, SSHv2 provides higher performance and security.

The device supports both versions and supports SSH interaction using the IPv4 and IPv6 addresses. Unless otherwise specified, SSH in this document refers to SSHv2.

2. SSH Exchange Process

Interaction between an SSH client and an SSH server is classified into seven phases: connection establishment, version negotiation, key exchange and algorithm negotiation, user authentication, session request, session interaction, and session closure.

- Connection establishment
  a The server listens on Transmission Control Protocol (TCP) port 22 and waits for a connection request from the client.
  b The client initiates a TCP connection request to port 22 of the server to establish a TCP connection with the server.
- Version negotiation
  a The server sends a version negotiation packet to the client.
  b The client receives and processes the packet and replies the version to be used to the server.
  c The server processes the reply from the client and determines whether version negotiation is successful.
- Key exchange and algorithm negotiation

  If the version negotiation succeeds, key exchange and algorithm negotiation are performed.

a The server and the client exchange the algorithm negotiation packet with each other and determine the final algorithm based on their capacity.

b The server and the client work together to generate a session key and session ID according to the key exchange algorithm and host key, which will be applied to subsequent user authentication, data encryption, and data decryption.

- User authentication

  After the encrypted channel is set up, user authentication is performed.

  a The client sends an authentication request to the server.

  b The server repeatedly conducts authentication for the client until the authentication succeeds or the server shuts down the connection because the maximum number of authentication attempts is reached.

- Session request

  a After the authentication is successful, the client sends a session request to the server.

  b The server waits and processes the client's session request. After the session request is successfully processed, the SSH client and server enter the session interaction phase.

- Session interaction

  In this phase, bidirectional transmission and processing of encrypted data are allowed.

  a The client sends commands that need to be executed to the server.

  b The server decrypts, parses, and processes these commands.

  c The server encrypts and sends the processing result to the client.

  d The client decrypts and processes the message from the server.

- Session closure

  The server and client disconnect the connection, and close the session.

3. User Authentication Mechanism

SSH supports password authentication, public key authentication, and other client authentication mechanisms.

- Password authentication

  Password authentication authenticates a client by verifying the username and password.

  a The client sends encrypted username and password to the server to request password authentication.

  b After receiving the request, the server decrypts the information, authenticates the client's identity through the Authentication, Authorization and Accounting (AAA) service (including local authentication and remote authentication), and sends an authentication success or failure message to the client.

- Public key authentication

  Public key authentication uses digital signature algorithms, such as Rivest-Shamir-Adleman (RSA) and Digital Signature Algorithm (DSA) to authenticate a client.

  a The client sends a public key authentication request to the server. This request contains information, including the username, public key, and public key algorithm.

  b After receiving the request, the server checks whether the public key is valid.

c If not, the server directly returns an authentication failure message. If yes, the server performs digital signature authentication on the client and returns a message indicating successful or failed authentication.

Note

Public key authentication applies only to SSHv2 clients.

4. Public Key Algorithm

A public key algorithm is an asymmetric encryption algorithm used to encrypt communication data between an SSH client and server to ensure data transfer security.

The device supports three public key algorithms: RSA, DSA, Elliptic Curves Cryptography (ECC), and ED25519.

- RSA

  RSA is based on most factorization algorithms and can be used for key and signature exchange. Due to low encryption and decryption efficiency, it does not apply to encryption and decryption of a large number of data.

- DSA

  Compared with RSA, DSA can be used only for signature and cannot be used for data encryption and decryption or key exchange. However, its efficiency is much higher than that of RSA.

- ECC

  ECC is based on the discrete logarithm algorithm and is difficult to be cracked down. In same security conditions, ECC has higher processing efficiency and lower storage and bandwidth requirements than RSA and DSA.

- ED25519

  ED25519 is the EdDSA signature scheme using SHAKE256 and Curve448 defined in RFC 8032. Compared with RSA, it is equally secure and has a shorter key, with a high signature and verification performance.

Protocols and Standards

- RFC 4251: The Secure Shell (SSH) Protocol Architecture
- RFC 4252: The Secure Shell (SSH) Authentication Protocol
- RFC 4253: The Secure Shell (SSH) Transport Layer Protocol
- RFC 4254: The Secure Shell (SSH) Connection Protocol
- RFC 4419: Diffie-Hellman Group Exchange for the Secure Shell (SSH) Transport Layer Protocol
- RFC 4716: The Secure Shell (SSH) Public Key File Format
- RFC 4819: Secure Shell Public Key Subsystem
- RFC 3526: More Modular Exponential (MODP) Diffie-Hellman groups for Internet Key Exchange (IKE)
- RFC 2409: The Internet Key Exchange (IKE)
- RFC 1950: ZLIB Compressed Data Format Specification version 3.3
- RFC 8030: EdDSA Instances

1.2 Configuration Task Summary

All the configuration tasks below are optional. Select the configuration tasks as required.

(1) Configuring SSH Server

- Enabling SSH Server
- (Optional) Configuring the SSH Server Version
- (Optional) Configuring the Listening Port
- (Optional) Configuring the Key and Key Exchange Algorithm of the SSH Server
- (Optional) Configuring the Encryption Algorithm
- (Optional) Configuring Authentication Parameters
- (Optional) Configuring an ACL
- (Optional) Configuring IP Address Blocking
- (Optional) Disconnecting an SSH Client Connection

(2) Configuring SCP Server

(3) Configuring SSH Client

- Establishing a Connection with the SSH Server
- (Optional) Restoring an Established SSH Client Session
- (Optional) Disconnecting a Suspended SSH Client Session

(4) Configuring SCP Client

(5) Configuring SSH Server

(6) Configuring SFTP Client

1.3 Configuring SSH Server

1.3.1 Overview

After the SSH server function is configured on a device, SSH clients can remotely log in to the device.

1.3.2 Configuration Tasks

SSH server configuration includes the following tasks:

- Enabling SSH Server
- (Optional) Configuring the SSH Server Version
- (Optional) Configuring the Listening Port
- (Optional) Configuring the Key and Key Exchange Algorithm of the SSH Server
- (Optional) Configuring the Encryption Algorithm
- (Optional) Configuring Authentication Parameters
- (Optional) Configuring an ACL
- (Optional) Configuring IP Address Blocking
- (Optional) Disconnecting an SSH Client Connection

1.3.3 Enabling SSH Server

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the SSH server function.

enable-server service ssh-server

The SSH server function is disabled by default.

(4) (Optional) Enable the SSHv1 function.

ip ssh-server compatible-ssh1x enable

The SSHv1 function is disabled by default.

To use the SSHv1 version, enable the SSHv1 function first.

1.3.4 Configuring the SSH Server Version

1. Overview

The SSH server supports the SSHv1 and SSHv2 versions. Compared with SSHv1, SSHv2 has higher performance and security. During actual configuration, select a proper SSH server version based on the client version and security requirements.

2. Restrictions and Guidelines

The SSH server is compatible with SSHv1 and SSHv2 clients by default, that is, clients of both versions can connect to the SSH server. After an SSH server version is configured, only clients of the corresponding version can connect to the server.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the SSH server version.

ip ssh-server version *version-type*

The SSH server is compatible with the SSHv1 and SSHv2 clients by default.

1.3.5 Configuring the Listening Port

1. Overview

An SSH server uses port 22 to listen to client connections by default. The listening port can be changed through configuration.

Port 22 is a protocol-defined standard port and vulnerable to attacks from malicious users. If an attacker accesses this port frequently in a short time, problems such as bandwidth waste and performance deterioration

may occur. As a result, other users cannot access the server (that is, the DoS attack). Changing the listening port can effectively prevent the problems.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the listening port of the SSH server.

ip ssh-server port *ssh-monitor-port*

The default listening port of the SSH server is port **22**.

1.3.6 Configuring the Key and Key Exchange Algorithm of the SSH Server

1. Overview

The key and key exchange algorithm of the SSH server are configured to establish a communication connection between the client and server.

2. Restrictions and Guidelines

SSHv1 uses an RSA key, and SSHv2 uses an RSA or a DSA key. If an RSA key is generated, both SSHv1 and SSHv2 can use the key. If a DSA key is generated, only SSHv2 can use the key.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Generate the public key of the SSH server.

ssh-crypto key generate { **dsa** | **ecc** | **rsa | ed25519** }

No public key is generated on the SSH server by default.

If no key is generated on the SSH server, this parameter must be configured. To delete the public key of the SSH server, run the **ssh-crypto key remove** { **dsa** | **ecc** | **rsa | ed25519** } command instead of the **unset** form of this command.

(4) Configure the DH key exchange algorithms supported by the SSH server.

ip ssh-server key-exchange { **curve25519_sha256** | **curve25519_sha256@libssh.org** | **dh_group_exchange_sha1** | **dh_group_exchange_sha256** | **dh_group14_sha1** | **dh_group14_sha256** | **dh_group16_sha512** | **dh_group18_sha512** | **dh_group1_sha1** | **ecdh_sha2_nistp256** | **ecdh_sha2_nistp384** | **ecdh_sha2_nistp521** }

The SSHv2 servers support curve25519_sha256, curve25519_sha256@libssh.org, dh_group_exchange_sha256, dh_group14_sha256, dh_group16_sha512, dh_group18_sha512 , ecdh_sha2_nistp256, ecdh_sha2_nistp384 and ecdh_sha2_nistp521 for key exchange, and SSHv1 servers support none by default.

(5) (Optional) Configure the minimum length of the key generated by the key exchange algorithm of the SSH server.

ip ssh-server dh-exchange min-len { **1024** | **2048** }

The default minimum length is **2048** bytes.

1.3.7 Configuring the Encryption Algorithm

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the encryption modes supported by the SSH server.

ip ssh-server cipher-mode { **3des-cbc** | **aes128-cbc** | **aes128-ctr** | **aes128-gcm** | **aes192-cbc** | **aes192-ctr** | **aes256-cbc** | **aes256-ctr** | **aes256-gcm** | **blowfish-cbc** | **cbc** | **ctr** | **des-cbc** | **gcm** | **chacha-only** | **others** } *

The encryption modes supported by the SSH server are **ctr**, **gcm** and **chacha-only** by default.

It is approved that **cbc** and **others** encryption algorithms can be decrypted in a limited period of time. Therefore, organizations or companies that have high security requirements can set the encryption modes supported by the SSH server to **ctr**, **gcm** and **chacha-only** to enhance the security level of the SSH server.

(4) Configure the message authentication algorithms supported by the SSH server.

ip ssh-server hmac-algorithm { **md5** | **md5-96** | **sha1** | **sha1-96** | **sha2-256** | **sha2-512** | **sha2-256-etm** | **sha2-512-etm** }

SSHv1 servers do not support any message authentication algorithms, and SSHv2 servers support the sha2-256, sha2-512, sha2-256-etm and sha2-512-etm message authentication algorithms by default.

1.3.8 Configuring Authentication Parameters

1. Overview

Parameters related to user authentication include the authentication timeout time, number of authentication attempts, and public key authentication.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the user authentication timeout time on the SSH server.

ip ssh-server time-out *timeout-time*

The default user authentication timeout time on the SSH server is **120** seconds.

If authentication still does not succeed when the user authentication timeout time expires, user authentication fails.

(4) Configure the number of authentication attempts allowed on the SSH server.

ip ssh-server authentication-retries *retry-times*

The default maximum number of authentication attempts allowed on the SSH server is **3**.

If the number of authentication attempts of a user by using the account and password exceeds the configured number of authentication attempts, user authentication fails.

(5) Associate with the public key file and username of a client.

ip ssh-server peer *username* **public-key** { **dsa** | **ecc** | **rsa** | **ed25519** } *filename-path*

If a client needs to use public key authentication, the client's public key and username need to be associated with the device. Only SSHv2 supports public key authentication.

1.3.9 Configuring an ACL

1. Overview

An ACL can filter out users who meet specific rules. This prevents unauthorized users from accessing the device through SSH and causing security risks.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ACL on the SSH server.

(IPv4 address)

ip ssh-server access-apply { *acl-name* | *acl-number* }

No IPv4 ACL is configured on the SSH server by default.

(IPv6 address)

ipv6 ssh-server access-apply *acl-name*

No IPv6 ACL is configured on the SSH server by default.

1.3.10 Configuring IP Address Blocking

1. Overview

When the number of authentication failures for login through SSH reaches the configured limit in an authentication failure count period, the source IP address is blocked. That is, the SSH client that uses this source IP address is not allowed to log in to the device to prevent the device being attacked. The SSH client can log in to the device only after the IP address is awakened. This function is enabled by default and can be manually disabled.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable IP address blocking on the SSH server.

ip ssh-server ip-block disable

The IP address blocking function of the SSH server is enabled by default.

Run the **unset** form of this command to disable this feature.

(4) (Optional) Configure the number of authentication failures for blocking an IP address and the time period for counting authentication failures on the SSH server.

ip ssh-server ip-block failed-times *failed-times* **period** *period-time*

The allowed maximum number of authentication failures is **6**, and the time period for counting authentication failures is **5** minutes by default.

(5) (Optional) Configure the time period for awakening blocked IP addresses on the SSH server.

ip ssh-server ip-block reactive *reactive-interval*

Blocked IP addresses are awakened every **10** minutes by default.

After the time period for awakening blocked IP addresses is reached, SSH clients using these IP addresses can log in to the SSH server again.

(6) (Optional) Clear user information with blocked IP addresses and authentication failures.

cleanup ssh ip-block { *ipv4-address* | *ipv6-address* | **all** }

After the user information is cleared, corresponding clients can log in to the SSH server again.

1.3.11 Disconnecting an SSH Client Connection

(1) Enter the privileged EXEC mode.

privilege

(2) Disconnect an established SSH client connection.

remove ssh { *session-id* | **vty** *session-id* }

Serving as an SSH server, the device may connect to multiple SSH clients. You can run this command to forcibly disconnect a client from the device. The client disconnection methods are as follows:

- Specify an SSH session ID. To display the SSH session ID of a client, run the **display ssh** command.
- Specify a Virtual Teletype (VTY) session ID. To display the VTY session ID of a client, run the **display users** command. This command can be used to disconnect SSH connections only.

1.4 Configuring SCP Server

1.4.1 Overview

Secure copy protocol (SCP) is a protocol that supports network file transfer. SCP is implemented based on RCP. RCP is responsible for file transfer, and SSH provides the authentication and encryption functions for RCP.

After the SCP server function is configured on the device, users can run the **scp** command to upload files to or download files from the device. Data exchanged during the process is encrypted for security.

1.4.2 Restrictions and Guidelines

The SSH server function must be enabled before the SCP server function is configured.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the SCP server function.

ip scp-server set enable

The SCP server function is disabled by default.

(4) (Optional) Configure the transmission path for uploading files to or downloading files from the SCP server.

ip scp-server set topdir { **flash:/***path* | **tmp:/***path* | **usb0:/***path* }

The default transmission path for file upload and download is **flash:/**.

1.5 Configuring SSH Client

1.5.1 Overview

When the local device is configured as an SSH client, the device can establish an SSH connection to another device that provides the SSH server service to remotely log in to the device for management operations.

1.5.2 Configuration Tasks

SSH client configuration includes the following tasks:

(1) Establishing a Connection with the SSH Server

(2) (Optional) Restoring an Established SSH Client Session

(3) (Optional) Disconnecting a Suspended SSH Client Session

1.5.3 Establishing a Connection with the SSH Server

(1) Enter the privileged EXEC mode.

privilege

(2) Establish a remote encrypted session with the SSH server as an SSH client.

ssh-server [**oob**] [**-v** { **1** | **2** } | **-c** { **3des** | **aes128-cbc** | **aes192-cbc** | **aes256-cbc** | **aes128-ctr** | **aes192-ctr** | **aes256-ctr** | **aes128-gcm** | **aes256-gcm** | **chacha-poly** } | **-l** *username* | **-m** { **hmac-md5-96** | **hmac-md5-128** | **hmac-sha1-96** | **hmac-sha1-160** | **hmac-sha2-256** | **hmac-sha2-512** | **hmac-sha2-256-etm** | **hmac-sha2-512-etm** } | **-p** *port-num*] * { *ipv4-address* | *hostname* } [**via** *mgmt-name*] [**/src-address** { **ip** *ipv4-address* | **ipv6** *ipv6-address* | **interface** *interface-type interface-number* }] [**/vrf** *vrf-name*]

SSHv1 supports only the DES (56-bit key) and 3DES (168-bit key) encryption algorithms. SSHv2 supports the following Advanced Encryption Standards (AES): ASE128-CBC, AES192-CBC, AES256-CBC, AES128-CTR, AES192-CTR, AES256-CTR, AES128-QCM, AES256-QCM and CHACHA-POLY. If you specify an unmatched encryption or authentication algorithm when selecting an SSH version, the unmatched algorithm will be ignored when a connection is established.

(3) Enter the global configuration mode.

config-system terminal

(4) (Optional) Configure the source interface of the SSH client.

ip ssh-client source-interface *interface-type interface-number*

No SSH client source interface is configured by default.

After the source interface is configured, the SSH client uses the IP address on this interface as the global source address during communication. If no source interface is configured, the source address of SSH packets will be obtained by querying the corresponding route based on the destination address. If no source interface or source IP address is independently specified for an SSH connection, the global configuration is used.

1.5.4 Restoring an Established SSH Client Session

1. Overview

After the device establishes an SSH session with the SSH server as a client, you can press Ctrl+Shift+6+X to exit the session temporarily. After exiting the session, you can run the corresponding command to restore the session.

2. Restrictions and Guidelines

To display information about the SSH connections established by the device as an SSH client, run the **display ssh-session** command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Restore an established SSH client session.

ssh-session *session-id*

1.5.5 Disconnecting a Suspended SSH Client Session

1. Overview

After the device connects to the SSH server as an SSH client, you can run the corresponding command to disconnect an SSH session with the specified session ID.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Disconnect a suspended SSH client session.

remove ssh-session *session-id*

You can run the **display ssh-session** command to display *session-id* of the device.

1.6 Configuring SCP Client

1.6.1 Overview

When the local device is configured as an SCP client, the device can run the **scp** command to upload files to or download files from a device that provides the SCP server service. Data exchanged during the process is encrypted for security.

1.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the source interface of the SCP client.

ip scp client source-interface *interface-type interface-number*

No SCP client source interface is configured by default.

After the source interface is configured, the SCP client uses the IP address on the interface as the global source address during communication. If no source interface is configured, the source address of SCP packets will be obtained by querying the corresponding route based on the destination address. If no source interface or source IP address is independently specified for an SCP connection, the global configuration is used.

(4) Upload files to or download files from the remote SCP server as an SCP client.

scp [**oob**] [**-v** { **1** | **2** } | **-c** { **3des** | **aes128-cbc** | **aes192-cbc** | **aes256-cbc** | **aes128-ctr** | **aes192-ctr** | **aes256-ctr** | **aes128-gcm** | **aes256-gcm** | **chacha-poly** } | **-m** { **hmac-md5-96** | **hmac-md5-128** | **hmac-sha1-96** | **hmac-sha1-160** | **hmac-sha2-256** | **hmac-sha2-512** | **hmac-sha2-256-etm** | **hmac-sha2-512-etm** } | **-p** *port-num*] * *source-file destination-file* [**via** *mgmt-name*] [**/src-address** { **ip** *ipv4-address* | **ipv6** *ipv6-address* | **interface** *interface-type interface-number* }] [**/vrf** *vrf-name*]

1.7 Configuring SFTP Server

1.7.1 Overview

After the SSH File Transfer Protocol (SFTP) server function is configured on the device, users can use commands of the SFTP client console to upload files to and download files from the device. All the data exchanged in this process is transmitted in encrypted mode to provide high security.

1.7.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the SFTP server function.

ip sftp server enable

The SFTP server function is disabled by default.

> ⚠ Caution
>
> To prevent high CPU usage, run the **ip sftp server max-sessions-num** command to limit the maximum number of sessions.

(4) (Optional) Configure the root path for uploading files to or downloading files from the SFTP server.

ip sftp server topdir { **flash:/***path* | **flash2:/***path* | **sd0:/***path* | **sata0:/***path* | **tmp:/***path* | **usb0:/***path* | **usb1:/***path* }

The default root path on the SFTP server is **flash:/**.

(5) (Optional) Configure the maximum number of SFTP session connections.

ip sftp server max-sessions-num *session-numbers*

The maximum number of SFTP session connections is 5 by default.

1.8 Configuring SFTP Client

1.8.1 Overview

After the local device is configured as the SFTP client, you can connect to the SFTP server by using an SFTP command to upload and download files locally or remotely in the client console. All the data exchanged in this process is transmitted in encrypted mode to provide high security.

1.8.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the source interface used by the SFTP client.

ip sftp client source-interface interface-type interface-number

No source interface is configured for the SFTP client by default.

After the source interface is configured, the SFTP client uses the IP address of this interface as the global source address. If no source interface is configured, the device obtains the source address of the SFTP packets by querying the destination address in the corresponding route. If no source interface or address is specified separately for an SFTP connection, the global configuration applies.

(4) (Optional) Configure an SSH version for global SFTP connections.

ip sftp client version { **1** | **2** }

SSHv2 is used for global SFTP connections by default.

(5) (Optional) Configure the global SSH encryption algorithms supported by the SFTP client.

ip sftp client cipher-mode { **3des | aes128-cbc | aes128-ctr | aes128-gcm | aes192-cbc | aes192-ctr | aes256-cbc | aes256-ctr | aes256-gcm** | **chacha-poly** }

The SFTP client supports the following encryption algorithms by default: 3DES-CBC, AES-128-CBC, AES128-CTR, AES128-GCM, AES-192-CBC, AES192-CTR, AES-256-CBC, AES256-CTR, AES256-GCM and CHACHA20-POLY1305.

(6) (Optional) Configure the global SSH message authentication algorithms supported by the SFTP client.

ip sftp client hmac-algorithm { **hmac-md5-128** | **hmac-md5-96** | **hmac-sha1-160** | **hmac-sha1-96** | **hmac -sha2-256** | **hmac -sha2-512** | **hmac -sha2-256-etm** | **hmac -sha2-512-etm** }

SSHv1 does not support any message authentication algorithms, and SSHv2 supports the following message authentication algorithms by default: MD5-128, SHA1-160, SHA1-96, MD5-96, SHA2-256, SHA2-512, SHA2-256-ETM, and SHA2-512-ETM.

(7) (Optional) configure the global Diffie–Hellman (DH) key exchange algorithms used by the SFTP client.

ip sftp client key-exchange { **curve25519_sha256** | **curve25519_sha256@libssh.org** | **dh_group_exchange_sha**1 | **dh_group_exchange_sha256** | **dh_group14_sha**1 | **dh_group14_sha256** | **dh_group16_sha512** | **dh_group18_sha512** | **dh_group**1**_sha**1 | **ecdh_sha**2**_nistp**256 | **ecdh_sha**2**_nistp**384 | **ecdh_sha**2**_nistp**521 }

SSHv1 does not support any key exchange algorithms, and SSHv2 supports the following key exchange algorithms by default: dh_group1_sha1, dh_group14_sha1, dh_group14_sha256, dh_group16_sha512, dh_group18_sha512, dh_group_exchange_sha1, dh_group_exchange_sha256, curve25519_sha256@libssh.org, curve25519_sha256, ecdh_sha2_nistp256, ecdh_sha2_nistp384, and ecdh_sha2_nistp521.

(8) (Optional) Configure the top directory on the SFTP client.

ip sftp client topdir { **flash:**/*path* | **flash2:**/*path* | **sd0:**/*path* | **sata0:**/*path* | **tmp:**/*path* | **usb0:**/*path* | **usb1:**/*path* }

The default top directory on the SFTP client is **flash:/**.

(9) (Optional) Configure a rate limit for file transfer using SFTP.

ip sftp client speed-limit *speed-limit-rate*

The default rate limit for file transfer using SFTP is 8192 kbit/s.

(10) Configure the device as the SFTP client, connect it to the SFTP server, and enter the SFTP client console to manage files.

sftp [**oob**] [**-v** { 1 | 2 } | **-c** { 3**des** | **aes**128-**cbc** | **aes**192-**cbc** | **aes**256-**cbc** | **aes**128-**ctr** | **aes**192-**ctr** | **aes**256-**ctr** | **aes**128-**gcm** | **aes256-gcm** | **chacha-poly** } | **-k** { curve25519_sha256 | curve25519_sha256@libssh.org | **dh_group_exchange_sha1** | **dh_group_exchange_sha256** | **dh_group14_sha1** | **dh_group14_sha256** | **dh_group16_sha512** | **dh_group18_sha512** | **dh_group1_sha1** | **ecdh_sha2_nistp256** | **ecdh_sha2_nistp384** | **ecdh_sha2_nistp521** } | **-m** { **hmac**-**md**5-96 | **hmac**-**md**5-128 | **hmac**-**sha**1-96 | **hmac**-**sha**1-160 | **hmac**-**sha**2-256 | **hmac**-**sha**2-512 | **hmac**-**sha**2-256-etm | **hmac**-**sha**2-512-etm } | **-p** *port-num*] * { *user@hostname* | *user@ip-address* } [**via mgmt** *mgmt-number*] [/**source** { **ip** *ipv4-address* | **ipv**6 *ipv6-address* | interface *interface-type interface-number* }] [/**vrf** *vrf-name*]

1.9 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** commands to output debugging information.

Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging function immediately after use.

Table 1-1 Monitoring

Command	Purpose
display ip ssh-config	Displays effective configurations of the SSH server.
display ssh	Displays information about established SSH connections.
display ssh ip-block { **all** | **list** }	Displays information about blocked IP addresses and authentication failures.
display ss crypto-key pubkey { **dsa** | **ecc** | **rsa** | **ed25519** }	Displays partial public key information of the SSH server.
display ssh-session	Displays information about established SSH client sessions.
debugging ssh	Debugs sessions of the SSH server.
debugging ssh client	Debugs sessions of the SSH client.

1.10 Configuration Examples

1.10.1 Configuring Line Password Authentication for SSH Users

1. Requirements

After the SSH server function is enabled on a device, users can use the password configured in line configuration mode to log in to the device.

2. Topology

Figure 1-1 Configuring Line Password Authentication for SSH Clients



3. Notes

- Enable the SSH server function.
- Configure the line password.
- Enable password authentication in the line.

4. Procedure

(1) Enable the SSH server function.

```
Device> privilege
Device# config-system terminal
Device(cfg)# enable-server service ssh-server
```

(2) Set the line password to **password1**.

```
Device(cfg)# line vty 0 4
Device(cfg-line)# password password1
```

(3) Enable password authentication in the line.

```
Device(cfg-line)# line-login
```

5. Verification

- Use the client software to log in to the device through SSH. The password is **password1**, and the username can be any value. Verify that the login is successful. Use the PuTTY software as an example. The login page is as follows:

```
login as: anyname
anyname@192.168.1.1's password:



Last login: Apr  2 2021 10:00:00 from 192.168.1.2 through ssh.
Device#
```

- Run the **display ip ssh-server** command to verify that the SSH server takes effect.

```
Device# display ip ssh-config
SSH Enable - version 2.0
SSH Port:                       22
SSH Cipher Mode:                ctr,gcm
SSH HMAC Algorithm:             sha2-256,sha2-512
Authentication timeout: 120 secs
Authentication retries: 3
SSH SCP Server:                 disabled
SSH dh-exchange min-len: 2048
SSH SFTP Server:                disable
SSH ip-block:                   enabled
```

- Run the **display ssh** command to display the SSH connection information. Verify that a connection with username **anyname** (used during login) exists.

```
Device# display ssh
Connection Version Encryption          Hmac              Compress     State                Username
         1       2.0 aes256-ctr        hmac-sha2-256 none         Session started anyname
```

- Run the **display users** command to display information about connected clients.

```
Device#display users
Line                User            Host(s)                 Idle          Location
---------------- ------------ -------------------- ---------- -------------------
   0 con 0          ---              idle                      00:00:16     ---
```

```
*  1 vty 0          ---          idle                 00:00:00     192.168.1.2
```

6. Configuration Files

```
!
enable-server service ssh-server
!
line console 0
line vty 0 4
 line-login
 password 7 $10$396$mnO2p3SVaxm7$
!
```

1.10.2 Configuring Local User Authentication for SSH Clients

1. Requirements

After the SSH server function is enabled on a device, users can use the local account created on the device to log in to the device through SSH.

2. Topology

Figure 1-2 Configuring Local User Authentication for SSH Clients



3. Notes

- Enable the SSH server function.
- Add a local user account.
- Enable local user authentication in the line.

4. Procedure

(1) Enable the SSH server function.

```
Device> privilege
Device# config-system terminal
Device(cfg)# set service ssh-server
```

(2) Add a local user whose username is **hostname1** and password is **password1**.

```
Device(cfg)# username hostname1 password password1
```

(3) Enable local user authentication in the line.

```
Device(cfg)# line vty 0 4
Device(cfg-line)# login local
```

5. Verification

- Use the client software to log in to the device through SSH. The username is **hostname1**, and the password is **password1**. Verify that the login is successful. Use the PuTTY software as an example. The login page is

as follows:

```
login as: hostname1
hostname1@192.168.1.1's password:



Last login: Apr    2 2021 10:00:00 from 192.168.1.2 through ssh.
Device>
```

- Run the **display ip ssh-server** command to verify that the SSH server takes effect.

```
Device# display ip ssh-server
SSH Enable - version 2.0
SSH Port:                       22
SSH Cipher Mode:                ctr,gcm
SSH HMAC Algorithm:             sha2-256,sha2-512
Authentication timeout: 120 secs
Authentication retries: 3
SSH SCP Server:                 disabled
SSH dh-exchange min-len: 2048
SSH ip-block:                   enabled
```

- Run the **display ssh-server** command to display the SSH connection information. Verify that a connection with username **hostname1** exists.

```
Device# display ssh-server
Connection Version Encryption          Hmac              Compress     State                Username
         1       2.0 aes256-ctr        hmac-sha2-256 none          Session started hostname1
```

- Run the **display users** command to display information about connected clients. Verify that a client whose username is **hostname1** exists.

```
Device#display users
Line                 User           Host(s)                Idle          Location
---------------- ------------ -------------------- ---------- -------------------
    0 con 0          ---            idle                    00:00:16     ---
*   1 vty 0          hostname1      idle                     00:00:00     192.168.1.2
```

6. Configuration Files

```
!
username hostname1 password 7 $10$2fe$Rd/C4ijEl+UN$
!
set service ssh-server
!
line console 0
line vty 0 4
 login local
!
```

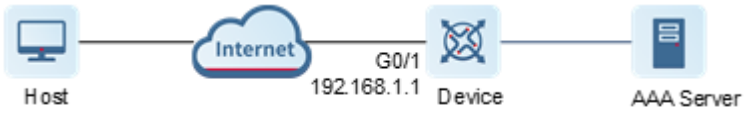
1.10.3 Configuring AAA Authentication for SSH Clients

1. Requirements

SSH clients can be authenticated using the AAA mode.

2. Topology

Figure 1-3 Configuring AAA Authentication for SSH Clients



3. Notes

- Configure the AAA server and add an AAA user on the AAA server.
- Enable the SSH server function.
- Enable the AAA service on the device and set the SSH authentication method to AAA server.
- Enable AAA authentication in the line.

4. Procedure

(1) Configure the AAA server and add an AAA user whose username is **hostname1** and password is **password1** on the AAA server. (For details about AAA server configuration, see *Configuring AAA*.)

(2) Enable the SSH server function.

```
Device> privilege
Device# config-system terminal
Device(cfg)# enable-server service ssh-server
```

(3) Enable AAA services.

```
Device(cfg)# aaa-config new-pattern
```

(4) Configure the address and shared key of the AAA server (a RADIUS server is used as an example here, the IP address is set to **192.168.1.3**, and the shared key is set to **radiuskey**).

```
Device(cfg)# rds-server host 192.168.1.3 share-key radiuskey
```

(5) Configure the login authentication method list **login-method** and select **radius** as the authentication method. The backup method **local** is configured. This ensures that a backup solution is available when the AAA server is faulty.

```
Device(cfg)# aaa-config authentication line-login login-method server-group radius native
```

(6) Enable AAA authentication in the line.

```
Device(cfg)# line vty 0 4
Device(cfg-line)# line-login authentication login-method
```

5. Verification

- Use the client software to log in to the device through SSH. The username is **hostname1**, and the password is **password1**. Verify that the login is successful. Use the PuTTY software as an example. The login page is

as follows:

```
login as: hostname1
hostname1@192.168.1.1's password:



Last login: Apr  2 2021 10:00:00 from 192.168.1.2 through ssh.
Device>
```

- Run the **display ip ssh-config** command to verify that the SSH server takes effect.

```
Device# display ip ssh-config
SSH Enable - version 2.0
SSH Port:                       22
SSH Cipher Mode:                ctr,gcm
SSH HMAC Algorithm:             sha2-256,sha2-512
Authentication timeout: 120 secs
Authentication retries: 3
SSH SCP Server:                 disabled
SSH dh-exchange min-len: 2048
SSH SFTP Server:                disable
SSH ip-block:                   enabled
```

- Run the **display ssh** command to display the SSH connection information. Verify that a connection with username **hostname1** exists.

```
Device# display ssh
Connection Version Encryption          Hmac              Compress     State                Username
         1       2.0 aes256-ctr        hmac-sha2-256 none           Session started hostname1
```

- Run the **display users** command to display information about connected clients. Verify that a client whose username is **hostname1** exists.

```
Device#display users
Line                 User           Host(s)                Idle          Location
---------------- ------------ -------------------- ---------- -------------------
    0 con 0          ---            idle                    00:00:16     ---
*   1 vty 0          hostname1      idle                     00:00:00     192.168.1.2
```

6. Configuration Files

```
!
aaa-config new-pattern
aaa-config authentication line-login login-method server-group radius native
!
rds-server host 192.168.1.3 share-key 7 $10$3b4$ln0s66i8XfEi$
!
line vty 0 4
 line-login authentication login-method
!
```

1.10.4 Configuring Public Key Authentication for SSH Clients

1. Requirements

When the device serves as the SSH server, users can use public key authentication to log in to the device without needing a password.

2. Topology

Figure 1-4 Configuring Public Key Authentication for SSH Clients



3. Notes

- Use the client software to generate a pair of keys (public key and private key).
- Copy the public key file to the flash memory of the device and associate the key with the user.

4. Procedure

(1) Configure a local user and enable local user authentication in a line. For details, see 1.10.2 Configuring Local User Authentication for SSH Clients.

(2) Use the client software to generate a pair of keys. The Ubuntu environment is used as an example. The generated private key is **id_rsa** and public key is **id_rsa.pub**.

```
root@ubuntu:~# ssh-keygen
Generating public/private rsa key pair.
Enter file in which to save the key (/root/.ssh/id_rsa):
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /root/.ssh/id_rsa.
Your public key has been saved in /root/.ssh/id_rsa.pub.
The key fingerprint is:
SHA256:6a/sWJjpx97QQEs9Aoiph5Ry452ybIyNyZUsT48Irrw root@ubuntu
The key's randomart image is:
+---[RSA 2048]----+
|   .o ..         |
|.o= .   . .      |
|o=.o..   + o     |
|+.+=o   o + .    |
|+O*oo    S       |
|++*o . = o       |
|o.      +.+ .    |
|..     . +o+     |
| E.    o+=.o     |
+----[SHA256]-----+
```

(3) Copy the public key file to the flash memory of the device (omitted).

(4) Associate the public key with **hostname1**.

```
Device> privilege
Device# config-system terminal
Device(cfg)# ip ssh-server peer hostname1 public-key rsa flash:test_key.pub
```

5. Verification

Verify that users can log in to the device through public key authentication. (The Ubuntu environment is used as an example.)

```
root@ubuntu:~/.ssh# ls
id_rsa    id_rsa.pub
root@ubuntu:~/.ssh# ssh-server -i id_rsa hostname1@192.168.1.1



Last login: Apr 2 2021 10:00:00 through console.
Device>
```

6. Configuration Files

```
!
user-account hostname1 password 7 $10$2fe$Rd/C4ijEl+UN$
!
ip ssh-server peer hostname1 public-key rsa flash:id_rsa.pub
!
enable-server service ssh-server
!
line console 0
line vty 0 4
 line-login local
!
```

1.10.5 Configuring SCP Server

1. Requirements

When the device serves as the SCP server, users can run the **scp** command to upload files to and download files from the device.

2. Topology

Figure 1-5 Configuring SCP Server



3. Notes

- Enable the SSH server function.
- Enable the SCP server function.

4. Procedure

(1) Configure a local user and enable local user authentication in a line. For details, see 1.10.2 Configuring Local User Authentication for SSH Clients.

(2) Enable the SSH server function.

```
Device> privilege
Device# config-system terminal
Device(cfg)# enable-server service ssh-server
```

(3) Enable the SCP server function.

```
Device(cfg)# ip scp-server set enable
```

5. Verification

- Verify that the client can run the **scp** command to upload files to the device. The Ubuntu environment is used as an example here.

(1) In the Ubuntu shell environment, run the **scp** command to upload **file.txt** to the device.

```
root@ubuntu:~# scp file.txt hostname1@192.168.1.1:/
hostname1@192.168.1.1's password:
file.txt                                      100%    14        0.0KB/s     00:00
```

(2) Check whether **file.txt** exists in the flash directory of the device.

```
Device> privilege
Device# dir flash:file.txt
Number   Properties       Size                  Time                        Name
------   ----------   ----------   ------------------------   --------------------
  1          -rwx          14B          Fri Apr 2 10:00:00 2021    file.txt
```

- Verify that the client can run the **scp** command to download files from the device. The Ubuntu environment is used as an example here.

(1) The **file.txt** file exists in the flash directory of the device.

```
Device> privilege
Device# dir flash:file.txt
Number   Properties       Size                  Time                        Name
------   ----------   ----------   ------------------------   --------------------
  1          -rwx          14B          Fri Apr 2 10:00:00 2021    file.txt
```

(3) In the Ubuntu shell environment, run the **scp** command to download **file.txt** from the device and rename it as **file1.txt**.

```
root@ubuntu:~# scp hostname1@192.168.1.1:/file.txt file1.txt
hostname1@192.168.1.1's password:
file.txt                                      100%    14        0.0KB/s     00:00
```

(4) In the Ubuntu shell environment, verify that the **file1.txt** file has been downloaded.

```
root@ubuntu:~# ls -l file1.txt
-rwxr--r-- 1 root root 14 Apr 2 10:00 file1.txt
```

6. Configuration Files

```
!
```

```
user-account hostname1 password 7 $10$2fe$Rd/C4ijEl+UN$
!
enable-server service ssh-server
!
ip scp-server set enable
!
line console 0
line vty 0 4
 line-login local
!
```

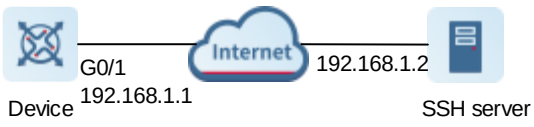
1.10.6 Configuring SSH Client

1. Requirements

When the device serves as an SSH client, the device can log in to the SSH server through SSH.

2. Topology

Figure 1-6 Configuring SSH Client



3. Notes

- Configure a valid user on the SSH server.
- Run the **ssh-server** command to log in to the SSH server.

4. Procedure

(1) Configure a valid user on the SSH server. (For details, see the SSH server configuration.)

(2) Run the **ssh-server** command to log in to the SSH server.

```
Device> privilege
Device# ssh -l admin 192.168.1.2
```

5. Verification

- Verify that the device logs in to the SSH server successfully.

```
Device# ssh -l hostname1 192.168.1.2
%Trying 192.168.1.2, 22,...open
hostname1@192.168.1.2's password:



Last login: Apr 2 2021 10:00:00 from 192.168.1.1 through ssh.
```

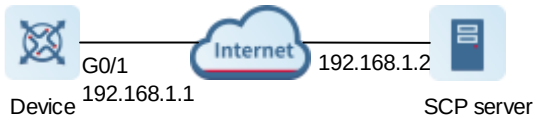
1.10.7 Configuring SCP Client

1. Requirements

When the device serves as an SCP client, the device can run the **scp** command to upload files to and download files from the SCP server.

2. Topology

Figure 1-7 Configuring SCP Client



3. Notes

- Configure a valid user on the SCP server.
- Run the **scp** command to upload files to or download files from the SCP server.

4. Procedure

(1) Configure a valid user on the SCP server. For details, see the SCP server configuration.

(2) Run the **scp** command to copy **file.txt** from the SCP server to the flash memory of the device.

```
Device> privilege
Device# scp hostname1@192.168.1.2:/file.txt flash:/file.txt
```

(3) Run the **scp** command to copy local file **file.txt** from the device to the SCP server.

```
Device> privilege
Device# scp hostname1@192.168.1.2:/file.txt flash:/file.txt
```

5. Verification

- Verify that the file on the SCP server is successfully copied to a local directory.

```
Device# scp oob flash:/file.txt hostname1@192.168.1.2:/file.txt

%Trying 192.168.1.2, 22,...open
hostname1@192.168.1.2's password:
Press Ctrl+C to quit
!
Transmission success.
```

- Verify that the local file is successfully copied to the SCP server.

```
Device# scp hostname1@192.168.1.2:/file.txt flash:/file.txt

%Trying 192.168.1.2,...open
hostname1@192.168.1.2's password:
Press Ctrl+C to quit
!
Transmission success.
```

1.10.8 Configuring SFTP Server

1. Requirements

The device functions as the SFTP server to receive file upload and download commands from the SFTP client.

2. Topology

Figure 1-8 Topology for Configuring the SFTP Server



3. Notes

- Enable the SSH service and configure the username and password.
- Enable the SFTP server function.

4. Procedure

(1) Enable the SSH service.

```
Device> privilege
Device# config-system terminal
Device(cfg)# enable-server service ssh-server
Device(cfg)# username aaa password Ruijie.123
```

(2) Enable the SFTP server function.

```
Device(cfg)# ip sftp server enable
```

(3) Connect the PC to the SFTP server using the SFTP client or the **sftp** command.

```
PC> sftp aaa@192.168.1.1
Connected to 192.168.1.1.
sftp>
```

5. Verification

- Run the **ls** command to check the files in the device directory.

```
sftp> ls -l
drwxrwxrwx       2 root         root                160 Dec 17    2020 addr
drwxr-xr-x       3 root         root                224 Dec 17    2020 cmpnt
-rw-rw-rw-       1 root         root              10162 Aug    8 15:30 config.text
-rw-r--r--       1 root         root               9002 Oct 22    2021 config.text.bak
-rwxrwxrwx       1 root         root                171 Aug    8 15:30 config_vsu.dat
......
sftp>
```

- Check that files are downloaded to the device directory.

```
sftp> get config.text
Fetching /data/config.text to config.text
/data/config.text                                  100% 2295KB     6.4MB/s     00:00
sftp>
```

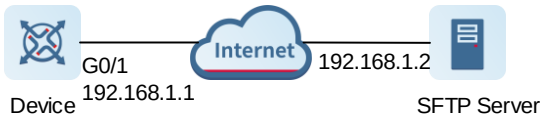
1.10.9 Configuring SFTP Client

1. Requirements

The device functions as the SFTP client. The **sftp** command can be used to establish a connection for accessing the console to upload files to and download files from the SFTP server.

2. Topology

Figure 1-9 Topology for Configuring the SFTP Client



3. Notes

- Configure users on the SFTP server.
- Run the **sftp** command to connect the device to the SFTP server and access the console.
- Run the **put** and **get** commands to upload and download files.

4. Procedure

(1) Configure available users on the SFTP server. (The configuration is omitted. For details, see the *SFTP Server User Manual*.)

(2) Run the **sftp** command on the device to access the client console.

```
Device> privilege
Device# sftp admin@192.168.1.2
%Trying 192.168.1.2, 22,...open
admin@192.168.1.2's password:
Connected to 192.168.1.2:22.
sftp>
```

(3) Run the **get** command to copy the **file.txt** file from the SFTP server to the flash drive of the device.

```
sftp> get file.txt
```

(4) Run the **put** command to copy the local file **file.txt** to the SFTP server.

```
sftp> put file.txt
```

5. Verification

- Check that the file on the SFTP server is copied to the local device.

```
sftp> get file.txt
Fetching /home/admin/file.txt to file.txt
```

```
/home/admin/file.txt                          100% 2295KB     1.7MB/s   00:00
sftp>
```

- Check that the local file is copied to the SFTP server.

```
sftp> put file.txt
Uploading file.txt to /home/admin/file.txt
file.txt                                      100% 2295KB     1.5MB/s   00:00
sftp>
```

Contents

1 Configuring CPP....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Basic Concepts....1
1.1.3 Processing Process....2
1.2 Configuration Task Summary....3
1.3 Configuring Basic Attributes of CPP....4
1.3.1 Overview....4
1.3.2 Restrictions and Guidelines....4
1.3.3 Procedure....5
1.4 Configuring Alarm Print Function of CPP....5
1.4.1 Overview....5
1.4.2 Restrictions and Guidelines....6
1.4.3 Procedure....6
1.5 Configuring Automatic Protection Function....6
1.5.1 Overview....6
1.5.2 Restrictions and Guidelines....7
1.5.3 Procedure....7
1.6 Configuring Automatic Protection Function for Ports....7
1.6.1 Overview....8
1.6.2 Restrictions and Guidelines....8
1.6.3 Procedure....8

1.7 Monitoring 9

1.8 Configuration Examples..... 11

1.8.1 Configuring Basic Functions of CPP 11

1 Configuring CPP

1.1 Introduction

1.1.1 Overview

Many malicious attacks exist in the network environment. CPU attack packets significantly increase the CPU usage of a device, which affects the CPU to process and respond to the CPU to normal services and thus causes device function exceptions. CPU protect policy (CPP) protects the CPU of a device. By controlling the traffic of different types of packets sent to the CPU of the device and scheduling their priorities, users can defend against malicious attacks in the network and ensure that the CPU processes normal protocol packets.

1.1.2 Basic Concepts

1. QoS

As a network security mechanism, Quality of Service (QoS) is a technology that is used to solve network delay and congestion.

Differentiated Service (DiffServ) is a typical model of QoS that provides different services to packets by classifying service streams.

2. Bandwidth

Bandwidth specifies the maximum data transmission rate and the rate threshold for packet transmission in this document. When the data transmission rate exceeds the bandwidth, packets beyond the bandwidth are discarded. Therefore, the transmission rates of different types of packets must be equal to or smaller than the bandwidth. Both the bandwidth and data transmission rate use the unit of packets per second (pps).

3. L2/L3/L4

L2/L3/L4 means a hierarchical structure of packets based on the TCP/IP model. L2 means layer-2 headers, namely, the Ethernet encapsulation part; L3 means layer-3 headers, namely, the IP encapsulation part; L4 means layer-4 headers, usually, the TCP/UDP encapsulation part.

4. Priority Queue

The strict priority (SP) queue scheduling algorithm is a QoS scheduling algorithm. This scheduling algorithm requires packets to be processed based on the queue priority from high to low. The CPU must process first the packets in a high-priority queue and then those in a low-priority queue.

Packets to be sent to the CPU are cached in the SP queue and scheduled according to the algorithms of the SP queue. Packet scheduling is the process of selecting and transmitting packets in a priority queue to the CPU.

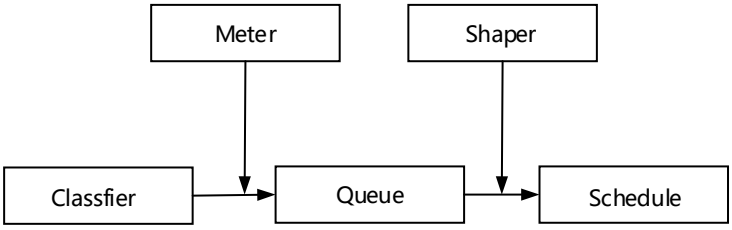
5. CPU Interface

A device caches packets before it sends them to the CPU. Sending the packets to the CPU is similar to a packet output process from the port. The CPU port is a virtual port through which packets are sent to the CPU. The priority queue and SP queue scheduling algorithms apply to only data packets that are transmitted through this CPU port.

1.1.3 Processing Process

The CPP function protects the CPU by using the standard QoS DiffServ model.

Figure 1-1 CPP Implementation Model



1. Classifier

The classifier classifies all packets to be sent to the CPU by the L2, L3 and L4 information of the packets. Classifying packets is the basis for implementing QoS policies. In subsequent actions, CPP can, according to the classification, implement different policies and provide different services. Packets on a device are classified by the supported protocols and management functions, for example, Bridge Protocol Data Unit (BPDU) packets of Spanning Tree Protocol (STP) and packets of Internet Control Message Protocol (ICMP). Packet types cannot be customized.

Note

Packet classification varies with products due to hardware difference. For details, see *Specifications* for details.

2. Meter

The meter limits the rates of different types of packets based on the set rate thresholds. You can set different rate thresholds for various packet types. When the rate of a packet type exceeds the threshold, the packets beyond the threshold are discarded.

By using the meter, you can control the rate of a packet type sent to the to the CPU within a threshold to prevent specific attack packets from exerting great impacts on the CPU resources. This is the level-1 protection of CPP.

3. Queue

Queues are used to classify packets at level 2. You can select the same queue for different packet types; queues cache packets in a device and provide services for the scheduler and shaper.

CPP queues are SP queues. The SPs of the packets are determined based on the time when they are added to a queue. Packets with a larger queue number have a higher priority.

4. Scheduler

The scheduler schedules packets based on SPs of queues. Packets in a high-priority queue are first scheduled.

Before being scheduled, the packets to be sent to the CPU are cached in queues. When being scheduled, the packets are sent to the CPU for processing.

Note

Only the SP scheduling policy is supported and cannot be modified.

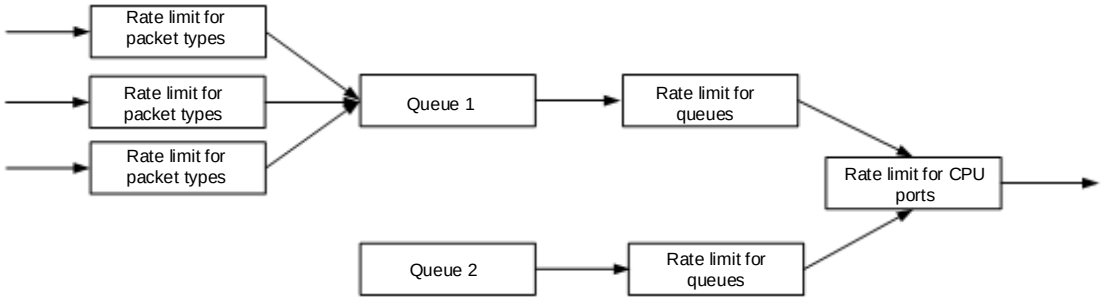
5. Shaper

The shaper shapes the packets to be sent to the CPU. When the actual rate of packets is greater than the shaping threshold, the packets must continue to cache in the queue and cannot be scheduled. When packet rates fluctuate, the shaper ensures that the rates of packets sent to the CPU are smooth (not more than the shaping threshold).

When the shaper is available, packets in a low-priority queue may be scheduled before those in a high-priority queue. If the rate of packets in a queue with a priority exceeds the shaping threshold, scheduling of the packets in this queue is stopped temporarily. Therefore, the shaper can prevent only the packets in high-priority queues from being scheduled.

Since the shaper limits the scheduling rates of packets, it provides the rate limit function. The shaper provides level-2 rate limit for priority queues and all packets sent to the CPU (CPU port). The shaper and meter functions provide 3-level rate limit and level-3 protection for the CPU.

Figure 1-2 3-Level Rate Limit of CPP



1.2 Configuration Task Summary

CPP configuration includes the following tasks:

- Notes
- Table 1-1 describes CPP functions supported by different products.

Table 1-1 CPP Functions Supported by Different Products

Function	Product Supporting the Function
Configuring Automatic Protection Function	S6980-64QC and S6580-24DC8QC
Configuring Automatic Protection Function for Ports	S6980-64QC and S6580-24DC8QC

- Configuring Basic Attributes of CPP
- Configuring Alarm Print Function of CPP
- Configuring Automatic Protection Function
- Configuring Automatic Protection Function for Ports

1.3 Notes

Table 1-1 describes CPP functions supported by different products.

Table 1-2 CPP Functions Supported by Different Products

Function	Product Supporting the Function
Configuring Automatic Protection Function	S6980-64QC and S6580-24DC8QC
Configuring Automatic Protection Function for Ports	S6980-64QC and S6580-24DC8QC

1.4 Configuring Basic Attributes of CPP

1.4.1 Overview

The implementation model of CPP includes five basic functional components: classifier, meter, queue, scheduler, and shaper. A device uses fixed classification policies for the classifier and SP scheduling algorithms for the scheduler.

In different network scenarios, users can configure personalized CPPs based on the meter, queue, and shaper to prevent packet attacks and network flapping from affecting normal operation of the CPU of a device.

- Configure a bandwidth for a specific packet type based on the meter to limit the packet transmission rate. Packets beyond the bandwidth are directly discarded.
- Configure a priority queue for a specific packet type based on the queue function. Packets in a high-priority queue are scheduled first.
- Configure a bandwidth for a CPU port and that for a specific priority queue to limit the packet transmission rate. Packets beyond the bandwidth are directly discarded.

1.4.2 Restrictions and Guidelines

- In one of the following situations, you must adjust the bandwidth for a specific packet type:
  - If a type of packets are discarded though they do not initiate attacks, you must increase the bandwidth for this packet type.
  - If a type of packets initiate attacks and cause CPU running exception, you must decrease the bandwidth for this packet type.
- In one of the following situations, you must adjust the priority queue for a specific packet type:
  - If a type of packets initiate attacks and cause exception in other packets in the same queue, you may put this type of packets into an unused queue.
  - A type of packets may not be discarded, and they are in the same queue with other types of packets in use. To prevent the bandwidth for other types of packets in the same queue from discarding packets, you can put this type of packets into a high-priority queue.
- In one of the following situations, you must adjust the bandwidth for a priority queue:
  - If the bandwidth for a type of packets increases, causing the available bandwidth for other packets in the same priority queue to decrease, you must appropriately increase the bandwidth for this priority queue.

- If attack packets are put into a priority queue and no other packets in use stay in the same queue, you should decrease the bandwidth for this priority queue.

- When you configure a bandwidth for a specific packet type, note that decreasing the bandwidth may affect the normal service traffic of this packet type. To apply CPP based on users, use the NFPP function. For more information about NFPP configuration, see *Configuring NFPP*.
- The bandwidth configuration for a packet type must match the bandwidth configuration for a priority queue. For the three levels of CPU protection, applying only one level of protection may cause negative effects. For example, if you increase the bandwidth for a packet type, you must also adjust the bandwidth for its priority queue. Otherwise, the other packets in the same priority queue may be affected.
- You are not advised to modify the bandwidth of the CPU port.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the basic attributes of CPP, and configure at least one task.

- Configure a bandwidth for a packet type.

  cpu-protect type *packet-type* **bandwidth** *bandwidth-value*

  Each packet type has a rate threshold (bandwidth) by default, and this rate threshold may be modified, but may not be removed.

- Configure a priority queue for a packet type.

  cpu-protect type *packet-type* **traffic-class** *traffic-class-number*

  Each packet type has a priority queue by default.

  The default priority queues for packet types depend on the actual product version.

- Configure a bandwidth for a priority queue.

  cpu-protect traffic-class *traffic-class-number* **bandwidth** *bandwidth-value*

  A shaping threshold (bandwidth) is determined for each priority queue by default, and this shaping threshold may be modified, but may not be removed.

- Configure a bandwidth for the CPU interface.

  cpu-protect cpu bandwidth *bandwidth-value*

  No bandwidth is configured for a CPU port by default.

1.5 Configuring Alarm Print Function of CPP

1.5.1 Overview

After the alarm print function of CPP is enabled, you can view the alarm logs of lost packets and detect packet attacks to the CPU of the device timely. For the normal service packets that are discarded falsely, you can adjust the CPP configuration to avoid subsequent packet loss.

The configuration effect of the alarm print function is as follows:

- After the alarm print function is enabled for a packet type, the corresponding alarm log is printed if packets of this type are lost.
- After the alarm print function is enabled for a priority queue, the alarm log is printed if packets in this queue are lost.
- After an alarm period is configured, a device periodically detects the lost packets sent to the CPU, and the alarm is printed only once for the packet loss in this specified period.

1.5.2 Restrictions and Guidelines

- You can configure the alarm print function with two methods. Select either of them to validate this function. Commands configured between the configuration methods produce the same function effect. If they are configured, the later configured command takes effect. To facilitate management, you are advised to adopt one configuration method.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an alarm period of CPP. Please configure only one task.

- **cpu-protect alarm period** *period-value*
- **cpp-warn warn-period** *interval*

The default alarm period of CPP is **10** minutes.

(4) Enable the alarm print function for a packet type. Please configure only one task.

- **cpu-protect alarm type** *packet-type* **enable**
- **cpp-warn type** *packet-type* **warn**

The alarm print function for packets is not enabled by default.

(5) Enable the alarm print function for a queue. Please configure only one task.

- **cpu-protect alarm traffic-class** *traffic-class-number* **enable**
- **cpp-warn traffic-class** *traffic-class-number* **warn**

The alarm print function for priority queues is not enabled by default.

1.6 Configuring Automatic Protection Function

Note

Only S6980-64QC and S6580-24DC8QC support configuration of automatic protection.

1.6.1 Overview

The automatic protection function allows a device to protect the application layer data based on sessions. The function ensures the normal operation of attacked services.

- After the automatic protection function is enabled based on packet types, the dynamic link protocol packets of the service module can get sufficient signaling to reduce protocol flapping when the packets are attacked.
- After it configures a domain for the customized security packet type, the function allows the matched security packets to get sufficient signaling to reduce protocol flapping when the packets are attacked.
- After it configures bandwidth for user trust packets, the function allows the trust packets to get sufficient signaling for forwarding, and limits the excessive bandwidth occupied by the packets.

1.6.2 Restrictions and Guidelines

- When the hardware capacity is insufficient, the whitelist entries configured by the Command Line Interface (CLI) or automatically delivered by services fail to be installed. A syslog is printed through the FP component on the device to indicate an entry capacity problem. Finally, the packet whitelist becomes invalid.
- You can enable the automatic protection function of protocol packets either in global configuration mode or for a packet type. If you do not configure the automatic protection function for a packet type, the global configuration of the automatic protection function prevails. If you configure the automatic protection function for a packet type, the configuration for this packet type prevails and the global configuration is not used.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the automatic protection function for protocol packets.

cpu-protect auto-defend [**type** *packet-type*] **enable**

When protocol flapping occurs frequently for a packet type, you can attempt to enable the automatic protection function for this packet type to prevent these packets from being affected by other protocol packets.

The automatic protection function for the dynamic routing protocol packets is disabled by default.

(4) Configure a trust packet type.

cpu-protect auto-defend trusted-host type *packet-type* { **src-ip** *ipv4-addres* | **src-ip6** *ipv6-address* | **src-mac** *mac-address* }

Each service is configured with a default trust packet type by default.

(5) Configure a bandwidth for packets under automatic protection.

cpu-protect auto-defend type *packet-type* **bandwidth** *value*

The bandwidth for trust packets under automatic protection is **80%** of the global bandwidth by default.

1.7 Configuring Automatic Protection Function for Ports

Note

Only S6980-64QC and S6580-24DC8QC support configuration of automatic protection for ports.

1.7.1 Overview

If a packet attack occurs on a port, the attack packets sent to the CPU through this port occupy bandwidth. As a result, other protocol packets cannot be sent to the CPU for processing and thus services are interrupted.

The automatic protection function can defend against packet attacks to CPU ports. The implementation effect is as follows:

- After the automatic protection function for a port is enabled, a device can detect packet attacks on this port.
- After it configures an automatic protection attack threshold for a port that transmits protocol packets, this function allows the packets to have sufficient signaling for forwarding and limits the excessive bandwidth occupied by the packets.
- After it configures a blacklist to impose a rate limiting threshold for the ports in the blacklist, this function allows service packets to get sufficient bandwidth on other ports.
- After it configures a monitoring period for a port, this function limits the rates of attack packets on this port within the period. After the period, this function limits the rates of attack packets if the protocol packets received on this port remain in excess of the attack threshold. Otherwise, the limit is removed.
- After the attack alarming function is enabled on a port, syslogs are generated and Simple Network Management Protocol (SNMP) Trap messages are sent when attacks occur or are cleared. The alarm message is triggered when they occur or are cleared. The alarm is not given repeatedly during a monitoring period.

1.7.2 Restrictions and Guidelines

- When the default hardware capacity is insufficient, the blacklist entries fail to be installed. A syslog is printed through the FP component on the device to indicate an entry capacity problem. Finally, the rate limit by the hardware may become ineffective and must be reconfigured by software.
- To detect a specific network environment, you can enable the packet statistics function on a port and check attacks based on the packet statistics. If any, enable the automatic protection function on the port and configure an attack threshold and a rate limiting threshold based on the environmental configuration.
- The blacklist can take effect on a port only when the automatic protection function is enabled on this port.
- You can enable the automatic protection function in global configuration mode or for a port. If you do not configure the automatic protection function for a port, the global configuration of the automatic protection function prevails. If you configure the automatic protection function for a port, the configuration for this port prevails and the global configuration is not used.
- You can configure the packet penalty threshold for an attacked port in global configuration mode or for a specific packet type. If you do not configure the penalty threshold for a specific packet type, you impose a penalty threshold on packets in the global configuration mode when an attack is detected on a port. If you configure a penalty threshold for a specific packet type, the configuration for this packet type prevails.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the automatic protection function for a port.

cpu-protect auto-port-defend [**interface** *interface-type interface-number* | **interface range** *interface-type start-number end-number*] **enable**

The automatic protection function for a port is disabled by default.

(4) Configure an attack threshold for a packet type under automatic protection on a port.

cpu-protect auto-port-defend type *packet-type* **attack-threshold** *threshold-value*

No attack threshold is configured for a packet type under automatic protection on a port by default.

(5) Configure a rate limiting threshold for all packets on an attacked port.

cpu-protect auto-port-defend limit threshold *threshold-value*

In the monitoring period, a rate limiting threshold will be applied to a port if an attack is detected on the port.

The rate limiting threshold for all packets on an attacked port is **5%** of the bandwidth for this interface by default.

(6) Configure a rate limiting threshold for specific packets on an attacked port.

cpu-protect auto-port-defend limit type *packet-type* **threshold** *threshold-value*

In the monitoring period, a rate limiting threshold is applied to the specific packets on a port if an attack is detected on the port.

The rate limiting threshold for specific packets on an attacked port is **5%** of the bandwidth for the port packets by default.

(7) Configure a blacklist for a port.

cpu-protect auto-port-defend blacklist interface *interface-type interface-number*

If a port that normally sends and receives packets is not identified as an attack interface, but affects traffic of services, you can add this port to a blacklist to reduce the traffic of this port.

A port is a trust port and is not in the blacklist by default.

(8) Configure a monitoring period for a port.

cpu-protect auto-port-defend monitor-period *value*

The default attack monitoring period of a port is **300** seconds.

(9) Enable the attack alarm function for a port.

cpu-protect auto-port-defend alarm enable

The attack alarm function for a port is disabled by default. After an attack occurs or is cleared, syslogs are not output and SNMP Trap messages are not sent to prompt users.

1.8 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

- Running the **cleanup** command may lose vital information and thus interrupt services.
- The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 Monitoring

Command	Purpose
cleanup cpu-protect counters [**device** *device-number*] [**slot** *slot-number*]	Clears CPP statistics.
cleanup cpu-protect counters mboard	Clears the CPP statistics on the master device.
cleanup cpu-protect statistics [**interface** *interface-type interface-number*]	Clears the CPP statistics on an interface.
display cpu-protect type *packet-type* [**device** *device-number*] [**slot** *slot-number*]	Displays the configuration and statistics of a packet type.
display cpu-protect traffic-class *traffic-class-number* [**device** *device-number*] [**slot** *slot-number*]	Displays the configuration and statistics of a priority queue.
display cpu-protect cpu	Displays the configuration on the CPU port.
display cpu-protect { **mboard** | **summary** }	Displays all the configuration and statistics on the master device.
display cpu-protect [**device** *device-number*] [**slot** *slot-number*]	Displays all the configuration and statistics of CPP.
display cpu-protect statistics [**interface** *interface-type interface-number*]	Displays the statistics on a CPP port.
display cpu-protect statistics type *packet-type*	Displays the statistics of a packet type on a CPP port.
display cpu-protect auto-defend	Displays automatic protection configuration and whitelist packet statistics.
display cpu-protect auto-defend summary	Displays the automatic protection configuration.
display cpu-protect auto-defend config	Displays the trust packet configuration.
display cpu-protect auto-port-defend	Displays the blacklist configuration on a port.
display cpu-protect auto-port-defend statistics	Displays the blacklist statistics on a port.
display cpu-protect auto-port-defend summary	Displays the automatic protection configuration on a port.
display cpu-protect auto-port-defend attacked summary	Displays the packet attack information on a port.
display cpu-protect linkage summary	Displays the protocol correlation state.

Note

The above monitoring command is compatible with chassis devices and cassette devices, as well as the standalone mode and VSU mode.

If the **device** and **slot** keywords are not entered, the **cleanup** command is used to clear the statistics of all nodes and interfaces in the system and the **display** command is used to display the configuration on the master device.

In the stand-alone mode, the **device** keyword is invisible. For chassis devices, the **slot** keyword is used to specify a line card; for cassette devices, the **slot** keyword is invisible.

In the VSU mode, the **device** keyword specifies a chassis or cassette device. If the **device** keyword is not entered, it specifies the master chassis or the master device. For chassis devices, if the **device** keyword is entered, the **slot** keyword must also be entered to mark a line card in a chassis; for cassette devices, the **slot** keyword is invisible.

Note

Only S6980-64QC and S6580-24DC8QC can display automatic protection-related configurations, allowlist packet statistics, automatic protection configuration, trusted packet configuration of automatic protection, blocklist configuration and statistics of automatic protection for ports, automatic protection configuration for ports, and protocol attack information of automatic protection for ports.

1.9 Configuration Examples

1.9.1 Configuring Basic Functions of CPP

1. Requirements

Address Resolution Protocol (ARP), IP, Open Shortest Path First (OSPF), dot1x, Virtual Router Redundancy Protocol (VRRP), Telnet and ICMP streams are available in the system. In the current configuration, ARP and dot1x streams are in priority queue 2; IP, ICMP and Telnet streams are in priority queue 4; OSPF streams are in priority queue 3; and VRRP streams are in priority queue 6. The bandwidth for each packet type is 10,000 pps; the bandwidth for each priority queue is 20,000 pps; and the bandwidth for the CPU port is 100,000 pps. If the device suffers ARP packet attacks, you must enable the CPP function to prevent the packet attacks and network flapping and ensure the normal running of other services on the device.

2. Notes

- Classify ARP attack packets into priority queue 1 and limit the bandwidth for ARP packets or the priority queue.
- Classify OSPF packets into priority queue 5.
- Classify IP attack packets (without the ARP entries) into priority queue 3 and limit the bandwidth for IP packets or the priority queue.

3. Procedure

Configure the basic attributes of CPP.

```
DeviceA> privilege
```

```
DeviceA# config-system terminal
DeviceA(cfg)# cpu-protect type arp traffic-class 1
DeviceA(cfg)# cpu-protect type arp bandwidth 5000
DeviceA(cfg)# cpu-protect type ospf traffic-class 5
DeviceA(cfg)# cpu-protect type v4uc-route traffic-class 3
DeviceA(cfg)# cpu-protect traffic-class 3 bandwidth 5000
DeviceA(cfg)# end
```

4. Verification

Run the **display cpu-protect** command to view the configuration and statistics.

```
DeviceA# display cpu-protect
%cpu port bandwidth: 100000(pps)
Traffic-class     Bandwidth(pps)    Rate(pps)    Drop(pps)
-------------     --------------    ---------    ---------
 0                20000             0            0
 1                20000             0            0
 2                20000             0            0
 3                5000              0            0
 4                20000             0            0
 5                20000             0            0
 6                20000             0            0
 7                20000             0            0
Packet Type            Traffic-class    Bandwidth(pps)    Rate(pps)    Drop(pps)    Total          Total Drop
-------------------    -------------    --------------    ---------    ---------    ---------      ----------
bpdu                    6                128               0            0            0              0
arp                     1                5000              0            0            0              0
tpp                     6                128               0            0            0              0
dot1x                   2                1500              0            0            0              0
gvrp                    5                128               0            0            0              0
rldp                    5                128               0            0            0              0
lacp                    5                256               0            0            0              0
rerp                    5                128               0            0            0              0
reup                    5                128               0            0            0              0
lldp                    5                768               0            0            0              0
cdp                     5                768               0            0            0              0
dhcps                   2                1500              0            0            0              0
dhcps6                  2                1500              0            0            0              0
dhcp6-client            2                1500              0            0            0              0
dhcp6-server            2                1500              0            0            0              0
dhcp-relay-c            2                1500              0            0            0              0
dhcp-relay-s            2                1500              0            0            0              0
option82                2                1500              0            0            0              0
tunnel-bpdu             2                128               0            0            0              0
tunnel-gvrp             2                128               0            0            0              0
unknown-v6mc            1                128               0            0            0              0
xgv6-ipmc               1                128               0            0            0              0
```

stargv6-ipmc	1	128	0	0	0	0
unknown-v4mc	1	128	0	0	0	0
xgv-ipmc	2	128	0	0	0	0
stargv-ipmc	2	128	0	0	0	0
udp-helper	1	128	0	0	0	0
dvmrp	4	128	0	0	0	0
igmp	2	1000	0	0	0	0
icmp	3	1600	0	0	0	0
ospf	5	2000	0	0	0	0
ospf3	5	2000	0	0	0	0
pim	4	1000	0	0	0	0
pimv6	4	1000	0	0	0	0
rip	4	128	0	0	0	0
ripng	4	128	0	0	0	0
vrrp	6	256	0	0	0	0
vrrpv6	6	256	0	0	0	0
ttl0	0	128	0	0	0	0
ttl1	0	2000	0	0	0	0
hop-limit	0	800	0	0	0	0
local-ipv4	3	4000	0	0	0	0
local-ipv6	3	4000	0	0	0	0
v4uc-route	3	800	0	0	0	0
v6uc-route	1	800	0	0	0	0
rt-host	4	3000	0	0	0	0
mld	2	1000	0	0	0	0
nd-snp-ns-na	1	3000	0	0	0	0
nd-snp-rs	1	1000	0	0	0	0
nd-snp-ra-redirect	1	1000	0	0	0	0
erps	5	128	0	0	0	0
mpls-ttl0	4	128	0	0	0	0
mpls-ttl1	4	128	0	0	0	0
mpls-ctrl	4	128	0	0	0	0
isis	4	2000	0	0	0	0
bgp	4	2000	0	0	0	0
cfm	5	512	0	0	0	0
web-auth	2	2000	0	0	0	0
fcoe-fip	4	1000	0	0	0	0
fcoe-local	4	1000	0	0	0	0
bfd	6	5120	0	0	0	0
micro-bfd	6	5120	0	0	0	0
micro-bfd-v6	6	5120	0	0	0	0
dldp	6	3200	0	0	0	0
other	0	4096	0	0	0	0
trill	4	1000	0	0	0	0
efm	5	1000	0	0	0	0
ipv6-all	0	2000	0	0	0	0

ip-option	0	800	0	0	0	0
mgmt	-	4000	4	0	4639	0
dns	2	200	0	0	0	0
sdn	0	5000	0	0	0	0
sdn_of_fetch	0	5000	0	0	0	0
sdn_of_copy	0	5000	0	0	0	0
sdn_of_trap	0	5000	0	0	0	0
vxlan-non-uc	1	512	0	0	0	0
local-telnet	3	1000	0	0	0	0
local-snmp	3	1000	0	0	0	0
local-ssh	3	1000	0	0	0	0

Contents

1 Configuring NFPP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Concepts 1
1.1.3 Rate Limiting Threshold and Attack Identification for Hosts 3
1.1.4 Rate Limiting and Attack Identification for Ports 4
1.1.5 Hardware Isolation 4
1.1.6 Monitoring Time 4
1.1.7 Trust Host 4
1.2 Configuration Task Summary 4
1.3 Configuring ARP Guard 5
1.3.1 Overview 5
1.3.2 Restrictions and Guidelines 5
1.3.3 Procedure (Global Configuration) 6
1.3.4 Procedure (Interface Configuration) 8
1.4 Configuring IP Guard 9
1.4.1 Overview 9
1.4.2 Restrictions and Guidelines 9
1.4.3 Procedure (Global Configuration) 10
1.4.4 Procedure (Interface Configuration) 11
1.5 Configuring TCP-SYN Guard 12
1.5.1 Overview 12

1.5.2 Restrictions and Guidelines 12
1.5.3 Procedure (Global Configuration) 13
1.5.4 Procedure (Interface Configuration) 14
1.6 Configuring ICMP Guard 15
1.6.1 Overview 15
1.6.2 Restrictions and Guidelines 15
1.6.3 Procedure (Global Configuration) 16
1.6.4 Procedure (Interface Configuration) 17
1.7 Configuring DHCP Guard 17
1.7.1 Overview 17
1.7.2 Restrictions and Guidelines 18
1.7.3 Procedure (Global Configuration) 18
1.7.4 Procedure (Interface Configuration) 19
1.8 Configuring DHCPv6 Guard 20
1.8.1 Overview 20
1.8.2 Restrictions and Guidelines 20
1.8.3 Procedure (Global Configuration) 21
1.8.4 Procedure (Interface Configuration) 21
1.9 Configuring ND Guard 22
1.9.1 Overview 22
1.9.2 Restrictions and Guidelines 22
1.9.3 Procedure (Global Configuration) 23
1.9.4 Procedure (Interface Configuration) 23
1.10 Customizing Guard 24

1.10.1 Overview 24
1.10.2 Restrictions and Guidelines 24
1.10.3 Procedure (Global Configuration) 26
1.10.4 Procedure (Interface Configuration) 27
1.11 Disabling Guard 28
1.11.1 Overview 28
1.11.2 Restrictions and Guidelines 28
1.11.3 Procedure 28
1.12 Configuring NFPP Logs 28
1.12.1 Overview 28
1.12.2 Restrictions and Guidelines 29
1.12.3 Procedure 29
1.13 Monitoring 29
1.14 Configuration Examples 31
1.14.1 Configuring CPU Guard 31

1 Configuring NFPP

1.1 Introduction

1.1.1 Overview

Many malicious attacks exist in the network environment. If no protection policies are available, attack packets occupy the CPU resources of a device. As a result, the CPU cannot process normal protocol packets and control packets and the device does not normally run. The network foundation protection policy (NFPP) protects devices from attacks and keeps a low CPU load for the devices under attack to ensure normal operation of services and stable operation of the entire network.

1.1.2 Basic Concepts

1. ARP Guard

In a LAN, the Address Resolution Protocol (ARP) can parse an IP address into a MAC address to ensure smooth communication. The Denial of Service (DoS) is a process of sending many invalid ARP packets to a gateway in a network to cause the gateway not to provide services for normal hosts. To avoid this type of attacks, configure a rate limiting threshold for the ARP packets and isolate the source of the attack.

2. IP Guard

Hacking attack and network virus intrusion start at a host that scans network activities. Many scanned packets occupy network bandwidth and thus affect normal communication of the network. To avoid this type of attacks, we configure the IP guard function for a device. This function also reduces the CPU usage of the device. IP attacks include the following two types:

- Scan changes in the destination IP address. This type of attacks consume network bandwidth and increase device load. This indicates most of upcoming hacking attacks.
- Send IP packets to an unconfigured destination IP address. This attack is designed for the CPU of a device. On an L3 device, if a destination IP address is configured, packets are forwarded by the chip without consuming CPU resources. Otherwise, the packets are sent to the CPU and the latter sends ARP requests to request the MAC address of the destination IP address. If too many packets sent to the CPU, this consumes many CPU resources. To avoid this type of attacks, we configure a rate limiting threshold for the IP packets and isolate the source of the attack.

3. ARP Source Suppression

When IP packets are sent to an unconfigured destination IP address at a high rate, they are forwarded to the CPU, triggering the device to send ARP requests to request the MAC address of the destination IP address. If too many packets sent to the CPU, this consumes many CPU resources. To avoid this type of attacks, isolate the source of the attack and enable the function of ARP source suppression. If IP packets that trigger ARP requests exceed the threshold in the specified period, the IP packets out of the threshold are discarded.

4. TCP SYN Guard

TCP-SYN-Flood attack is also called half-open connection attack. An attacker sends SYN packets to the target device without responding to the SYN+ACK packets returned by the target device. If it does not receive the SYN+ACK responses, the target device waits and thus a half-open connection is formed. The attacker allows the target device to generate many half-open connections with this method, forcing the target device to waste resources on the half-open connections.

The available TCP connections of a server are limited. If a malicious attacker sends this type of connection requests consecutively in an attack, the available TCP connections of this server become congested. Thus, the available system resources and available bandwidth drop dramatically, affecting normal network services. Consequently, a DoS occurs.

To avoid the TCP-SYN-Flood attack, configure the TCP SYN guard function on the device to limit the sending rates of TCP SYN packets.

5. ICMP Guard

The Internet Control Message Protocol (ICMP) is a common means of network fault diagnosis. According to the principle of ICMP, when a host sends an ICMP Echo Request packet, the receiver returns an ICMP Echo Reply upon receipt of the packet. This process requires CPU coordination. If an attacker sends many ICMP requests to the target device, many CPU resources are consumed, which may interrupt normal device operation in extreme conditions. This is called an ICMP flood attack. To avoid it, configure a rate limiting threshold for the ICMP packets and isolate the source of the attack.

6. DHCP Guard

The Dynamic Host Configuration Protocol (DHCP) is widely used in a LAN to dynamically assign IP addresses. A DHCP starvation attack is the most popular DHCP attack that broadcasts DHCP requests by using a false MAC address. When a network attacker sends enough DHCP requests, it can exhaust the address space of the DHCP server within a period. In this case, the legal host fails to request IP addresses from the DHCP server and thus cannot access a network. To avoid this type of attacks, configure a rate limiting threshold for the DHCP packets and isolate the detected source of the attack.

7. DHCPv6 Guard

DHCPv6 is widely used in a LAN to dynamically assign IPv6 addresses. Like DHCPv4, DHCPv6 has security problems. Therefore, the attack methods of DHCPv4 are applicable to DHCPv6. When a network attacker sends enough DHCPv6 requests, it can exhaust the address space of the DHCPv6 server within a period. In this case, the legal host fails to request IPv6 addresses from the DHCP server and thus cannot access a network. To avoid this type of attacks, configure a rate limiting threshold for DHCPv6 packets and isolate the detected source of the attack.

8. ND Guard

Neighbor discovery (ND) parses addresses, discovers routes and prefixes, and redirects routes in an IPv6 network. A ND process involves five types of packets: Neighbor Solicitation (NS), Neighbor Advertisement (NA), Router Solicitation (RS), Router Advertisement (RA), and Redirect packets. The five types of packets are called ND packets collectively.

ND Snooping listens to ND packets in a network, filters out invalid ND packets, and monitors IPv6 users and binds these users to ports to avoid theft of IPv6 addresses. ND Snooping requires ND packets to be sent to the

CPU. If they are sent at a high rate, this initiates an attack on the CPU. Therefore, we must configure ND guard to limit the rates of ND packets.

9. Customized Guard

There are many types of network protocols, including routing protocols such as OSPF, BGP, and RIP. To exchange packets between devices running different protocols, the packets must be sent to the CPU and processed by the protocol. Each protocol running on a device gives attackers a chance to initiate attacks. If an attacker sends many protocol packets to a network device, this consumes many CPU resources of the device, even affecting device operation.

Network protocols are diverse and grow constantly. Different protocols are demanded in different user environments. Therefore, protocol-specific guard functions have some limitations. To avoid this type of attacks, configure the customized guard function for the device to allow users to flexibly define and configure guard types to meet guard requirements in different user environments.

1.1.3 Rate Limiting Threshold and Attack Identification for Hosts

By configuring a rate limiting threshold for attack packets of hosts, a device can identify different attack scenarios such as host attacks, ARP scanning, and IP scanning. Host attacks can be identified in the following ways:

- Identify hosts according to the source IP address, VLAN ID, and port.
- Identify hosts according to the source MAC address, VLAN ID, and port on the link layer.

Each host is configured with a rate limiting threshold and an attack or attack threshold. The rate limiting threshold must be smaller than the attack threshold.

- If a single host sends attack packets at a rate higher than the rate limiting threshold, packets out of the threshold are discarded.
- If a single host sends attack packets at a rate higher than the attack threshold, the host behavior is identified as an attack and recorded into the log. Meanwhile, a Trap message is sent.
- If the received ARP packets exceed the scanning threshold in the configured period, the source MAC address in the packets is unchanged and the source IP address is changing on the link layer, or the source MAC address and source IP address on the link layer are unchanged but the destination IP address is changing, an ARP scanning attack is suspected.
- If the received IP packets exceed the scanning threshold in the configured period and the source IP address in the packets is unchanged and the destination IP address is changing, an IP scanning attack is suspected.

Note

- When NFPP detects attacks of a specific packet type, NFPP can send an alarm message to users. To avoid frequent alarm messages, NFPP does not resend an alarm within 60 seconds running after the message is generated.
- To prevent frequent log printing from consuming CPU resources, NFPP writes attacks related log content into the buffer and prints the log content from the buffer at the specified rate. NFPP does not limit the rates of Trap messages.
- Only ARP guard and IP scanning guard support the scanning prevention function.

1.1.4 Rate Limiting and Attack Identification for Ports

Each port is configured with a rate limiting threshold and an attack threshold. The rate limiting threshold must be smaller than the attack threshold.

- If a single port sends packets at a rate higher than the rate limiting threshold, packets out of the threshold are discarded.
- If a single port sends packets at a rate higher than the attack threshold, the port behavior is recorded into the log. Meanwhile, a Trap message is sent.

1.1.5 Hardware Isolation

Hardware is isolated by the guard policy after an attack is detected. Isolation requires a filter of hardware. Attack packets from an isolated device are not sent to the CPU for processing to ensure normal operation of the device.

Hardware isolation falls into host based isolation and port-based isolation. The port-based isolation configures a rate limit for physical ports. Only ARP guard and ND guard support hardware isolation based on physical ports.

To isolate attackers, you must configure policies for hardware entries. Because hardware resources are limited, a log is printed to remind users when such resources are exhausted.

Note

Only ARP guard can be configured with the global isolation and forwarding function and the rate limit function for port-based isolation and forwarding.

1.1.6 Monitoring Time

Monitoring hosts can detect attacks in the system. If the isolation time is 0 (indicating no isolation), the guard module automatically monitors attackers through software based on the configured monitoring time. If the isolation time is not 0, the guard module isolates attackers through hardware.

During software monitoring, when the isolation time is configured as a non-zero value, the guard module automatically isolates the attackers under software monitoring through hardware and configures the timeout time of the software monitoring as the isolation time. The monitoring time takes effect only when the isolation time is 0.

1.1.7 Trust Host

To stop monitoring a host is to trust the host. You can configure the host as a trusted host to allow it to send packets to the CPU.

1.2 Configuration Task Summary

NFPP configuration includes the following tasks:

(1) Configuring ARP Guard

(2) Configuring IP Guard

(3) Configuring TCP-SYN Guard

(4) Configuring ICMP Guard

(5) Configuring DHCP Guard

(6) Configuring DHCPv6 Guard

(7) Configuring ND Guard

(8) Customizing Guard

(9) Disabling Guard

(10) Configuring NFPP Logs

1.3 Configuring ARP Guard

1.3.1 Overview

ARP guard is used to detect, process, and prevent ARP based packet attacks. It includes the following functions:

- Identify ARP attacks based on hosts or physical ports. This function can determine an attack host based on the source IP address/MAC address (link layer), VLAN ID, and physical port. A rate limiting threshold and an attack threshold can be configured for each attack identification method. When the ARP packets are sent at a rate higher than the rate limiting threshold, the packets out of the threshold are discarded. When the ARP packets are sent at a rate higher than the attack threshold, an alarm is printed and a Trap message is sent to remind users of ARP attacks.
- Identify ARP scanning attacks. For the ARP packets received within 10 seconds, if the source MAC address is unchanged and the source IP address is changing on the link layer, or the former and source IP address on the link layer are unchanged but the destination IP address is changing more than the specified scanning threshold of times, an ARP scanning attack is suspected and an alarm is printed to remind users.
- Suppress the ARP source. In this type of attacks, ARP requests are triggered through IP packets to consume CPU resources. If the number of IP packets with inexistent destination IP address exceeds the threshold in the specified period, the IP packets out of the threshold are discarded.
- Isolate the source of ARP attacks based on hosts or ports.
  - This function configures hardware entries to isolate the identified attack hosts. Thus, attack packets are not sent to the CPU or forwarded.
  - After isolation takes effect, you can enable the isolation forwarding function to limit the isolation range so that only the management plane of the target CPU is isolated and packets on the data plane are still forwarded.
  - If the port-based isolation entries take effect, the default action is to discard all attack packets. After you enable the function of port-based rate limit forwarding, you can allow some of packets to pass through by changing the isolation action.
- Monitor host attacks through software. After this function guard identifies any attacks, it can monitor the attack host through software in the configured monitoring time.

1.3.2 Restrictions and Guidelines

- ARP guard is to solve the ARP DoS attacks on a device. This function can reduce the impact of this type of attacks on the device, but does not cope with ARP spoofing attacks or ARP attacks in a network.
- If this function is configured in global configuration mode and interface configuration mode, the interface

configuration prevails. For example, if the ARP guard function is enabled on an interface and then disabled in global configuration mode, the function remains effective on this interface.

- After the ARP guard function is disabled, the system automatically clears the entries related to the monitored hosts, scanning hosts, and isolated ports.
- The rate limiting threshold configured based on a MAC address has a higher priority than that configured based on an IP address, and the rate limiting threshold configured based on an IP address has a higher priority than that configured based on a port. To optimize the ARP guard effect, users are advised to configure rate limiting thresholds and attack thresholds for hosts based on the following rules:
  Attack threshold based on source MAC address > Rate limiting threshold based on source MAC address > Attack threshold based on IP address > Rate limiting threshold based on IP address
- The configured attack threshold must be greater than the rate limiting threshold. Otherwise, the configuration fails.
- An ARP scanning table can store only 256 latest records. When the ARP scanning table is full, new records overwrite old ones. ARP scanning attacks may be misjudged. Therefore, when ARP guard does not isolate a detected host that is suspected to initiate an ARP scanning attack, this function only provides some information for the reference of users.
- ARP source suppression conflicts with the isolation function of IP guard scan. If the latter is enabled, you cannot enable the function of ARP source suppression.
- When host attack packets are sent at a rate higher than the rate limiting threshold of CPP, you can configure isolation time to discard the packets and prevent them from occupying bandwidth. When you configure isolation time, you must note that:
  - If you change the isolation time to a non-zero value from 0, attack hosts under software monitoring are automatically isolated by hardware, and the configured monitoring time becomes invalid.
  - If you change the isolation time to 0 from a non-zero value, entries of the isolated attacker are directly deleted without monitoring the attacker through software.
- After you enable the isolation function, the recorded attack host occupies the security module entry.

  You are advised to configure an isolation interval of at least 180 seconds. When both port-based rate limiting and isolation are enabled, isolation will be triggered only when port-based rate limiting and isolation intervals expire.
- When you configure the maximum number of monitored hosts, you must note that:
  - As monitored hosts increase, processing them occupies more CPU resources.
  - If the configured maximum number of hosts is smaller than that of the actually monitored hosts, the existing monitored hosts are not automatically deleted. An alarm is printed to remind users of configuration failure. In this case, users must manually clear some monitored hosts.
  - When the monitored hosts reach the threshold, a log is printed to remind users. In this case, you must manually clear some monitored hosts.

1.3.3 Procedure (Global Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the NFPP configuration mode.

nfpp

(4) Enable the function of global ARP guard.

arp-guard enable

The function of global ARP guard is enabled by default.

(5) (Optional) Configure the global isolation time of ARP guard.

arp-guard isolate-period [*interval* | **permanent**]

The default global isolation time of ARP guard is **0**, and the isolation function is disabled by default.

(6) (Optional) Enable the function of global isolation forwarding of ARP guard.

arp-guard isolate-forwarding enable

The function of global isolation forwarding of ARP guard is enabled by default.

Note

This command is not supported by S6580-24DC8QC.

(7) (Optional) Enable the function of port-based rate limit forwarding of ARP guard.

arp-guard speedlimit-forwarding enable

The function of port-based rate limit forwarding of ARP guard is disabled by default.

Note

This command is not supported by S6580-24DC8QC.

(8) Configure the monitoring time of ARP guard.

arp-guard monitor-period *interval*

The default monitoring time of ARP guard is **600** seconds.

(9) Configure the maximum number of monitored hosts of ARP guard in global configuration mode.

arp-guard monitored-host-limit *limit-number*

The maximum number of monitored hosts of ARP guard in global configuration mode is **20000** by default.

(10) Configure a global rate limiting threshold of ARP guard.

arp-guard speed-limit { **per-port** *speed-limit* | **per-src-ip** *speed-limit* | **per-src-mac** *speed-limit* }

The default global rate limiting thresholds of ARP guard for each interface, each source IP address, and each source MAC address are 1,000 pps, 30 pps, and 30 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

(11) Configure a global attack threshold of ARP guard.

arp-guard attack-threshold { **per-port** *attack-threshold* | **per-src-ip** *attack-threshold* | **per-src-mac** *attack-threshold* }

The default global attack thresholds of ARP guard for each interface, each source IP address, and each source MAC address are 2,500 pps, 1,500 pps, and 1,500 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

(12) Configure a global scanning threshold of ARP guard.

arp-guard scan-threshold *scan-threshold*

The default global scanning threshold of ARP guard is 100 packets per 10 seconds.

(13) Enable the function of ARP source suppression.

arp-guard suppress-mode enable

The function of ARP source suppression is disabled by default.

(14) Configure a threshold of ARP source suppression.

arp-guard suppress-threshold *suppress-threshold*

The default threshold of ARP source suppression is **5** pps.

(15) Configure the period of ARP source suppression.

arp-guard suppress-period *interval*

No period of ARP source suppression is configured by default.

1.3.4 Procedure (Interface Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the configuration mode of the L2 Ethernet interface.

  interface *ethernet-type interface-number*

- Enter the configuration mode of the L3 Ethernet interface.

  interface *ethernet-type interface-number*

(4) Enable the ARP guard function on an interface.

nfpp arp-guard enable

The ARP guard function is disabled on an interface by default.

(5) Configure the isolation time of ARP guard on an interface.

nfpp arp-guard isolate-period [*interval* | **permanent**]

No isolation time of ARP guard is configured on an interface by default. The global isolation time of ARP guard is used.

(6) Configure a rate limiting threshold and an attack threshold of ARP guard on an interface.

nfpp arp-guard policy { **per-port** *speed-limit attack-threshold* | **per-src-ip** *speed-limit attack-threshold* | **per-src-mac** *speed-limit attack-threshold* }

No local rate limiting threshold or local attack threshold of ARP guard is configured on an interface by default. The global rate limiting threshold and global attack threshold of ARP guard are used.

(7) Configure a local scanning threshold of ARP guard on an interface.

nfpp arp-guard scan-threshold *scan-threshold*

No local scanning threshold of ARP guard is configured on an interface by default. The global scanning threshold of ARP guard is used.

1.4 Configuring IP Guard

1.4.1 Overview

After the IP guard function is enabled, a device can detect several common types of IP attacks timely, and ensure its stable operation by rate limit and host isolation. IP guard includes the following functions:

- Identify IP attacks based on hosts or physical ports. This function can determine an attack host based on the source IP address, VLAN ID, and physical port. A rate limiting threshold and an attack threshold can be configured for each attack identification method. When the IP packets are sent at a rate higher than the rate limiting threshold, the packets out of the threshold are discarded. When the IP packets are sent at a rate higher than the attack threshold, an alarm is printed and a Trap message is sent to remind users of IP attacks.
- Identify IP scanning attacks. For the IP packets received within 10 seconds, if the source IP address is unchanged and the destination IP address (not the local address) is changing more than the specified scanning threshold of times, an IP scanning attack is suspected and an alarm is printed to remind users.
- Isolate IP attackers based on hosts. This function configures hardware entries to isolate identified attack hosts. Thus, attack packets are not sent to the CPU or forwarded.
- Monitor attack hosts through software.
  - After this function identifies any attacks, it can monitor the attack host through software in the configured monitoring time.
  - This function can configure a specified host as a trusted host based on its IP address so that the host is free from software monitoring and allowed to send packets to the CPU.

1.4.2 Restrictions and Guidelines

- IP guard is to solve IP attacks whose destination IP address is not a local IP address. If the destination IP address is a local IP address, the rates of IP packets are limited by the function of CPU protect policy (CPP).
- If this function is configured in global configuration mode and interface configuration mode, the interface configuration prevails. For example, if the IP guard function is enabled on an interface and then disabled in global configuration mode, the function is still enabled on this interface.
- After the IP guard function is disabled, the system automatically clears the entries related to the monitored hosts.
- The rate limiting threshold configured based on a source IP address has a higher priority than that configured based on a port.
- The configured attack threshold must be greater than the rate limiting threshold. Otherwise, the configuration fails.
- The isolation function of IP guard conflicts with the function of ARP source suppression. Therefore, you cannot configure the two functions at the same time.

- If you enable them at the same time, you cannot configure the isolation time of IP guard.
- If you enable the function of ARP source suppression and configure the global or local isolation time for IP guard, you cannot enable the IP guard function.

To configure the isolation function of IP guard, make sure to disable the function of ARP source suppression.

- When host attack packets are sent at a rate higher than the rate limiting threshold of CPP, you can configure isolation time to discard the packets and prevent them from occupying bandwidth. When you configure isolation time, you must note that:
  - If you change the isolation time to a non-zero value from 0, attack hosts under software monitoring are automatically isolated by hardware, and the configured monitoring time becomes invalid.
  - If you change the isolation time to 0 from a non-zero value, entries related to the isolated attacker are directly deleted without monitoring the attacker through software.
- After you enable the isolation function, the recorded attack host occupies the security module entry.

  You are advised to configure an isolation interval of at least 180 seconds. When both port-based rate limiting and isolation are enabled, isolation will be triggered only when port-based rate limiting and isolation intervals expire.
- When you configure the maximum number of monitored hosts, you must note that:
  - As monitored hosts increase, processing them occupies more CPU resources.
  - If the configured maximum number of hosts is smaller than that of the actually monitored hosts, the existing monitored hosts are not automatically deleted. An alarm is printed to remind users of configuration failure. In this case, users must manually clear some monitored hosts.
  - When the monitored hosts reach the threshold, a log is printed to remind users. In this case, you must manually clear some monitored hosts.
- IP guard can be configured with trusted hosts. When entries in the table of monitored hosts match the trusted hosts (their IP addresses are the same), the system automatically deletes the entries of the IP addresses. IP guard can be configured with a maximum of 500 trusted hosts.

1.4.3 Procedure (Global Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the NFPP configuration mode.

nfpp

(4) Enable the global IP guard function.

ip-guard enable

The global IP guard function is enabled by default.

(5) (Optional) Configure the global isolation time of IP guard.

ip-guard isolate-period [*interval* | **permanent**]

The default global isolation time of IP guard is **0**, and the isolation function is disabled by default.

If the isolation function is enabled, attack host information occupies a hardware entry of the security module.

(6) Configure the monitoring time of IP guard.

ip-guard monitor-period *interval*

The default monitoring time of IP guard is **600** seconds.

(7) Configure the maximum number of monitored hosts of IP guard.

ip-guard monitored-host-limit *number*

The maximum number of monitored hosts of IP guard is **20000** by default.

(8) (Optional) Configure a global rate limiting threshold of IP guard.

ip-guard speed-limit { **per-port** *speed-limit* | **per-src-ip** *speed-limit* }

The default global rate limiting thresholds of IP guard for each interface and each source IP address are 100 pps and 20 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

(9) Configure a global attack threshold of IP guard.

ip-guard attack-threshold { **per-port** *attack-threshold* | **per-src-ip** *attack-threshold* }

The default global attack thresholds of IP guard for each interface and each source IP address are 400 pps and 200 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

(10) Configure a global scanning threshold of IP guard.

ip-guard scan-threshold *scan-threshold*

The default global scanning threshold of IP guard is 100 packets per 10 seconds.

(11) (Optional) Configure the trusted hosts of IP guard.

ip-guard trusted-host *ipv4-address mask*

No host is configured as a trusted host of IP guard by default.

1.4.4 Procedure (Interface Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the configuration mode of the L2 Ethernet interface.

  interface *ethernet-type interface-number*

- Enter the configuration mode of the L3 Ethernet interface.

  interface *ethernet-type interface-number*

(4) Enable the IP guard function on an interface.

nfpp ip-guard enable

The IP guard function is disabled on an interface by default.

(5) Configure the local isolation time of IP guard on an interface.

nfpp ip-guard isolate-period [*interval* | **permanent**]

No isolation time of IP guard is configured by default.

(6) Configure a local rate limiting threshold and a local attack threshold of IP guard on an interface.

nfpp ip-guard policy { **per-port** *speed-limit attack-threshold* | **per-src-ip** *speed-limit attack-threshold* }

No local rate limiting threshold or local attack threshold of IP guard is configured on an interface by default. The global rate limiting threshold and global attack threshold of IP guard are used.

(7) Configure a scanning threshold of IP guard on an interface.

nfpp ip-guard scan-threshold *scan-threshold*

No local scanning threshold of IP guard is configured on an interface by default. The global scanning threshold of IP guard is used.

1.5 Configuring TCP-SYN Guard

Note

This command is not supported by the S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, or S6510-32CQ.

1.5.1 Overview

An attacker can launch an attack against the device connected through TCP connections. To detect and prevent the TCP-SYN flooding attacks, you can enable the TCP-SYN guard function on the devices. TCP-SYN guard includes the following functions:

- Identify TCP-SYN attacks based on hosts or physical ports. This function can determine an attack host based on the source IP address, VLAN ID, and physical port. A rate limiting threshold and an attack threshold can be configured for each attack identification method. When the TCP-SYN packets are sent at a rate higher than the rate limiting threshold, the packets out of the threshold are discarded. When the TCP-SYN packets are sent at a rate higher than the attack threshold, an alarm is printed and a Trap message is sent.
- Isolate TCP-SYN attackers based on hosts. This function configures hardware entries to isolate identified attack hosts. Thus, attack packets are not sent to the CPU or forwarded.
- Monitor attack hosts through software.
  - After this function identifies any attacks, it can monitor the attack host through software in the configured monitoring time.
  - This function can configure a specified host as a trusted host of TCP-SYN guard based on its IP address so that it is free from software monitoring and allowed to send packets to the CPU.

1.5.2 Restrictions and Guidelines

- If this function is configured in global configuration mode and interface configuration mode, the interface configuration prevails.
- The rate limiting threshold configured based on a source IP address has a higher priority than that configured based on a port.
- The configured attack threshold must be greater than the rate limiting threshold. Otherwise, the configuration fails.

- When host attack packets are sent at a rate higher than the rate limiting threshold of CPP, you can configure isolation time to discard the packets and prevent them from occupying bandwidth. When you configure isolation time, you must note that:
  - If you change the isolation time to a non-zero value from 0, attack hosts under software monitoring are automatically isolated by hardware, and the configured monitoring time becomes invalid.
  - If you change the isolation time to 0 from a non-zero value, entries related to the isolated attacker are directly deleted without monitoring the attacker through software.
- After you enable the isolation function, the recorded attack host occupies the security module entry.

  You are advised to configure an isolation interval of at least 180 seconds. When both port-based rate limiting and isolation are enabled, isolation will be triggered only when port-based rate limiting and isolation intervals expire.
- When you configure the maximum number of monitored hosts, you must note that:
  - As monitored hosts increase, processing them occupies more CPU resources.
  - If the configured maximum number of hosts is smaller than that of the actually monitored hosts, the existing monitored hosts are not automatically deleted. An alarm is printed to remind users of configuration failure. In this case, users must manually clear some monitored hosts.
  - When the monitored hosts reach the threshold, a log is printed to remind users. In this case, you must manually clear some monitored hosts.
- TCP-SYN guard can be configured with trusted hosts. When entries in the table of monitored hosts match the trusted hosts (their IP addresses are the same), the system automatically deletes the entries of the IP addresses. TCP-SYN guard can be configured with a maximum of 500 trusted hosts.

1.5.3 Procedure (Global Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the NFPP configuration mode.

nfpp

(4) Enable the global TCP-SYN guard function.

tcp-syn-guard enable

Global TCP-SYN guard is enabled by default.

(5) (Optional) Configure the global isolation time of TCP-SYN guard.

tcp-syn-guard isolate-period [*interval* | **permanent**]

The default global isolation time of TCP-SYN guard is **0**, and the isolation function is disabled by default.

(6) Configure the monitoring time of TCP-SYN guard.

tcp-syn-guard monitor-period *interval*

The default monitoring time of TCP-SYN guard is **600** seconds.

(7) Configure the maximum number of monitored hosts of TCP-SYN guard.

tcp-syn-guard monitored-host-limit *number*

The maximum number of monitored hosts of TCP-SYN guard is **20000** by default.

(8) Configure a global rate limiting threshold of TCP-SYN guard.

tcp-syn-guard speed-limit { **per-port** *speed-limit* | **per-src-ip** *speed-limit* }

The default global rate limiting thresholds of TCP-SYN guard for each interface and each source IP address are 100 pps and 20 pps respectively on S6980-64QC, and not configured on S6920-4C, S6510-48VS8CQ, or S6250-48XS8CQ.

(9) Configure a global attack threshold of TCP-SYN guard.

tcp-syn-guard attack-threshold { **per-port** *attack-threshold* | **per-src-ip** *attack-threshold* }

The default global attack thresholds of TCP-SYN guard for each interface and each source IP address are 400 pps and 200 pps respectively on S6980-64QC, and not configured on S6920-4C, S6510-48VS8CQ, or S6250-48XS8CQ.

(10) (Optional) Configure the trusted hosts of TCP-SYN guard.

tcp-syn-guard trusted-host *ipv4-address mask*

No host is configured as a trusted host of TCP-SYN guard by default.

1.5.4 Procedure (Interface Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the configuration mode of the L2 Ethernet interface.

  interface *ethernet-type interface-number*

- Enter the configuration mode of the L3 Ethernet interface.

  interface *ethernet-type interface-number*

(4) Enable the TCP-SYN guard function on an interface.

nfpp tcp-syn-guard enable

The TCP-SYN guard function is not configured on an interface by default. The global TCP-SYN guard function is used.

(5) Configure the isolation time of TCP-SYN guard on an interface.

nfpp tcp-syn-guard isolate-period { *interval* | **permanent** }

No local isolation time of TCP-SYN guard is configured by default. The global isolation time of TCP-SYN guard is used.

(6) Configure the local rate limiting threshold and local attack threshold of TCP-SYN guard on an interface.

nfpp tcp-syn-guard policy { **per-port** *speed-limit attack-threshold* | **per-src-ip** *speed-limit attack-threshold* }

No local rate limiting threshold or local attack threshold of TCP-SYN guard is configured on an interface by default. The global rate limiting threshold and global attack threshold of TCP-SYN guard are used.

1.6 Configuring ICMP Guard

1.6.1 Overview

The ICMP guard function configured for a device is used to prevent an attacker from consuming system resources by sending many ICMP requests. ICMP guard includes the following functions:

- Identify ICMP attacks based on hosts or physical ports. This function can determine an attack host based on the source IP address, VLAN ID, and port. A rate limiting threshold and an attack threshold can be configured for each attack identification method. When the ICMP packets are sent at a rate higher than the rate limiting threshold, the packets out of the threshold are discarded. When the ICMP packets are sent at a rate higher than the attack threshold, an alarm is printed and a Trap message is sent.
- Isolate ICMP attackers based on hosts. This function configures hardware entries to isolate identified attack hosts. Thus, attack packets are not sent to the CPU or forwarded.
- Monitor attack hosts through software.
  - After this function identifies any attacks, it can monitor the attack host through software in the configured monitoring time.
  - This function can configure a specified host as a trusted host of ICMP guard based on its IP address so that it is free from software monitoring and allowed to send packets to the CPU.

1.6.2 Restrictions and Guidelines

- If this function is configured in global configuration mode and interface configuration mode, the interface configuration prevails.
- After the ICMP guard function is disabled, the system automatically clears the entries related to the monitored hosts.
- The rate limiting threshold configured based on a source IP address has a higher priority than that configured based on a port.
- The configured attack threshold must be greater than the rate limiting threshold. Otherwise, the configuration fails.
- When host attack packets are sent at a rate higher than the rate limiting threshold of CPP, you can configure isolation time to discard the packets and prevent them from occupying bandwidth. When you configure isolation time, you must note that:
  - If you change the isolation time to a non-zero value from 0, attack hosts under software monitoring are automatically isolated by hardware, and the configured monitoring time becomes invalid.
  - If you change the isolation time to 0 from a non-zero value, entries related to the isolated attacker are directly deleted without monitoring the attacker through software.
- After you enable the isolation function, the recorded attack host occupies the security module entry.

  You are advised to configure an isolation interval of at least 180 seconds. When both port-based rate limiting and isolation are enabled, isolation will be triggered only when port-based rate limiting and isolation intervals expire.
- When you configure the maximum number of monitored hosts, you must note that:
  - As monitored hosts increase, processing them occupies more CPU resources.
  - If the configured maximum number of hosts is smaller than that of the actually monitored hosts, the

existing monitored hosts are not automatically deleted. An alarm is printed to remind users of configuration failure. In this case, users must manually clear some monitored hosts.

- When the monitored hosts reach the threshold, a log is printed to remind users. In this case, you must manually clear some monitored hosts.

- ICMP guard can be configured with trusted hosts. When entries in the table of monitored hosts match the trusted hosts (their IP addresses are the same), the system automatically deletes the entries of the IP addresses. ICMP guard can be configured with a maximum of 500 trusted hosts.

1.6.3 Procedure (Global Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the NFPP configuration mode.

nfpp

(4) Enable the function of global ICMP guard.

icmp-guard enable

The function of global ICMP guard is enabled by default.

(5) (Optional) Configure the global isolation time of ICMP guard.

icmp-guard isolate-period [*interval* | **permanent**]

The default global isolation time of ICMP guard is **0**, and the isolation function is disabled by default.

(6) Configure the monitoring time of ICMP guard.

icmp-guard monitor-period *interval*

The default monitoring time of ICMP guard is **600** seconds.

(7) Configure the maximum number of monitored hosts.

icmp-guard monitored-host-limit *number*

The maximum number of monitored hosts is **20000** by default.

(8) Configure the global rate limiting threshold of ICMP guard.

icmp-guard speed-limit { **per-port** *speed-limit* | **per-src-ip** *speed-limit* }

The default global rate limiting thresholds of ICMP guard for each interface and each source IP address are 1,000 pps and 800 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

(9) Configure the global attack threshold of ICMP guard.

icmp-guard attack-threshold { **per-port** *attack-threshold* | **per-src-ip** *attack-threshold* }

The default global attack thresholds of ICMP guard for each interface and each source IP address are 1,500 pps and 1,200 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

(10) (Optional) Configure the trusted hosts of ICMP guard.

icmp-guard trusted-host *ipv4-address mask*

No host is configured as a trusted host by default.

1.6.4 Procedure (Interface Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the configuration mode of the L2 Ethernet interface.

  interface *ethernet-type interface-number*

- Enter the configuration mode of the L3 Ethernet interface.

  interface *ethernet-type interface-number*

(4) Enable the ICMP guard function on an interface.

nfpp icmp-guard enable

The ICMP guard function is not configured on an interface by default. The global ICMP guard function is used.

(5) Configure the local isolation time of ICMP guard on an interface.

nfpp icmp-guard isolate-period { *interval* | **permanent** }

No local isolation time of ICMP guard is configured on an interface by default. The global isolation time of ICMP guard is used.

(6) Configure the local rate limiting threshold and local attack threshold of ICMP guard on an interface.

nfpp icmp-guard policy { **per-port** *speed-limit attack-threshold* | **per-src-ip** *speed-limit attack-threshold* }

No local rate limiting threshold or local attack threshold of ICMP guard is configured on an interface by default. The global rate limiting threshold and global attack threshold of ICMP guard are used.

1.7 Configuring DHCP Guard

1.7.1 Overview

In a DHCP starvation attack, all IP addresses of a DHCP server are exhausted. As a result, the DHCP server fails to provide the address assignment service, thus affecting the network access of legal users. To address this type of attacks, you must deploy the DHCP guard function on a device.

DHCP guard includes the following functions:

- Identify DHCP attacks based on hosts or physical ports. This function can determine an attack host based on the source MAC address, VLAN ID, and port on the link layer. A rate limiting threshold and an attack threshold can be configured for each attack identification method. When the DHCP packets are sent at a rate higher than the rate limiting threshold, the packets out of the threshold are discarded. When the DHCP packets are sent at a rate higher than the attack threshold, an alarm is printed and a Trap message is sent.
- Isolate DHCP attackers based on hosts. This function configures hardware entries to isolate identified attack hosts. Thus, attack packets are not sent to the CPU or forwarded.
- Monitor attack hosts through software. After an attack is identified, this function monitors the attack host

through software in the configured monitoring time.

1.7.2 Restrictions and Guidelines

- If this function is configured in global configuration mode and interface configuration mode, the interface configuration prevails.
- After the DHCP guard function is disabled, the system automatically clears the entries related to the monitored hosts.
- The configured attack threshold must be greater than the rate limiting threshold. Otherwise, the configuration fails.
- When host attack packets are sent at a rate higher than the rate limiting threshold of CPP, you can configure isolation time to discard the packets and prevent them from occupying bandwidth. When you configure isolation time, you must note that:
  - If you change the isolation time to a non-zero value from 0, attack hosts under software monitoring are automatically isolated by hardware, and the configured monitoring time becomes invalid.
  - If you change the isolation time to 0 from a non-zero value, entries related to the isolated attacker are directly deleted without monitoring the attacker through software.
- After you enable the isolation function, the recorded attack host occupies the security module entry.

  You are advised to configure an isolation interval of at least 180 seconds. When both port-based rate limiting and isolation are enabled, isolation will be triggered only when port-based rate limiting and isolation intervals expire.
- When you configure the maximum number of monitored hosts, you must note that:
  - As monitored hosts increase, processing them occupies more CPU resources.
  - If the configured maximum number of hosts is smaller than that of the actually monitored hosts, the existing monitored hosts are not automatically deleted. An alarm is printed to remind users of configuration failure. In this case, users must manually clear some monitored hosts.
  - When the monitored hosts reach the threshold, a log is printed to remind users. In this case, you must manually clear some monitored hosts.
- If a port is configured as trusted by the DHCP Snooping function, the DHCP guard function becomes invalid to avoid misjudging DHCP traffic on the trusted port. For trusted port configuration in DHCP Snooping, see *Configuring DHCP Snooping*.

1.7.3 Procedure (Global Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the NFPP configuration mode.

nfpp

(4) Enable the function of global DHCP guard.

dhcp-guard enable

The function of global DHCP guard is enabled by default.

(5) (Optional) Configure the global isolation time of DHCP guard.

dhcp-guard isolate-period { *interval* | **permanent** }

The default global isolation time of DHCP guard is **0**, and the isolation function is disabled by default.

(6) Configure the monitoring time of DHCP guard.

dhcp-guard monitor-period *interval*

The default monitoring time of DHCP guard is **600** seconds.

(7) Configure the maximum number of monitored hosts.

dhcp-guard monitored-host-limit *number*

The maximum number of monitored hosts of DHCP guard is **20000** by default.

(8) Configure the global rate limiting threshold of DHCP guard.

dhcp-guard speed-limit { **per-port** *speed-limit* | **per-src-mac** *speed-limit* }

The default global rate limiting thresholds of DHCP guard for each interface and each source MAC address are 1,000 pps and 5 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

(9) Configure the global attack threshold of DHCP guard.

dhcp-guard attack-threshold { **per-port** *attack-threshold* | **per-src-mac** *attack-threshold* }

The default global attack thresholds of DHCP guard for each interface and each source MAC address are 1,200 pps and 10 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

1.7.4 Procedure (Interface Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the configuration mode of the L2 Ethernet interface.

  interface *ethernet-type interface-number*

- Enter the configuration mode of the L3 Ethernet interface.

  interface *ethernet-type interface-number*

(4) Enable the DHCP guard function on an interface.

nfpp dhcp-guard enable

The DHCP guard function is disabled on an interface by default. The global DHCP guard function is enabled.

(5) Configure isolation time of DHCP guard on an interface.

nfpp dhcp-guard isolate-period [*interval* | **permanent**]

No local isolation time of DHCP guard is configured on an interface by default. The global isolation time of DHCP guard is used.

(6) Configure the local rate limiting threshold and local attack threshold of DHCP guard on an interface.

nfpp dhcp-guard policy { **per-port** *speed-limit attack-threshold* | **per-src-mac** *speed-limit attack-threshold* }

No local rate limiting threshold or local attack threshold of DHCP guard is configured on an interface by default. The global rate limiting threshold and global attack threshold of DHCP guard are used.

1.8 Configuring DHCPv6 Guard

1.8.1 Overview

Like DHCPv4, DHCPv6 has security problems. Therefore, the attack methods of DHCPv4 are applicable to DHCPv6. To prevent this type of attacks, we must also deploy the DHCPv6 guard policy. DHCPv6 guard includes the following functions:

- Identify DHCPv6 attacks based on hosts or physical ports. This function can determine an attack host based on the source MAC address, VLAN ID, and port on the link layer. A rate limiting threshold and an attack threshold can be configured for each attack identification method.
- Isolate DHCPv6 attackers based on hosts. This function configures hardware entries to isolate identified attack hosts. Thus, attack packets are not sent to the CPU or forwarded.
- Monitor attack hosts through software. After an attack is identified, this function monitors the attack host through software in the configured monitoring time.

1.8.2 Restrictions and Guidelines

- If this function is configured in global configuration mode and interface configuration mode, the interface configuration prevails.
- After the DHCPv6 guard function is disabled, the system automatically clears the entries related to the monitored hosts.
- The configured attack threshold must be greater than the rate limiting threshold. Otherwise, the configuration fails.
- When host attack packets are sent at a rate higher than the rate limiting threshold of CPP, you can configure isolation time to discard the packets and prevent them from occupying bandwidth. When you configure isolation time, you must note that:
  - If you change the isolation time to a non-zero value from 0, attack hosts under software monitoring are automatically isolated by hardware, and the configured monitoring time becomes invalid.
  - If you change the isolation time to 0 from a non-zero value, entries related to the isolated attacker are directly deleted without monitoring the attacker through software.
- After you enable the isolation function, the recorded attack host occupies the security module entry.

  You are advised to configure an isolation interval of at least 180 seconds. When both port-based rate limiting and isolation are enabled, isolation will be triggered only when port-based rate limiting and isolation intervals expire.
- When you configure the maximum number of monitored hosts, you must note that:
  - As monitored hosts increase, processing them occupies more CPU resources.
  - If the configured maximum number of hosts is smaller than that of the actually monitored hosts, the

existing monitored hosts are not automatically deleted. An alarm is printed to remind users of configuration failure. In this case, users must manually clear some monitored hosts.

- When the monitored hosts reach the threshold, a log is printed to remind users. In this case, you must manually clear some monitored hosts.

- If a port is configured as trusted by the DHCPv6 Snooping function, the DHCPv6 guard function becomes invalid to avoid misjudging DHCPv6 traffic on the trusted port. For trusted port configuration in DHCPv6 Snooping, see *Configuring DHCPv6 Snooping*.

1.8.3 Procedure (Global Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the NFPP configuration mode.

nfpp

(4) Enable the function of global DHCPv6 guard.

dhcpv6-guard enable

The function of global DHCPv6 guard is enabled by default.

(5) Configure the monitoring time of DHCPv6 guard.

dhcpv6-guard monitor-period *interval*

The default monitoring time of DHCPv6 guard is **600** seconds.

(6) Configure the maximum number of monitored hosts.

dhcpv6-guard monitored-host-limit *number*

The maximum number of monitored hosts of DHCPv6 guard is **20000** by default.

(7) Configure the global rate limiting threshold of DHCPv6 guard.

dhcpv6-guard speed-limit { **per-port** *speed-limit* | **per-src-mac** *speed-limit* }

The default global rate limiting thresholds of DHCPv6 guard for each interface and each source MAC address are 1,000 pps and 5 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

(8) Configure the global attack threshold of DHCPv6 guard.

dhcpv6-guard attack-threshold { **per-port** *attack-threshold* | **per-src-mac** *attack-threshold* }

The default global attack thresholds of DHCPv6 guard for each interface and each source MAC address are 1,200 pps and 10 pps respectively on S6920-4C, S6510-48VS8CQ, S6250-48XS8CQ, and S6980-64QC.

1.8.4 Procedure (Interface Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the configuration mode of the L2 Ethernet interface.

  interface *ethernet-type interface-number*

- Enter the configuration mode of the L3 Ethernet interface.

  interface *ethernet-type interface-number*

(4) Enable the DHCPv6 guard function on an interface.

nfpp dhcpv6-guard enable

The DHCPv6 guard function is disabled on an interface by default. The global DHCPv6 guard function is used.

(5) Configure the local rate limiting threshold and local attack threshold of DHCPv6 guard on an interface.

nfpp dhcpv6-guard policy { **per-port** *speed-limit attack-threshold* | **per-src-mac** *speed-limit attack-threshold* }

No local rate limiting threshold is local attack threshold of DHCPv6 guard is configured on an interface by default. The global rate limiting threshold and global attack threshold of DHCPv6 guard are used.

1.9 Configuring ND Guard

1.9.1 Overview

ND packets of ND guard fall into three types by use:

- Type 1: Neighbor requests and neighbor advertisements, which are used to parse addresses.
- Type 2: Route requests, which are used to discover gateways for hosts.
- Type 3: Route advertisements, which are used to advertise gateways and prefixes; and redirection packets, which are used to advertise better next hops.

ND guard identifies ND packets based on physical ports. This function prevents attacks by configuring rate limiting thresholds for attack ports and isolating the ports.

1.9.2 Restrictions and Guidelines

- If this function is configured in global configuration mode and interface configuration mode, the interface configuration prevails.
- The configured attack threshold must be greater than the rate limiting threshold. Otherwise, the configuration fails.
- After you enable the ND Snooping function by running the **ipv6 nd snooping enable** command, all ND packets are NDSNP packets. In this case, you must configure rate limiting thresholds and attack thresholds for the NDSNP packets to validate the guard.
- After you enable the isolation function, the recorded attack host occupies the security module entry. If the port-based isolation entries take effect, the default action is to discard all attack packets. After you enable the function of port-based rate limit forwarding, you can allow some of packets to pass through by changing the isolation action.

You are advised to configure an isolation interval of at least 180 seconds. When both port-based rate limiting and isolation are enabled, isolation will be triggered only when port-based rate limiting and isolation intervals expire.

- Rate limit by hardware is ineffective to NDSNP packets.

1.9.3 Procedure (Global Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the NFPP configuration mode.

nfpp

(4) Enable the global ND guard function.

nd-guard enable

The function of global ND guard is enabled by default.

(5) Enable the function of port-based rate limit forwarding of ND guard.

nd-guard speedlimit-forwarding enable

The function of port-based rate limit forwarding of ND guard is enabled by default.

Note

This command is not supported by S6580-24DC8QC.

(6) Configure the global rate limiting threshold of ND guard.

nd-guard speed-limit per-port { **ndsnp** *speed-limit* | **ns-na** *speed-limit* | **ra-redirect** *speed-limit* | **rs** *speed-limit* }

The default global rate limiting thresholds of ND guard for NDSNP packets, neighbor requests and advertisements, route advertisements and redirection packets, and route requests are not configured, 1,000 pps, 500 pps, and 500 pps respectively on S6920-4C, S6510-48VS8CQ, and S6250-48XS8CQ, 3,000 pps, 1,000 pps, 500 pps, and 500 pps respectively on S6980-64QC.

(7) Configure the global attack threshold of ND guard.

nd-guard attack-threshold per-port { **ndsnp** *attack-threshold* | **ns-na** *attack-threshold* | **ra-redirect** *attack-threshold* | **rs** *attack-threshold* }

The default global attack thresholds of ND guard for NDSNP packets, neighbor requests and advertisements, route advertisements and redirection packets, and route requests are not configured, 2500 pps, 800 pps, and 800 pps respectively on S6920-4C, S6510-48VS8CQ, and S6250-48XS8CQ, 6,000 pps, 2,500 pps, 800 pps, and 800 pps respectively on S6980-64QC.

1.9.4 Procedure (Interface Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the configuration mode of the L2 Ethernet interface.

  interface *ethernet-type interface-number*

- Enter the configuration mode of the L3 Ethernet interface.

  interface *ethernet-type interface-number*

(4) Enable the ND guard function on an interface.

nfpp nd-guard enable

The ND guard function is disabled on an interface by default. The global ND guard function is enabled.

(5) Configure the local rate limiting threshold and local attack threshold of ND guard on an interface.

nfpp nd-guard policy per-port { **ndsnp** *speed-limit attack-threshold* | **ns-na** *speed-limit attack-threshold* | **ra-redirect** *speed-limit attack-threshold* | **rs** *speed-limit attack-threshold* }

No local rate limiting threshold or local attack threshold of ND guard is configured on an interface by default. The global rate limiting threshold and global attack threshold of ND guard are used.

1.10 Customizing Guard

1.10.1 Overview

Users can customize guard types and guard policies to solve network attacks in special application scenarios.

Customized guard includes the following functions:

- Allow users to combine packet type fields and use them as packet types of customized guard to meet requirements of guard scenarios.
- Configure guard policies based on hosts and ports. Host guard policies include those based on the source IP address, VLAN ID and port, or those based on the source MAC address, VLAN ID and port on the link layer. A rate limiting threshold and an attack threshold can be configured for each guard policy. When the packets are sent at a rate higher than the rate limiting threshold, the packets out of the threshold are discarded. When the packets are sent at a rate higher than the attack threshold, an alarm is printed and a Trap message is sent.
- Monitor attack hosts through software.
  - After this function identifies any attacks, it can monitor the attack host through software in the configured monitoring time.
  - This function can configure a specified host as a trusted host based on its IPv4, IPv6, or MAC address so that it is free from software monitoring and allowed to send packets to the CPU without rate limit and processing.

1.10.2 Restrictions and Guidelines

- Customized guard allows users to combine packet type fields to include most existing protocol types and expand new protocol types. Improper configuration of the packet type fields may cause a network exception. Therefore, users must well understand network protocols before they configure this function.

- Customized guard has a higher packet detection priority than basic guard types. When packets match the customized guard policy and the policy of a basic guard type, they preferentially undergo rate limit and attack identification according to the customized guard policy. To ensure the configuration effect of basic guard types, try to refer to the configuration guide when you configure the packet type fields of customized guard.
- Customized guard policies must be configured with different type names and packet types. Matched packet types must differ from the packet types of existing guard types, including those of basic guard types such as ARP guard, IP guard, ICMP guard, TCP-SYN guard, DHCP guard, DHCPv6 guard, and ND guard.
- You can configure packet types of customized guard by combining the following fields. These fields include Ethernet link layer type (etype), source MAC address (src-mac), destination MAC address (dst-mac), IPv4 or IPv6 protocol number (protocol), source IPv4 or IPv6 address (src-ip/src-ipv6), destination IPv4 or IPv6 address (dst-ip/dst-ipv6), source port on the transport layer (src-port), and destination port on the transport layer (dst-port).

  To configure the matched fields of packet fields, conform with the following limit rules. Otherwise, the configuration fails:

  - The IPv4 or IPv6 protocol number (protocol) is configurable only when the Ethernet link layer type (etype) is IPv4 or IPv6.
  - The source IPv4 address (src-ip) and destination IPv4 address (dst-ip) are configurable only when the Ethernet link layer type (etype) is IPv4.
  - The source IPv6 address (src-ipv6) and destination IPv6 address (dst-ipv6) are configurable only when the Ethernet link layer type (etype) is IPv6.
  - The source port on the transport layer (src-port) and destination port on the transport layer (dst-port) are configurable only when the IPv4 or IPv6 protocol number (protocol) is TCP or UDP.
- For the three types of customized guard policies that are based on the source IP address, source MAC address, and port, you must configure at least one type. That is, you must configure the global rate limiting threshold and global attack threshold of customized guard. Otherwise, the customized guard function does not take effect. You can configure a host policy based on the source IP address, VLAN ID, and port only when the specified Ethernet link layer type (etype) is IPv4 or IPv6.
- The rate limiting threshold based on a source MAC address has a higher priority than that based on a source IP address. The configured attack threshold must be greater than the rate limiting threshold. Otherwise, the configuration fails.
- Only after you configure packet types and guard policies can you enable the customized guard function. If you do not complete any configuration, you are reminded that the current customized guard is not completely configured and the configuration fails.
- The host policy of customized guard on an interface must be consistent with the global host policy configuration. For example, if no global rate limiting threshold and global attack threshold are configured based on the source IP address, you cannot configure the local rate limiting threshold and local attack threshold on an interface.
- If this function is configured in global configuration mode and interface configuration mode, the interface configuration prevails.
- After the customized guard function is disabled, the system automatically clears the entries related to the monitored hosts.

- When you configure the maximum number of monitored hosts, you must note that:
  - As monitored hosts increase, processing them occupies more CPU resources.
  - If the configured maximum number of hosts is smaller than that of the actually monitored hosts, the existing monitored hosts are not automatically deleted. An alarm is printed to remind users of configuration failure. In this case, users must manually clear some monitored hosts.
  - When the monitored hosts reach the threshold, a log is printed to remind users. In this case, you must manually clear some monitored hosts.
- When you configure trusted hosts, note that:
  - You must configure matched packet types before you configure trusted hosts.
  - The trusted hosts of customized guard must be consistent with the matched packet types. For example, only when the Ethernet link layer type (etype) is IPv4, can you add IPv4 addresses as trusted IP addresses. Similarly, when the matched packet type is IPv4, IPv6 addresses cannot be added as trusted IP addresses.
  - You can configure all hosts in a network segment as trusted hosts by configuring masks.
  - When entries in the table of monitored hosts match the trusted hosts, the system automatically deletes the entries of the IP addresses.
  - ICMP guard can be configured with a maximum of 500 trusted hosts.

1.10.3 Procedure (Global Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the NFPP configuration mode.

nfpp

(4) Customize a guard type, and enter the configuration mode of the NFPP customized guard type.

define *define-name*

(5) Configure the matched packet types of the customized guard type.

match-rule { **dst-ip** *destination-ipv4-address* [**dst-ip-mask** *mask*] | **dst-ipv6** *destination-ipv6-address* [**dst-ipv6-masklen** *prefix-length*] | **dst-mac** *destination-mac* [**dst-mac-mask** *destination-mac-mask*] | **dst-port** *port-number* | **etype** *type* | **protocol** *protocol* | **src-ip** *source-ipv4-address* [**src-ip-mask** *mask*] | **src-ipv6** *source-ipv6-address* [**src-ipv6-masklen** *prefix-length*] | **src-mac** *source-mac-address* | **src-port** *port-number* } *

(6) Configure the global rate limiting threshold and global attack threshold of the customized guard type.

global-policy { **per-port** *speed-limit attack-threshold* | **per-src-ip** *speed-limit attack-threshold* | **per-src-mac** *speed-limit attack-threshold* }

No global rate limiting threshold or global attack threshold of the customized guard type is configured by default.

They must be configured. Otherwise, you cannot enable the customized guard function.

(7) Configure the monitoring time of the customized guard type.

monitor-**period** *interval*

The default monitoring time of a customized guard type is **600** seconds.

(8) Configure the maximum number of monitored hosts of the customized guard type.

monitored-**host**-**limit** *number*

The maximum number of monitored hosts is **20000** by default.

(9) Configure the trusted hosts of the customized guard type.

trusted-host { *ipv4-address mask* | *ipv6-address*/*prefix-length* | *mac-address mask* }

No host is configured as a trusted host of the customized guard type by default.

(10) Enable the function of global customized guard.

define *define-name* **enable**

The function of global customized guard is disabled by default.

1.10.4 Procedure (Interface Configuration)

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the configuration mode of the L2 Ethernet interface.

  interface *ethernet-type interface-number*

- Enter the configuration mode of the L3 Ethernet interface.

  interface *ethernet-type interface-number*

(4) Enable the customized guard function on an interface.

nfpp define *define-name* **enable**

The customized guard function is disabled on an interface by default. The global customized guard function is used.

(5) Configure the local rate limiting threshold and local attack threshold of the customized guard type on an interface.

nfpp define *define-name* **policy** { **per-port** *speed-limit attack-threshold* | **per-src-ip** *speed-limit attack-threshold* | **per-src-mac** *speed-limit attack-threshold* }

No local rate limiting threshold or local attack threshold of the customized guard type is configured on an interface by default. The global rate limiting threshold and global attack threshold of the customized guard type are used.

1.11 Disabling Guard

1.11.1 Overview

Users can disable or enable all basic types of global guard function by running one command. Basic guard types supported by this function include ARP guard, IP guard, ICMP guard, TCP-SYN guard, DHCP guard, DHCPv6 guard, and ND guard.

1.11.2 Restrictions and Guidelines

- This function applies to only basic guard types and does not affect the customized guard types.
- This function applies to the guard functions in global configuration mode and does not affect the guard functions in interface configuration mode.
- After you run the one-click enabling/disabling command, the configuration of this command cannot be displayed by running the **display current-config** command.
- The one-click enabling/disabling command is to actually enable or disable basic guard types in global configuration mode. After you run this command to disable basic guard types, the **display current-config** command allows you to display the global disabling command of each basic guard function. After you run this command to enable basic guard types, the **display current-config** command allows you to display the default configuration.
- This command cannot be saved, but its running result can be saved and take effect after device restart.

1.11.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable all the basic types of global guard of NFPP.

all-guard enable

1.12 Configuring NFPP Logs

1.12.1 Overview

When many attacks are launched to a device, the generated prompts may affect user experience. You can configure the size of a log buffer, system message rate, and filtering conditions to restrictively record and display logs.

- NFPP retrieves logs from a special buffer at a rate, uses them to generate system messages, and clears the logs from the buffer. Therefore, you can configure the rate of generating system messages from logs of the log buffer through NFPP to control the message report frequency and avoid log spamming.
- You can also configure filtering conditions to obtain required log information accurately and filter out logs that do not comply with the filtering rules.
- To monitor whether an attack occurs, you can print logs on a screen in real time.

1.12.2 Restrictions and Guidelines

- After active/standby switchover of a device, the device will clear logs in a buffer and record new logs again.
- Even if a device does not suffer an attack, logs are continuously retrieved from the buffer to generate system messages.
- When the log buffer overflows, new logs overwrite the stale logs, and the log buffer displays an entry with attributes being "-". In this case, you must increase the log buffer size or improve the generation rate of system messages.

1.12.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the NFPP configuration mode.

nfpp

(4) Configure the size of a log buffer.

log_buf entries *number*

The default size of the NFPP log buffer is **256** entries.

(5) Configure the rate of generating system messages from logs of the log buffer through NFPP.

log_buf log_num *message-number* **interval_time** *interval*

No rate of generating system messages from logs of the log buffer is configured through NFPP and NFPP logs are not written into the buffer by default.

If the values of the *message-number* and *interval* parameters are **0**, logs are used by NFPP to immediately generate system messages without being written into the buffer.

(6) Configure NFPP to record the logs of a specified VLAN and a specified interface.

logging { **interface** *interface-type interface-number* | **vlan** *vlan-range }*

NFPP records the logs of all VLANs and interfaces by default.

(7) Enable the function of screen log output.

log_buf enable

The function of screen log output is disabled by default and logs are saved in the buffer.

1.13 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** command may lose vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup nfpp arp-guard scan	Clears the scanning table of ARP guard.
cleanup nfpp arp-guard hosts	Clears the monitored hosts of ARP guard.
cleanup nfpp ip-guard hosts	Clears the monitored hosts of IP guard.
cleanup nfpp tcp-syn-guard hosts	Clears the monitored hosts of TCP-SYN guard.
cleanup nfpp nd-guard hosts	Clears the monitored hosts of ND guard.
cleanup nfpp icmp-guard hosts	Clears the monitored hosts of ICMP guard.
cleanup nfpp dhcp-guard hosts	Clears the monitored hosts of DHCP guard.
cleanup nfpp dhcpv6-guard hosts	Clears the monitored hosts of DHCPv6 guard.
cleanup nfpp define *define-name* **hosts**	Clears the monitored hosts of customized guard.
cleanup nfpp log	Clears NFPP log buffer.
display nfpp arp-guard summary	Displays ARP guard configuration.
display nfpp arp-guard hosts	Displays the monitored hosts of ARP guard.
display nfpp arp-guard scan	Displays the scanning table of ARP guard.
display nfpp ip-guard summary	Displays the IP guard configuration.
display nfpp ip-guard hosts	Displays the monitored hosts of IP guard.
display nfpp ip-guard trusted-host	Displays the trusted hosts of IP guard.
display nfpp tcp-syn-guard summary	Displays the configuration parameters of TCP-SYN guard.
display nfpp tcp-syn-guard hosts	Displays the monitored hosts of TCP-SYN guard.
display nfpp tcp-syn-guard trusted-host	Displays the scanning table of TCP-SYN guard.
display nfpp icmp-guard summary	Displays the configuration parameters of ICMP guard.
display nfpp icmp-guard hosts	Displays the monitored hosts of ICMP guard.
display nfpp icmp-guard trusted-host	Displays the trusted hosts of ICMP guard.
display nfpp dhcp-guard summary	Displays the configuration parameters of DHCP guard.
display nfpp dhcp-guard hosts	Displays the monitored hosts of DHCP guard.
display nfpp dhcpv6-guard summary	Displays the DHCPv6 guard configuration.
display nfpp dhcpv6-guard hosts	Displays the monitored hosts of DHCPv6 guard.
display nfpp nd-guard summary	Displays the ND guard configuration.

Command	Purpose
display nfpp nd-guard hosts	Displays the trusted hosts of ND guard.
display nfpp define summary [*name*]	Displays the customized guard configuration.
display nfpp define hosts *name*	Displays the monitored hosts of the customized guard.
display nfpp define trusted-host *name*	Displays the trusted hosts of the customized guard.
display nfpp log summary	Displays the statistics of the NFPP log buffer.
display nfpp log buffer [**statistics**]	Displays the log buffer of NFPP.

1.14 Configuration Examples

1.14.1 Configuring CPU Guard

1. Requirements

In a network system, protocols such as ARP, IP, TCP-SYN, ICMP, DHCP, DHCPv6, and ND are popular targets of attacks. This may cause packet forwarding failure and high CPU usage. You can enable guard policies to ensure the stable running of devices.

2. Notes

- Configure the rate limiting threshold, attack threshold and isolation time of ARP guard.
- Configure the rate limiting threshold, attack threshold and isolation time of IP guard.
- Configure the rate limiting threshold, attack threshold and isolation time of TCP-SYN guard.
- Configure the rate limiting threshold, attack threshold and isolation time of ICMP guard.
- Configure the rate limiting threshold, attack threshold and isolation time of DHCP guard.
- Configure the rate limiting threshold and attack threshold of DHCPv6 guard.
- Configure the rate limiting threshold and attack threshold of ND guard.

3. Procedure

Configure ARP guard on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# nfpp
DeviceA(cfg-nfpp)# arp-guard speed-limit per-src-mac 5
DeviceA(cfg-nfpp)# arp-guard attack-threshold per-src-mac 10
DeviceA(cfg-nfpp)# arp-guard isolate-period 180
```

Configure IP guard on Device A.

```
DeviceA(cfg-nfpp)# ip-guard speed-limit per-src-ip 20
DeviceA(cfg-nfpp)# ip-guard attack-threshold per-src-ip 30
DeviceA(cfg-nfpp)# ip-guard isolate-period 180
```

```
DeviceA(cfg-nfpp)# ip-guard trusted-host 192.168.201.46 255.255.255.255
```

Configure TCP-SYN guard on Device A.

```
DeviceA(cfg-nfpp)# tcp-syn-guard speed-limit per-src-ip 20
DeviceA(cfg-nfpp)# tcp-syn-guard attack-threshold per-src-ip 30
DeviceA(cfg-nfpp)# tcp-syn-guard isolate-period 180
DeviceA(cfg-nfpp)# tcp-syn-guard trusted-host 192.168.201.46 255.255.255.255
```

Configure ICMP guard on Device A.

```
DeviceA(cfg-nfpp)# icmp-guard speed-limit per-src-ip 20
DeviceA(cfg-nfpp)# icmp-guard attack-threshold per-src-ip 30
DeviceA(cfg-nfpp)# icmp-guard isolate-period 180
```

Configure DHCP guard on Device A.

```
DeviceA(cfg-nfpp)# dhcp-guard speed-limit per-src-ip 20
DeviceA(cfg-nfpp)# dhcp-guard attack-threshold per-src-ip 30
DeviceA(cfg-nfpp)# dhcp-guard isolate-period 180
DeviceA(cfg-nfpp)# dhcp-guard trusted-host 192.168.201.46 255.255.255.255
```

Configure DHCPv6 guard on Device A.

```
DeviceA(cfg-nfpp)# dhcpv6-guard speed-limit per-src-ip 20
DeviceA(cfg-nfpp)# dhcpv6-guard attack-threshold per-src-ip 30
DeviceA(cfg-nfpp)# dhcpv6-guard trusted-host 192.168.201.46 255.255.255.255
```

Configure ND guard on Device A.

```
DeviceA (cfg-nfpp)# nd-guard speed-limit per-port ns-na 30
DeviceA (cfg-nfpp)# nd-guard attack-threshold per-port ns-na 50
```

4. Verification

Run the **display nfpp arp-guard summary** command to display the configuration.

```
DeviceA# display nfpp arp-guard summary
(Format of column Speed-limit and Attack-threshold is per-src-ip/per-src-mac/per-port.)
Interface Status    Isolate-period Speed-limit          Attack-threshold Scan-threshold
Global       Disable 180                  4/5/100             8/10/200             15

Maximum count of monitored hosts: 1000
Monitor period: 600s
```

Run the **display nfpp arp-guard hosts** command to display the monitored hosts.

```
DeviceA# display nfpp arp-guard hosts
If col_filter 1 displays '*', it means "hardware do not isolate host".
 VLAN       interface     IP address             MAC address         remain-time(s)
 ----       ---------     ----------             -----------         --------------
 1            Gi0/43          5.5.5.16               -                    175
Total: 1 host
```

Run the **display nfpp arp-guard scan** command to display the scanned hosts.

```
DeviceA# display nfpp arp-guard scan
VLAN    interface           IP address             MAC address         timestamp
```

```
----  ---------  ----------  -----------  ---------
1     Gi0/5      -           001a.a9c2.4609  2013-4-30 23:50:32
1     Gi0/5      192.168.206.2  001a.a9c2.4609  2013-4-30 23:50:33
1     Gi0/5      -           001a.a9c2.4609  2013-4-30 23:51:33
1     Gi0/5      192.168.206.2  001a.a9c2.4609  2013-4-30 23:51:34
Total: 4 record(s)
```

Note

If attack traffic carries the VLAN ID 0, NFPP may consider the corresponding host as an attack host and configure a rate limiting threshold for the attack traffic. Run the **display nfpp arp-guard host** command to display the attack hosts. The VLAN ID field may be displayed as a port PVID or hidden.

Run the **display nfpp ip-guard summary** command to display the configuration.

```
DeviceA# display nfpp ip-guard summary
(Format of column Speed-limit and Attack-threshold is per-src-ip/per-src-mac/per-port.)
Interface Status   Isolate-period Speed-limit      Attack-threshold Scan-threshold
Global    Disable 180            20/-/100        30/-/200         100

Maximum count of monitored hosts: 1000
Monitor period: 600s
```

Run the **display nfpp ip-guard hosts** command to display the monitored hosts.

```
DeviceA# display nfpp ip-guard hosts
If col_filter 1 displays '*', it means "hardware do not isolate host".
 VLAN    interface   IP address          Reason    remain-time(s)
 ----    ---------   ----------          ------    --------------
  1      Gi0/5       192.168.201.47      ATTACK    160
Total: 1 host
```

Run the **display nfpp ip-guard trusted-host** command to display the trusted hosts.

```
DeviceA# display nfpp ip-guard trusted-host
IP address              mask
----------              ----
192.168.201.46          255.255.255.255
Total: 1 record(s)
```

Note

If attack traffic carries the VLAN ID 0, NFPP may consider the corresponding host as an attack host and configure a rate limiting threshold for the attack traffic. Run the **display nfpp ip-guard host** command to display the attack hosts. The VLAN ID field may be displayed as a port PVID or hidden.

Run the **display nfpp tcp-syn-guard summary** command to display the configuration.

```
DeviceA# display nfpp tcp-syn-guard summary
(Format of column Speed-limit and Attack-threshold is per-src-ip/per-src-mac/per-port.)
Interface Status   Isolate-period Speed-limit      Attack-threshold Scan-threshold
Global    Disable 180            20/-/100        30/-/200         100
```

```
Maximum count of monitored hosts: 1000
Monitor period: 600s
```

Run the **display nfpp tcp-syn-guard hosts** command to display the monitored hosts.

```
DeviceA# display nfpp tcp-syn-guard hosts
If col_filter 1 displays '*', it means "hardware do not isolate host".
 VLAN      interface     IP address            Reason      remain-time(s)
 ----      ---------     ----------            ------      --------------
  1          Gi0/5          192.168.201.47      ATTACK      160
Total: 1 host
```

Run the **display nfpp tcp-syn-guard trusted-host** command to display the trusted hosts.

```
DeviceA# display nfpp tcp-syn-guard trusted-host
IP address                mask
----------                ----
192.168.201.46          255.255.255.255
Total: 1 record(s)
```

Note

If attack traffic carries the VLAN ID 0, NFPP may consider the corresponding host as an attack host and configure a rate limiting threshold for the attack traffic. Run the **display nfpp tcp-syn-guard host** command to display the attack hosts. The VLAN ID field may be displayed as a port PVID or hidden.

Run the **display nfpp icmp-guard summary** command to display the configuration.

```
DeviceA# display nfpp icmp-guard summary
(Format of column Speed-limit and Attack-threshold is per-src-ip/per-src-mac/per-port.)
Interface Status   Isolate-period Speed-limit          Attack-threshold Scan-threshold
Global      Disable 180                 20/-/100            30/-/200             100

Maximum count of monitored hosts: 1000
Monitor period: 600s
```

Run the **display nfpp icmp-guard hosts** command to display the monitored hosts.

```
DeviceA# display nfpp icmp-guard hosts
If col_filter 1 displays '*', it means "hardware do not isolate host".
 VLAN      interface     IP address            Reason      remain-time(s)
 ----      ---------     ----------            ------      --------------
  1          Gi0/5          192.168.201.47      ATTACK      160
Total: 1 host
```

Run the **display nfpp icmp-guard trusted-host** command to display the trusted hosts.

```
DeviceA# display nfpp icmp-guard trusted-host
IP address                mask
----------                ----
192.168.201.46          255.255.255.255
Total: 1 record(s)
```

Note

If attack traffic carries the VLAN ID 0, NFPP may consider the corresponding host as an attack host and configure a rate limiting threshold for the attack traffic. Run the **display nfpp icmp-guard host** command to display the attack hosts. The VLAN ID field may be displayed as a port PVID or hidden.

Run the **display nfpp dhcp-guard summary** command to display the configuration.

```
DeviceA# display nfpp dhcp-guard summary
(Format of column Speed-limit and Attack-threshold is per-src-ip/per-src-mac/per-port.)
Interface Status    Isolate-period Speed-limit          Attack-threshold Scan-threshold
Global    Disable 180                20/-/100           30/-/200         100

Maximum count of monitored hosts: 1000
Monitor period: 600s
```

Run the **display nfpp dhcp-guard hosts** command to display the monitored hosts.

```
DeviceA# display nfpp dhcp-guard hosts
If col_filter 1 displays '*', it means "hardware do not isolate host".
 VLAN      interface     IP address            Reason      remain-time(s)
 ----      ---------     ----------            ------      --------------
  1          Gi0/5          192.168.201.47      ATTACK      160
Total: 1 host
```

Run the **display nfpp dhcp-guard trusted-host** command to display the trusted hosts.

```
DeviceA# display nfpp dhcp-guard trusted-host
IP address               mask
----------               ----
192.168.201.46           255.255.255.255
Total: 1 record(s)
```

Note

If attack traffic carries the VLAN ID 0, NFPP may consider the corresponding host as an attack host and configure a rate limiting threshold for the attack traffic. Run the **display nfpp dhcp-guard host** command to display the attack hosts. The VLAN ID field may be displayed as a port PVID or hidden.

Run the **display nfpp dhcpv6-guard summary** command to display the configuration.

```
DeviceA# display nfpp dhcpv6-guard summary
(Format of column Speed-limit and Attack-threshold is per-src-ip/per-src-mac/per-port.)
Interface Status    Isolate-period Speed-limit          Attack-threshold Scan-threshold
Global    Disable 180                20/-/100           30/-/200         100

Maximum count of monitored hosts: 1000
Monitor period: 600s
```

Run the **display nfpp dhcpv6-guard hosts** command to display the monitored hosts.

```
DeviceA# display nfpp dhcpv6-guard hosts
```

```
If col_filter 1 displays '*', it means "hardware do not isolate host".
 VLAN      interface     IP address              Reason       remain-time(s)
 ----      ---------     ----------              ------       --------------
   1         Gi0/5          192.168.201.47       ATTACK       160
Total: 1 host
```

Run the **display nfpp dhcpv6-guard trusted-host** command to display the trusted hosts.

```
DeviceA# display nfpp dhcpv6-guard trusted-host
IP address                mask
----------                ----
192.168.201.46            255.255.255.255
Total: 1 record(s)
```

Note

If attack traffic carries the VLAN ID 0, NFPP may consider the corresponding host as an attack host and configure a rate limiting threshold for the attack traffic. Run the **display nfpp dhcpv6-guard host** command to display the attack hosts. The VLAN ID field may be displayed as a port PVID or hidden.

Run the **display nfpp nd-guard summary** command to display the configuration.

```
DeviceA# display nfpp nd-guard summary
(Format of column Speed-limit and   Attack-threshold is NS-NA/RS/RA-REDIRECT/ND-SNP.)
Interface Status   Speed-limit           Attack-threshold
Global      Disable 30/25/25/25
```

Contents

1 Configuring Storm Control....1

1.1 Introduction....1

1.1.1 Overview....1

1.1.2 Principles....1

1.2 Configuring Storm Control....2

1.3 Monitoring....2

1.4 Configuration Examples....3

1 Configuring Storm Control

1.1 Introduction

1.1.1 Overview

When a local area network (LAN) has excess broadcast, multicast, or unknown unicast data flows, the network speed will slow down and packet transmission will have an increased timeout probability. This situation is called a LAN storm. A storm may occur when topology protocol execution or network configuration is incorrect.

Storm control can be implemented to limit broadcast, multicast, or unknown unicast data flows. When the rate of broadcast, multicast, or unknown unicast data flows received over a device port exceeds the specified range, the device transmits only packets in the specified range and discards packets beyond the range until the packet rate falls within the range. This prevents flooded data from entering the LAN and causing a storm.

1.1.2 Principles

1. Storm Control Based on the Percentage-based Bandwidth Threshold

If the rate of data flows received over a device port exceeds the specified bandwidth, excess data flows are discarded until the rate falls within the threshold.

Note

When the storm suppression threshold is set to the percentage-based bandwidth, products may convert the configured value to a multiple of the step supported by the product. Therefore, the actual effective suppression threshold on a port may be smaller than the configured suppression threshold.

2. Storm Control Based on the Packets-per-Second Threshold

If the rate of data flows received over a device port exceeds the configured packets-per-second threshold, excessive data flows are discarded until the rate falls within the threshold.

3. Storm Control Based on the Kilobits-per-Second Threshold

If the rate of data flows received over a device port exceeds the configured kilobits-per-second threshold, excessive data flows are discarded until the rate falls within the threshold.

Note

When the storm suppression threshold is set to kilobits-per-second, products may convert the configured value to a multiple of the step supported by the product. Therefore, the actual effective suppression threshold on a port may be smaller than the configured suppression threshold.

4. Storm Control Action

Within the storm control detection interval (10s by default), when the average rate of received packets exceeds the specified allowed packet rate, you can configure the storm control action to handle the interface. The action will be triggered if any one among the configured broadcast, multicast, or unknown unicast thresholds is violated.

1.2 Configuring Storm Control

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

(4) Enable packet storm control.

storm-control-enable { **broadcast** | **multicast** | **unicast** } [**level** *percent* | **pps** *packets* | *rate*]

Packet storm control is disabled by default.

(5) (Optional) Configure the storm control action.

storm-control-enable action { **shutdown** | **trap** }

No storm control action is configured by default.

The storm control action can be configured only after storm control is enabled.

(6) (Optional) Configure storm control logging.

storm-control-enable enable log

Storm control logging is disabled by default.

(7) (Optional) Configure the interval for storm control detection.

storm-control-enable interval

The default interval for storm control detection is 10s.

1.3 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-1 Monitoring

Command	Purpose
display storm-control-enable [*interface-type interface-number*]	Displays storm control information.

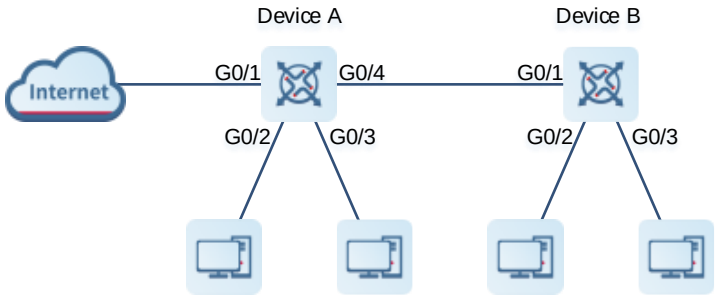
1.4 Configuration Examples

1. Requirements

Enable storm control to prevent excessive broadcast, multicast, or unknown unicast data flows in a LAN, which will cause the network speed to slow down.

2. Topology

Figure 1-1 Topology of Configuring Storm Control



3. Notes

Enable storm control on access devices A and B.

4. Procedure

(1) Enable storm control on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface list gigabitethernet 0/2,0/3,0/4
DeviceA(cfg-if-range)# storm-control-enable broadcast
DeviceA(cfg-if-range)# storm-control-enable multicast
DeviceA(cfg-if-range)# storm-control-enable unicast-pkt
```

(2) Enable storm control on Device B.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceB(cfg)# interface list gigabitethernet 0/1,0/2,0/3
DeviceB(cfg-if-range)# storm-control-enable broadcast
DeviceB(cfg-if-range)# storm-control-enable multicast
DeviceB(cfg-if-range)# storm-control-enable unicast-pkt
```

5. Verification

(1) Check whether storm control is enabled on Device A.

```
DeviceA# display storm-control-enable
Interface                    Broadcast Control Multicast Control Unicast Control Action
------------------------- ----------------- ----------------- ---------------- --------
```

```
        GigabitEthernet 0/1          Disabled          Disabled          Disabled       none
        GigabitEthernet 0/2           default           default           default       none
        GigabitEthernet 0/3           default           default           default       none
        GigabitEthernet 0/4          default           default           default       none
```

(2) Check whether storm control is enabled on Device B.

```
DeviceA# display storm-control-enable
Interface                    Broadcast Control Multicast Control Unicast Control Action
------------------------- ----------------- ----------------- ---------------- --------
        GigabitEthernet 0/1           default           default           default       none
        GigabitEthernet 0/2           default           default           default       none
        GigabitEthernet 0/3           default           default           default       none
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 0/2
 storm-control-enable broadcast
 storm-control-enable multicast
 storm-control-enable unicast-pkt
!
interface GigabitEthernet 0/3
 storm-control-enable broadcast
 storm-control-enable multicast
 storm-control-enable unicast-pkt
!
interface GigabitEthernet 0/4
 storm-control-enable broadcast
 storm-control-enable multicast
 storm-control-enable unicast-pkt
!
End
```

- Device B configuration file

```
hostname DeviceB
!
interface GigabitEthernet 0/1
 storm-control-enable broadcast
 storm-control-enable multicast
 storm-control-enable unicast-pkt
!
interface GigabitEthernet 0/2
 storm-control-enable broadcast
 storm-control-enable multicast
 storm-control-enable unicast-pkt
!
```

```
interface GigabitEthernet 0/3
 storm-control-enable broadcast
 storm-control-enable multicast
 storm-control-enable unicast-pkt
!
end
```

Contents

1 Configuring uRPF....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.1.3 Protocols and Standards....2
1.2 Restrictions and Guidelines....2
1.3 Configuration Task Summary....3
1.4 Enabling uRPF on an Interface....3
1.4.1 Overview....3
1.4.2 Restrictions and Guidelines....3
1.4.3 Procedure....3
1.5 Configuring uRPF Packet Loss Rate Advertisement....4
1.5.1 Overview....4
1.5.2 Restrictions and Guidelines....4
1.5.3 Configuration Tasks....4
1.5.4 Configuring Global Packet Loss Rate Parameters....4
1.5.5 Configuring Packet Loss Rate Parameters on an Interface....5
1.6 Monitoring....5
1.7 Configuration Examples....6
1.7.1 Configuring the uRPF Strict Mode....6
1.7.2 Configuring the uRPF Loose Mode....8

1 Configuring uRPF

Note

S6980-64QC and S6920-4C do not support the uRPF function.

1.1 Introduction

1.1.1 Overview

Unicast Reverse Path Forwarding (uRPF) is a function that protects the network against source address spoofing.

uRPF obtains the source address and inbound interface of a received packet and searches for a forwarding entry in the forwarding table based on the source address. If no entry is configured, the packet is discarded. In uRPF strict mode, if the outbound interface of the forwarding entry does not match the inbound interface of the packet, the packet is also discarded.

1.1.2 Principles

1. Basic Concepts

- uRPF strict mode

  uRPF obtains the source address and inbound interface of a received packet and searches for a forwarding entry in the forwarding table based on the source address. If the entry does not exist, the packet is discarded. If the outbound interface of the forwarding entry does not match the inbound interface of the packet, the packet is also discarded. The strict mode requires that the inbound interface of a received packet must be the outbound interface of the route entry to the source address of the packet.

  The uRPF strict mode is often deployed on a point-to-point (P2P) interface, and inbound and outbound data streams must go through the network of the P2P interface. uRPF blocks the packets with spoofed source addresses at the access layer or aggregation layer to prevent sending such packets from PCs to the core network.

- uRPF loose mode

  uRPF reversely searches for a route based on the source IP address of a received packet. The outbound interface of the next hop on the route does not need to be the inbound interface of the received packet. However, the route cannot be a route of a host on the local network.

  The loose mode is applicable to asymmetric routes or multi-homed networks with asymmetric traffic. On a multi-homed network, the user network is connected to multiple Internet service providers (ISPs), and the inbound and outbound traffic is not symmetric. To prevent invalid packets from attacking the user network, deploy the uRPF loose mode on the outbound interface connected to ISPs.

- uRPF packet loss rate

  The uRPF packet loss rate is the number of packets discarded due to the uRPF check per second. The unit is packets/second, that is, pps.

- Interval for calculating the uRPF packet loss rate

It refers to the interval from the previous time the packet loss rate is calculated to the current time the packet loss rate is calculated.

- Sampling interval of the uRPF packet loss rate

  It refers to the interval at which the number of lost packets is collected for calculating the packet loss rate. This interval must be greater than or equal to the interval for calculating the packet loss rate.

- Threshold of the uRPF packet loss rate

  It refers to the maximum packet loss rate that is acceptable. When the packet loss rate exceeds the threshold, alarms can be sent to users through syslogs or trap messages. You can adjust the threshold of the packet loss rate based on the actual network conditions.

- Alarm interval of the uRPF packet loss rate

  It refers to the interval at which alarms are sent to users. You can adjust the alarm interval based on the actual network conditions to prevent frequent output of logs or trap messages.

2. Enabling uRPF

Enable uRPF to perform a uRPF check on IP packets, thus protecting the device against source address spoofing.

uRPF can be configured in interface configuration mode.

- Interface configuration mode

  uRPF is performed for packets received on the configured interface.

3. Notifying the uRPF Packet Loss Rate

To conveniently monitor lost packets after uRPF is enabled, the device uses syslogs or trap messages to proactively notify users of the packet loss information detected in the uRPF check.

In the period from the time when uRPF is enabled to the time when the sampling interval arrives, the packet loss rate is the number of lost packets measured within the sampling interval divided by the uRPF enabling time. After that, the packet loss rate is calculated as follows:
Current packet loss rate = (Current number of lost packets measured at the calculation interval – Number of lost packets measured before the sampling interval)/Sampling interval

1.1.3 Protocols and Standards

- RFC 2827: Network Ingress Filtering: DDOS Attacks which employ IP Source Address Spoofing
- RFC 3704: Ingress Filtering for Multi-homed Networks

1.2 Restrictions and Guidelines

- uRPF is implemented with the help of existing unicast routes in the network. Therefore, unicast routes must be configured in the network.
- uRPF checks only source addresses of packets whose destination addresses are unicast addresses and does not check packets whose destination addresses are multicast or IPv4 broadcast addresses.
- uRPF cannot be enabled on a range of interfaces.
- uRPF does not check DHCP/BOOTP packets whose source IP address is 0.0.0.0 and destination IP address is 255.255.255.255.

- uRPF does not check loopback packets sent by the local device to itself.

1.3 Configuration Task Summary

uRPF configuration includes the following tasks:

(1) Enabling uRPF.

- Enabling uRPF on an Interface

(2) (Optional) Configuring uRPF Packet Loss Rate Advertisement. Configure at least one of the tasks.

- Configuring Global Packet Loss Rate Parameters
- Configuring Packet Loss Rate Parameters on an Interface

1.4 Enabling uRPF on an Interface

1.4.1 Overview

uRPF is performed for packets received on the configured interface to prevent source address spoofing.

1.4.2 Restrictions and Guidelines

- Both the uRPF strict mode and loose mode can be configured on an interface.
- uRPF cannot be configured in global configuration mode and interface configuration mode simultaneously.
- The default route is not used for a uRPF check by default. You can configure **allow-def-route** to use the default route for a uRPF check if necessary.
- After the uRPF IPv4 command is configured, a switch will perform a uRPF check on both IPv4 packets. Currently.
- The uRPF strict mode supports matching of equal-cost routes. If a packet received on an interface matches an equal-cost route during the uRPF check, the packet will be processed according to the uRPF loose mode.
- After uRPF is enabled, the route forwarding capacity of a device will be reduced by half.
- After uRPF is enabled on interfaces, a uRPF check is performed for all packets received on physical ports corresponding to these interfaces, which increases the scope of packets checked by uRPF. If a packet received on a tunnel port is also received on the preceding physical ports, the packet is also checked by uRPF. In such a scenario, be cautious in enabling uRPF.

Specification

A switch supports configuration of uRPF on a Layer 3 Ethernet interface or link aggregation, also support configuration of uRPF on a switch virtual interface (SVI). For details about the switch products, contact technical support engineers.

- uRPF does not support association with the ACL option.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

(4) Enable uRPF for an interface. Configure at least one of the tasks.

- Enable IPv4 uRPF for an interface.

  ip check unicast-pkt src-address reachable-via { **any** | **rx** } [**allow-def-route**] [*acl-number*]

  IPv4 uRPF is disabled for an interface by default.

  Switches support configuration of IPv4 uRPF on a Layer 3 Ethernet interface or link aggregation.

1.5 Configuring uRPF Packet Loss Rate Advertisement

1.5.1 Overview

After uRPF packet loss information monitoring is enabled, a device can proactively send syslogs or trap messages to notify users of the packet loss information detected in the uRPF check so that the users can monitor the network status conveniently.

1.5.2 Restrictions and Guidelines

uRPF must be enabled.

1.5.3 Configuration Tasks

Configure at least one of the tasks.

- Configuring Global Packet Loss Rate Parameters
- Configuring Packet Loss Rate Parameters on an Interface

1.5.4 Configuring Global Packet Loss Rate Parameters

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure global packet loss rate parameters. Configure at least one of the tasks.

- Configure the interval of calculating the uRPF IPv4 packet loss rate.

  ip check urpf drop-pkt-rate compute interval *interval*

  The default interval for calculating the uRPF IPv4 packet loss rate is **30** seconds.

- Configure the alarm interval for the uRPF IPv4 packet loss rate.

  ip check urpf drop-pkt-rate notify hold-down *hold-down-time*

The default alarm interval for the uRPF IPv4 packet loss rate is **300** seconds.

1.5.5 Configuring Packet Loss Rate Parameters on an Interface

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

(4) Configure packet loss rate parameters on an interface. Configure at least one of the tasks.

- Enable uRPF packet loss information monitoring.

  ip check urpf drop-pkt-rate notify

  uRPF packet loss information monitoring is disabled by default.

- Configure the threshold of the uRPF packet loss rate.

  ip check urpf notification threshold *rate-value*

  The default threshold of the uRPF packet loss rate is **1000** packets per second.

  If the threshold is **0**, a notification is sent for every packet that is discarded because it fails in the uRPF check.

1.6 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** commands to output debugging information.

⚠ Caution

- The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** commands to clear information.

⚠ Caution

- Running the **cleanup** command during operation of the device may lose vital information and interrupt services.

Table 1-1 Monitoring

Command	Purpose
display ip urpf [**interface** *interface-type interface-number*]	Displays the IPv4 uRPF configurations and statistics.
cleanup ip urpf [**interface** *interface-type interface-number*]	Clears IPv4 uRPF packet loss statistics.
debugging urpf event	Debugs uRPF events.
debugging urpf timer	Debugs uRPF timers.

1.7 Configuration Examples

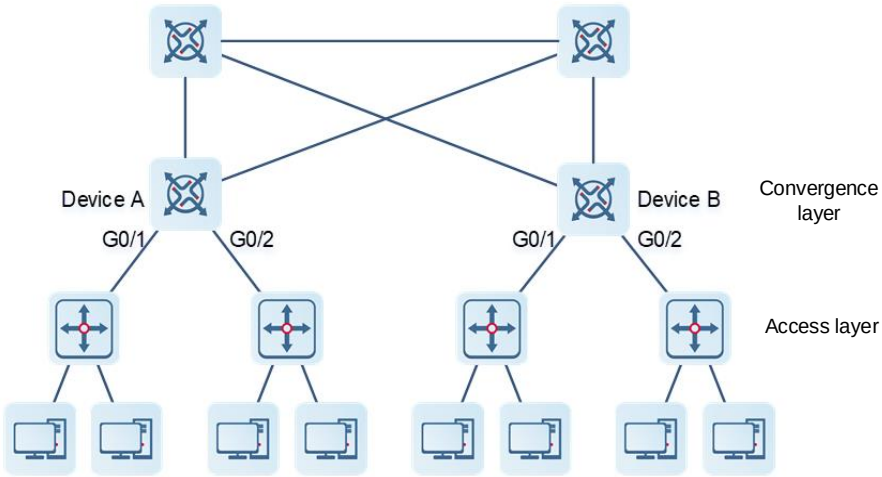
1.7.1 Configuring the uRPF Strict Mode

1. Requirements

In uRPF strict mode, packets with spoofed source addresses are blocked at the access layer or aggregation layer to prevent sending such packets from PCs to the core network.

2. Topology

Figure 1-1 Configuring the uRPF Strict Mode



3. Notes

Enable the uRPF strict mode on the aggregation devices, including Devices A and B.

4. Procedure

(1) Configure Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
```

```
DeviceA(cfg)# interface gigabitethernet0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 195.52.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# ip check unicast-pkt src-address reachable-via rx
DeviceA(cfg-if-GigabitEthernet 0/1)# ip check urpf drop-pkt-rate notify
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
DeviceA(cfg)# interface gigabitethernet0/2
DeviceA(cfg-if-GigabitEthernet 0/2)# ip address 195.52.2.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/2)# ip check unicast-pkt src-address reachable-via rx
DeviceA(cfg-if-GigabitEthernet 0/2)# ip check urpf drop-pkt-rate notify
DeviceA(cfg-if-GigabitEthernet 0/2)# exit
```

(2) Configure Device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitethernet0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 195.52.3.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# ip check unicast-pkt src-address reachable-via rx
DeviceB(cfg-if-GigabitEthernet 0/1)# ip check urpf drop-pkt-rate notify
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
DeviceB(cfg)# interface gigabitethernet0/2
DeviceB(cfg-if-GigabitEthernet 0/2)# ip address 195.52.4.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/2)# ip check unicast-pkt src-address reachable-via rx
DeviceB(cfg-if-GigabitEthernet 0/2)# ip check urpf drop-pkt-rate notify
DeviceB(cfg-if-GigabitEthernet 0/2)# exit
```

5. Verification

If source address spoofing exists on the network, run the **display ip urpf** command to display the number of spoofing packets discarded by uRPF.

Display the number of spoofing packets discarded by uRPF on GigabitEthernet 0/1 of Device A.

```
DeviceA# display ip urpf interface gigabitethernet 0/1
IP check src-address reachable-via RX
IP check URPF drop-pkt-rate notify is enabled
IP check URPF notification threshold is 1000pps
Number of drop packets in this interface is 124
Number of drop-pkt-rate notification counts in this interface is 0
```

Display the number of spoofing packets discarded by uRPF on GigabitEthernet 0/2 of Device A.

```
DeviceA# display ip urpf interface gigabitethernet 0/2
IP check src-address reachable-via RX
IP check URPF drop-pkt-rate notify is enabled
IP check URPF notification threshold is 1000pps
Number of drop packets in this interface is 133
Number of drop-pkt-rate notification counts in this interface is 0
```

Display the number of spoofing packets discarded by uRPF on GigabitEthernet 0/1 of Device B.

```
DeviceB# display ip urpf interface gigabitethernet 0/1
IP check src-address reachable-via RX
```

```
IP check URPF drop-pkt-rate notify is enabled
IP check URPF notification threshold is 1000pps
Number of drop packets in this interface is 124
Number of drop-pkt-rate notification counts in this interface is 0
```

Display the number of spoofing packets discarded by uRPF on GigabitEthernet 0/2 of Device B.

```
DeviceB# display ip urpf interface gigabitethernet 0/2
IP check src-address reachable-via RX
IP check URPF drop-pkt-rate notify is enabled
IP check URPF notification threshold is 1000pps
Number of drop packets in this interface is 250
Number of drop-pkt-rate notification counts in this interface is 0
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 0/1
 ip address 195.52.1.1 255.255.255.0
 ip check unicast-pkt src-address reachable-via rx
 ip check urpf drop-pkt-rate notify
!
interface GigabitEthernet 0/2
 ip address 195.52.2.1 255.255.255.0
 ip check unicast-pkt src-address reachable-via rx
 ip check urpf drop-pkt-rate notify
!
end
```

- Device B configuration file

```
hostname DeviceB
!
interface GigabitEthernet 0/1
 ip address 195.52.3.1 255.255.255.0
 ip check unicast-pkt src-address reachable-via rx
 ip check urpf drop-pkt-rate notify
!
interface GigabitEthernet 0/2
 ip address 195.52.4.1 255.255.255.0
 ip check unicast-pkt src-address reachable-via rx
 ip check urpf drop-pkt-rate notify
!
end
```

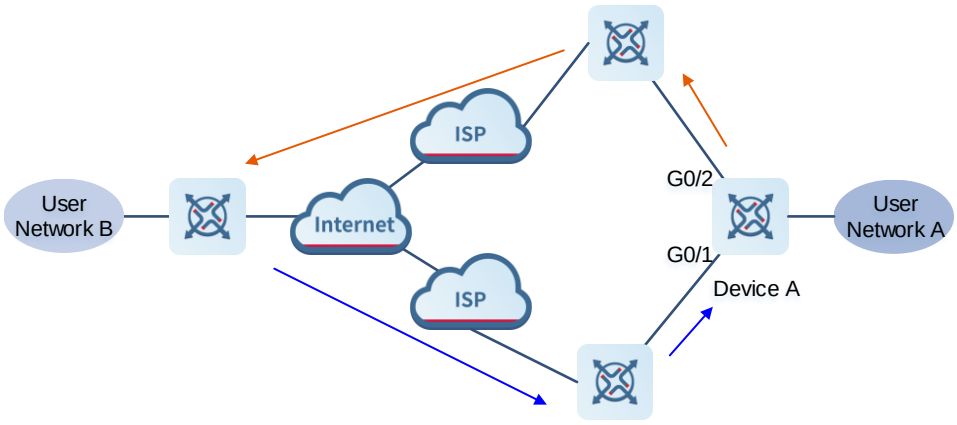
1.7.2 Configuring the uRPF Loose Mode

1. Requirements

To prevent invalid packets from attacking the user network, deploy the uRPF loose mode.

2. Topology

Figure 1-2 Configuring the uRPF Loose Mode



3. Notes

Configure the uRPF loose mode on the outbound interfaces connected to ISPs.

4. Procedure

Enable the uRPF loose mode on GigabitEthernet 0/1 and GigabitEthernet 0/2.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet0/1)# ip address 195.52.1.2 255.255.255.252
DeviceA(cfg-if-GigabitEthernet0/1)# ip check unicast-pkt src-address reachable-via any
DeviceA(cfg-if-GigabitEthernet0/1)# ip check urpf drop-pkt-rate notify
DeviceA(cfg-if-GigabitEthernet0/1)# exit
DeviceA(cfg)# interface gigabitethernet 0/2
DeviceA(cfg-if-GigabitEthernet0/2)# ip address 152.95.1.2 255.255.255.252
DeviceA(cfg-if-GigabitEthernet0/2)# ip check unicast-pkt src-address reachable-via any
DeviceA(cfg-if-GigabitEthernet0/2)# ip check urpf drop-pkt-rate notify
DeviceA(cfg-if-GigabitEthernet0/2)# end
```

5. Verification

If source address spoofing exists on the network, run the **display ip urpf** command to display the number of spoofing packets discarded by uRPF.

```
DeviceA# display ip urpf
IP check URPF drop-pkt-rate compute interval is 300s
IP check URPF drop-pkt-rate notify hold-down is 300s
Interface gigabitEthernet0/1
IP check src-address reachable-via ANY
IP check URPF drop-pkt-rate notify is enabled
IP check URPF notification threshold is 1000pps
Number of drop packets in this interface is 4121
```

```
Number of drop-pkt-rate notification counts in this interface is 2
Interface gigabitEthernet0/2
IP check src-address reachable-via ANY
IP check URPF drop-pkt-rate notify is enabled
IP check URPF notification threshold is 1000pps
Number of drop packets in this interface is 352
Number of drop-pkt-rate notification counts in this interface is 0
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 0/1
  ip address 195.52.1.2 255.255.255.252
  ip check unicast-pkt src-address reachable-via any
  ip check urpf drop-pkt-rate notify
!
interface GigabitEthernet 0/2
  ip address 152.95.1.2 255.255.255.252
  ip check unicast-pkt src-address reachable-via any
  ip check urpf drop-pkt-rate notify
!
end
```

Contents

1 Configuring Security Log Auditing 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Restrictions and Guidelines 2
1.3 Configuration Task Summary 2
1.4 Configuring Security Log Auditing 2
1.4.1 Overview 2
1.4.2 Enabling Security Log Auditing 2
1.4.3 Configuring the Local Storage Time for Security Logs 3
1.4.4 Configuring the Handling Time of Aged Security Logs 3
1.4.5 Configuring the Local Storage Capacity for Security Logs 3
1.5 Monitoring 4
1.6 Configuration Examples 5
1.6.1 Configuring Security Log Auditing 5

1 Configuring Security Log Auditing

1.1 Introduction

1.1.1 Overview

Security log auditing is used to record key operations on a device and audit and backtrack these operations afterwards to improve the device security and meet the national security standards.

Note

The following sections describe content related to security log auditing only.

1.1.2 Principles

1. Basic Concepts

- Key operation

  After enabling the security log auditing function, the device records logs for key operations, including account management, login events, system events, configuration file changes, and security log events. Key operations include:

  - Adding/deleting accounts
  - Editing authentication information
  - Modifying configurations (such as the DNS address and IP address)
  - User login/logout
  - Restarting/Stopping the device
  - Uploading/downloading files (when supported)
  - Editing user permissions (when supported)
  - Enabling/disabling log auditing and deleting logs

2. Enabling Security Log Auditing

After the security log auditing function is enabled, the device can record logs for key operations and locally store the logs.

When a user performs a key operation, such as account management, login/logout, system restart, configuration file change, or log auditing function enabling/disabling, the corresponding service module will collect information (such as time, username, IP address, operation log content, and operation result), generate a log, and write the log to the local database through the log auditing framework.

At 03:00:00 a.m. each day, the system checks whether any logs for key operations exceed the storage time. If yes, the logs will be deleted.

A maximum of 10,000 logs can be stored by default. After the number of stored logs exceeds the storage limit, new logs will overwrite the earliest ones.

3. Enabling the Function of Sending Security Logs to the Syslog Server

After a device is configured with the functions of syslog, and the security log auditing and sending functions are enabled, the log auditing framework will send generated logs to the syslog server to store.

1.2 Restrictions and Guidelines

- To protect the flash storage chip, newly generated logs will be cached in the memory and written to a file every 1 hour. When a restart command, such as **reboot**, is executed, logs will be written to a file. If many logs are generated in this hour and some logs are not written to the file in 1 hour, the logs will be lost.
- When logs are exported to a file, download the file in time. The file will be automatically deleted after 1 hour.

1.3 Configuration Task Summary

Configuration of security log auditing includes the following tasks:

(1) Enabling Security Log Auditing

(2) (Optional) Configuring the Local Storage Time for Security Logs

(3) (Optional) Configuring the Handling Time of Aged Security Logs

(4) (Optional) Configuring the Local Storage Capacity for Security Logs

1.4 Configuring Security Log Auditing

1.4.1 Overview

The security log auditing function can record logs for key operations on a device and locally store the logs. Log query and export are supported.

1.4.2 Enabling Security Log Auditing

1. Overview

After the security log auditing function is enabled, the device will record logs for key operations, including account management, login events, system events, configuration file changes and security log events, and store the logs to the local database.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable security log auditing.

security-log audit-enable

Security log auditing is enabled by default.

1.4.3 Configuring the Local Storage Time for Security Logs

1. Overview

After the local storage time is configured for security logs, the logs that have exceeded the storage time will be deleted.

2. Restrictions and Guidelines

Security logs that have exceeded the storage time will be deleted.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the local storage time for security logs.

security-log data-store-days *data-store-time*

The default local storage time for security logs is **65535** days.

1.4.4 Configuring the Handling Time of Aged Security Logs

1. Overview

After the handling time of aged security logs is configured, the system checks whether any logs have exceeded the storage time at the configured handling time (deviation: 5 minutes) and deletes expired logs.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the handling time of aged security logs.

security-log auto-vacuum-time *hh*:*mm*:*ss*

The default handling time of aged security logs is **03:00:00** every day.

1.4.5 Configuring the Local Storage Capacity for Security Logs

1. Overview

After the local storage capacity for security logs is configured, new logs will overwrite earliest ones when the capacity of locally stored logs exceeds the storage limit.

2. Restrictions and Guidelines

If flash space is limited, you can run this command to decrease the storage capacity for security logs.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the local storage capacity for security logs.

security-log data-store-items *log-number*

The default local storage capacity for security logs is **10000**.

1.5 Monitoring

Run the **delete** commands to clear information.

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** commands to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 Monitoring

Command	Purpose
debugging security-log errors	Debugs log auditing.
debugging security-log recv-info	Debugs received logs.
debugging security-log sql-info	Debugs the log database.
display security-log	Displays all logs.
display security-log config	Displays configurations.
display security-log detail { **all** | { **from** *YY//MM/DD hh*:*mm*:*ss* **to** *YY//MM/DD hh*:*mm*:*ss* } } [**hostname** *hostname*] [**log-type** { **acc_mnt** | **config** | **login** | **other** | **sec_log** | **sys** }] [**peerinfo** *peerinfo*] [**user** *username*] { [**order-by** { **log-type** | **time** } { **asc** | **desc** } [**start-item** *start-item* **end-item** *end-item*]] }	Displays the detailed log information, which can be filtered by time, log type, username, host name, and terminal information.
display security-log detail export { **all** | { **from** *YY//MM/DD hh*:*mm*:*ss* **to** *YY//MM/DD hh*:*mm*:*ss* } } [**hostname** *hostname*] [**log-type** { **acc_mnt** | **config** | **login** | **other** | **sec_log** | **sys** }] [**peerinfo** *peerinfo*] [**user** *username*] { [**order-by** { **log-type** | **time** } { **asc** | **desc** } [**start-item** *start-item* **end-item** *end-item*]] }	Exports the detailed log information, which can be filtered by time, log type, username, host name, and terminal information.

Command	Purpose
display security-log detail stat { **all** | { **from** *YY//MM/DD hh:mm:ss* **to** *YY//MM/DD hh:mm:ss* } } [**hostname** *hostname*] [**log-type** { **acc_mnt** | **config** | **login** | **other** | **sec_log** | **sys** }] [**peerinfo** *peerinfo*] [**user** *username*]	Displays the log statistics, which can be filtered by time, log type, username, host name, and terminal information.
display security-log info	Displays the statistics during log processing.
display security-log statistics	Displays the statistics.
security-log delete all	Deletes all logs.

1.6 Configuration Examples

1.6.1 Configuring Security Log Auditing

1. Requirements

To improve device security and audit and backtrack exceptions (such as key configurations being tampered) afterwards, key operations on a device need to be recorded. After the security log auditing function is enabled, the device records key operations, including account management, login events, system events, configuration file changes, and security log events. After logging in to the device, the administrator can monitor recent user behaviors through the records.

2. Topology

Figure 1-1 Configuring Security Log Auditing



3. Notes

- Enable security log auditing.
- Configure parameters related to security log auditing.

4. Procedure

Enable security log auditing.

```
Device> privilege
Device# config-system terminal
Device(cfg)# security-log audit-enable
```

Set the local storage time of security logs to **300** days.

```
Device(cfg)# security-log data-store-days 300
```

Set the handling time of aged security logs to **05:05:00** every day.

```
Device(cfg)# security-log auto-vacuum-time 05:05:00
```

Set the local storage capacity for security logs to **1000**.

```
Device(cfg)# security-log data-store-items 1000
```

5. Verification

Log in to the device, perform key operations, and run **display security-log detail all** 30s later to check the log auditing records.

```
Device# display security-log detail all
time, username, peerinfo, hostname, log-type: content
2019-10-22 10:00:02, admin, vty0(192.168.111.111), Device, SEC_LOG: SECURITY_LOG deleted all security log successfully
2019-10-22 10:00:02, admin, vty0(192.168.111.111), Device, SEC_LOG: SECURITY_LOG disabled security log audit configuration unsuccessfully
2019-10-22 10:00:03, ---, console, Device, SEC_LOG: SECURITY_LOG enabled security log audit configuration successfully
```

Run the **display security-log config** command to verify that the local storage time of security logs is 300 days, the handling time of aged security logs is 05:05:00 every day, and the local storage capacity for security logs is 1000.

```
Device# display security-log config
Security-log audit: enable
Limit number: 10000
Store days: 300
Auto vacuum time: 05:05:00
```

6. Configuration Files

```
hostname Device
!
security-log data-store-days 300
security-log data-store-items 1000
security-log auto-vacuum-time 05:05:00
!
end
```

Reliability Configuration

1. Configuring REUP
2. Configuring RLDP
3. Configuring DLDP
4. Configuring M-LAG
5. Configuring VRRP
6. Configuring VRRP Plus
7. Configuring BFD
8. Configuring Track
9. Configuring IP Event Dampening
10. Configuring HAM
11. Configuring DAG

Contents

1 Configuring REUP 1

1.1 Introduction 1

1.1.1 Topology 1

1.1.2 Basic Concepts 2

1.1.3 Dual-Link Backup 2

1.1.4 MAC Address Update 4

1.1.5 VLAN Load Balancing 5

1.1.6 Link State Tracking 6

1.2 Configuration Task Summary 7

1.3 Configuring Basic Features of REUP 8

1.3.1 Overview 8

1.3.2 Restrictions and Guidelines 8

1.3.3 Procedure 8

1.4 Configuring the Preemption Function of REUP 8

1.4.1 Overview 8

1.4.2 Restrictions and Guidelines 9

1.4.3 Procedure 9

1.5 Configuring the MAC Address Update Function 9

1.5.1 Overview 9

1.5.2 Restrictions and Guidelines 10

1.5.3 Procedure 10

1.6 Configuring VLAN Load Balancing 11

1.6.1 Overview 11
1.6.2 Restrictions and Guidelines 11
1.6.3 Procedure 11
1.7 Configuring Link State Tracking 12
1.7.1 Overview 12
1.7.2 Restrictions and Guidelines 12
1.7.3 Procedure 12
1.8 Monitoring 13
1.9 Configuration Examples 14
1.9.1 Configuring Basic Features of REUP 14
1.9.2 Configuring the MAC Address Update Function 16
1.9.3 Configuring VLAN Load Balancing 19
1.9.4 Configuring Link State Tracking 22

1 Configuring REUP

Rapid Ethernet Uplink Protection (REUP) provides a rapid uplink protection function. In the dual uplink networking, REUP is used to ensure normal communication between links, block redundant links, avoid link loops, and implement fast backup.

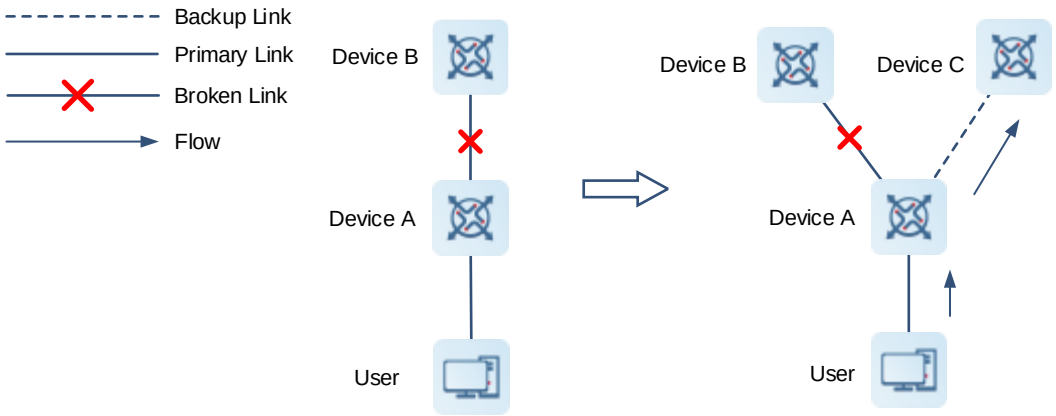
1.1 Introduction

1.1.1 Topology

The left topology in Figure 1-1 shows single uplink networking, where there is only one link between the user and the distribution device to implement communication. When the link between access device A and distribution device B fails, the link between the user and device B is disconnected, which interrupts all external services of the user.

The topology can be transformed into REUP dual uplink deployment topology, as shown in the right topology in Figure 1-1. In this topology, one link functions as the active link for normal communication, and the other link functions as a backup link and is not used for communication. When the active link fails, traffic is switched to the backup link. This helps reduce the impact of a single link failure on the network and improve reliability.

Figure 1-1 REUP Topology



REUP and the Spanning Tree Protocol (STP) are mutually exclusive on ports. Generally, a device runs STP in the downlink direction and runs REUP in the uplink direction to implement uplink backup and data protection in the case of faults. REUP can still provide basic link redundancy when STP is disabled. It also provides millisecond-level fault recovery faster than STP. Its main features are as follows.

- The uplink ports of REUP are configured in pairs, namely, an uplink port pair includes a forward port and a standby port. The standby port does not forward data packets. When the forward port fails, the standby port immediately becomes the forward port to forward data.
- REUP sends an address update packet to the upstream device so that the upstream device can update the MAC address information immediately. This function ensures that L2 data streams are restored rapidly after a link gets faulty.

1.1.2 Basic Concepts

1. REUP Pair

You can specify a port as the backup port of another port to configure an REUP pair, in which one port is the active port and the other is the backup port. When both ports are normal, one port is configured as the forward port and the other as the standby port.

2. MAC Address Update Message

According to the MAC address learning principle, before L2 link switching a user's MAC address entry exists on the forwarding link, and no packet is forwarded by the blocked link, so there is no user's MAC address entry on the blocked link. After switching, packets cannot be forwarded because they are still forwarded according to the existing MAC address entry of the user. MAC address update messages refer to FLUSH packets sent to an uplink device through private multicast packets. When an uplink device, on which the function of receiving MAC address update messages is enabled, receives a MAC address update message, the device updates the MAC entry on the relevant interface.

3. MAC Address Update Group

Multiple ports can be added to a group. If one port in the group receives a MAC address update message, the MAC address entries on other ports in the group are updated at the same time. In this case, the group is called a MAC address update group.

4. MAC Address Update Packet

MAC address update packets are sent to uplink devices of other manufacturers to ask them to update their MAC address entries, so as to implement compatibility with them.

5. Link State Tracking Group

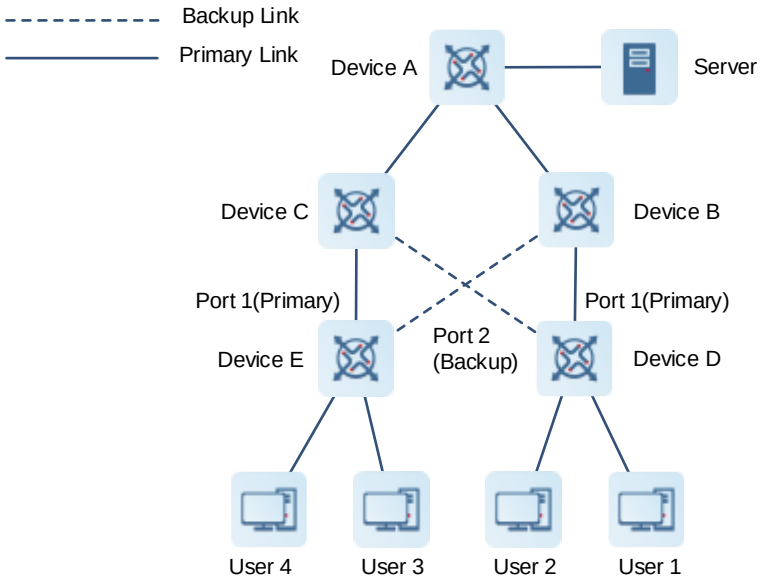
The uplink and downlink ports of a device can be added to a group at the same time, which is called a link state tracking group. If all uplink ports in the group are down, all downlink ports in this group are forced down.

1.1.3 Dual-Link Backup

When an active link is faulty, the link in the backup state will rapidly device to the forward state and start to forward data, to minimize the service interruption caused by the link failure. REUP adopts the real-time backup and batch backup mechanisms to ensure that the configuration and operating status of the standby device are completely consistent with those of the active device. When the master/slave switching occurs in a virtual device unit (VSU) scenario, the data flow is not cut off, the operating status of REUP is consistent with that before the switching, and link switching is not performed again.

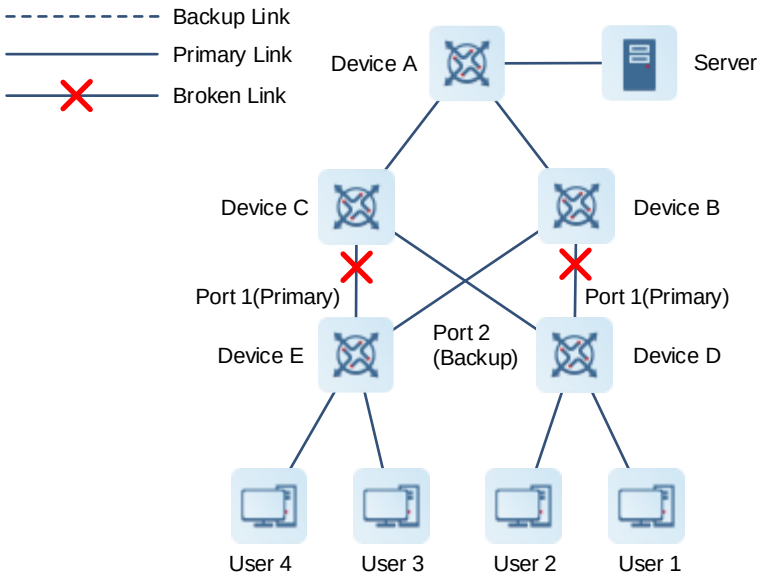
- You can specify a port as the backup port of another port to configure an REUP pair. When the two ports are normal, one link is in the forwarding state (forwarding data packets) and the other link is in the backup state (not forwarding data packets).
- When the active link is faulty, the link in the backup state rapidly switches to the forwarding state and starts to forward data.
- When the faulty link recovers, it enters the backup state and does not forward data packets.
- The preemption mode can be configured to specify whether the link recovering from a fault preempts to become the link in the forwarding state.

Figure 1-2 Configuring Two Active Links



As shown in Figure 1-2, ports 1 and 2 of device D (device E) are connected to uplink switches B and C (switches C and B) respectively, and REUP is configured on ports 1 and 2. When the links are normal, port 1 is in the forwarding state and forwards data packets and port 2 is in the backup state and does not forward data packets.

Figure 1-3 Topology with Device D and Device E1Faulty



As shown in Figure 1-3, once port 1 fails, port 2 immediately starts to forward data packets and restores uplink transmission of the device. In the non-preemption mode, when the link of port 1 recovers, port 1 is in the backup state and does not forward data packets, and port 2 continues forwarding data packets. You must enable the REUP dual-link backup function on a port so that the port undertakes the REUP link switching when the port is faulty.

Note

- REUP, Ethernet Ring Protection Switching (ERPS), and Rapid Ethernet Ring Protection (RERP) do not share ports.
- On devices enabled with REUP, the storm control function must be disabled on all L2 ports.

1.1.4 MAC Address Update

During link switching, the MAC address entries on a port are updated to speed up packet convergence. As shown in Figure 1-2, the MAC address update process is analyzed as follows:

- Ports 1 and 2 of device D (device E) have the REUP dual-link backup enabled. Port 1 works as the active port. During normal communication, device A learns the MAC addresses of users 1 and 2 (users 3 and 4) from the port connecting to device B (device C).
- When port 1 of device D (device E) fails, port 2 quickly changes to the forwarding state and starts to forward data packets. In this case, device A has not learned the MAC addresses of users 1 and 2 (users 3 and 4) from the port connecting to device B (device C). The data packets sent by the server to users 1 and 2 (users 3 and 4) are forwarded to device C (B) by device A, causing loss of the packets from the server to users 1 and 2 (users 3 and MAC 4).

The following measures can be taken to avoid the above problem:

- Enable the MAC address update function on device D (device E). When port 2 starts to forward packets, device D (device E) sends a MAC address update message to port 2.
- After receiving the MAC address update message from port 2, device A updates the MAC address entries on the port of device A. In this way, device A forwards the packets sent by the server to the users, to the port connecting to device C (device B) to speed up packet transmission convergence.
- In addition, a MAC address update group can be configured, that is, multiple ports can be classified into the same group. When a port in this group receives a MAC address update message, the MAC address entries on other ports in the group are updated, to reduce the side effect of flooding caused by MAC address update.
- To be compatible with upstream devices not supporting MAC address update messages, device D (device E) sends MAC address update packets for users 1 and user 2 (users 3 and 4) upstream when port 2 switches to the forwarding state. In this way, device A updates the MAC addresses of users 1 and 2 (users 3 and 4) to the corresponding port and the downstream data transmission of device A is restored.

Based on the above process analysis, the relevant configuration to be involved includes the following items:

Table 1-1 Description of MAC Address Update Configuration Items

Configuration Item	Description
Sending MAC address updates	If the function of sending MAC address update messages is not enabled, no MAC address update message is sent when REUP dual-link backup switching is performed.

Configuration Item	Description
Receiving MAC address updates	If the function of receiving MAC address update messages is not enabled, a device cannot receive MAC address update messages from downlink devices during REUP dual-link backup switching and will not update the MAC addresses.
Specifying a VLAN for sending MAC address updates	If the VLAN in which an interface sends MAC address update messages is configured, the messages are sent in the configured VLAN; otherwise, the messages are sent in the default VLAN to which the interface belongs.
Specifying a VLAN for receiving MAC address updates	If no VLAN in which an interface receives MAC address update messages is configured, MAC address update messages are received in all the configured VLANs; otherwise, MAC address update messages are received in the specified VLAN.
MAC address update group	If no MAC address update group is configured, MAC address update is not performed when MAC address update packets are received.
Maximum number of MAC address update messages sent per second	A greater number of messages configured indicates more CPU time used for sending messages and fewer downstream packets to be lost.

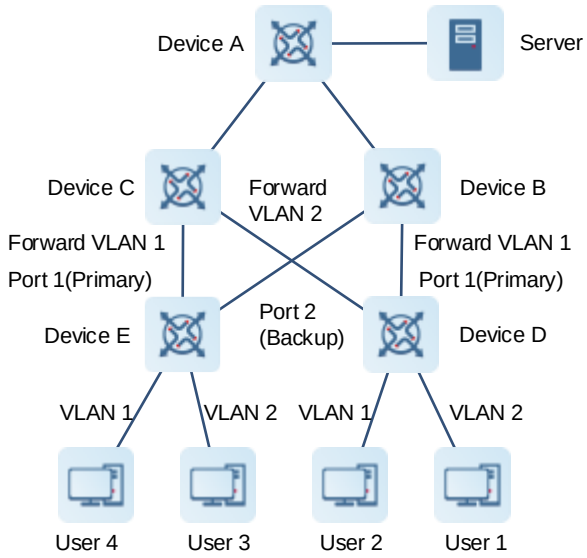
1.1.5 VLAN Load Balancing

The VLAN load balancing function allows the two ports of a REUP pair to forward data packets of mutually exclusive VLANs so as to make full use of the link bandwidth.

As shown in Figure 1-4, the MAC address update process is analyzed as follows:

- Configure REUP dual-link backup and enable the VLAN load balancing of REUP on ports 1 and 2 of device D, and map VLAN 1 to instance 1 and VLAN 2 to instance 2. Data of VLAN 1 (instance 1) is transmitted through port 1 and data of VLAN 2 (instance 2) is transmitted through port 2. Perform the same configuration on device E.
- When a port is faulty, the other port takes over the transmission of all VLAN data.
- When the faulty port recovers and works properly within the preemption delay, the transmission of VLAN data in the charge of the faulty port is switched back to the recovered port.

Figure 1-4 Configuring VLAN Load Balancing



Therefore, the VLAN load balancing function needs to be enabled on the ports to make full use of the link bandwidth for packet forwarding when both links are normal.

Caution

- The instance mapping of REUP VLAN load balancing is controlled by the Multiple Spanning Tree Protocol (MSTP) module in a unified manner. For details about how to configure the instances, see *Configuring MSTP*.
- The VLAN load balancing function can be configured only on trunk, uplink, or hybrid ports.

1.1.6 Link State Tracking

Link state tracking means that, when an upstream link is faulty, the downstream link switches services to the backup upstream link so that the backup port can continue to forward packets. Link state tracking provides the function of notifying a downlink device to perform link switching when all the upstream links are faulty. You can configure the upstream and downstream ports of a link state tracking group and bind the link states of multiple downstream ports to multiple uplink ports to implement link state synchronization. In default configuration, when all uplinks in a tracking group are faulty, the downlink ports are shut down forcibly to ensure that the downlink transmission is switched from the active link to the backup link.

As shown in Figure 1-5, when the uplink of device B is faulty, link state tracking rapidly shuts down the downstream port of device B so that the upstream transmission of device D is switched to device C.

Figure 1-5 Topology with the Active Uplink Faulty

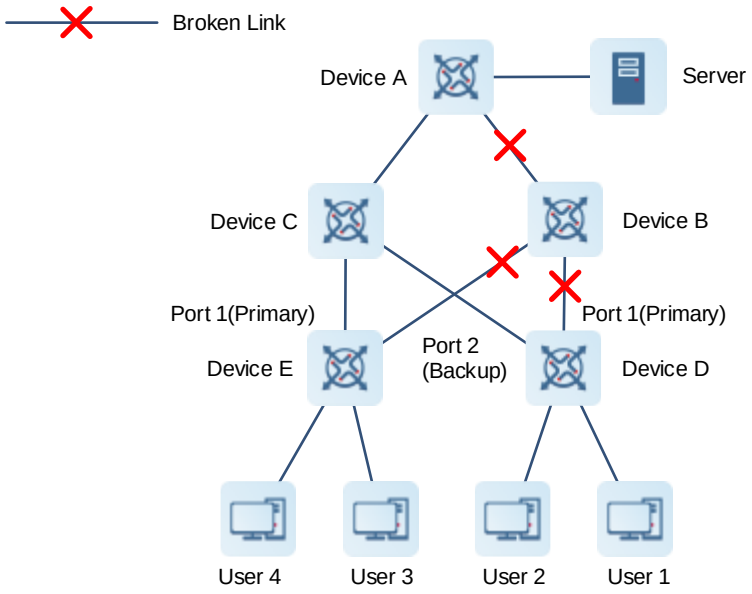


Table 1-2 Description of Link State Tracking Configuration Items

Configuration Item	Description
Link state tracking function	If the link state tracking function is not enabled, the state of a corresponding uplink port cannot be detected and packet forwarding switching cannot be implemented in time.
Downlink up delay function for a link state tracking group	By enabling the downlink up delay function, you can avoid frequent downlink switching caused by uplink flapping in a link state tracking group. That is, when the uplink becomes up, the downlink becomes up after a delay.
Adding a port to a link state tracking group	If a port is not added to a link state tracking group, the state of a corresponding uplink port cannot be detected and packet forwarding switching cannot be implemented in time.

1.2 Configuration Task Summary

REUP configuration includes the following tasks:

(1) Configuring Basic Features of REUP

(2) (Optional) Configuring the Preemption Function of REUP

(3) Configuring the MAC Address Update Function

(4) (Optional) Configuring VLAN Load Balancing

(5) Configuring Link State Tracking

1.3 Configuring Basic Features of REUP

1.3.1 Overview

When a link is faulty, another normal link switches to the forward state immediately to forward packets.

1.3.2 Restrictions and Guidelines

- One port belongs to only one REUP pair. Each active link has only one backup link. A backup link can be used as the backup link of only one active link. The active and backup links must use different ports.
- REUP supports Layer 2 physical ports and link aggregations, but does not support member ports of a link aggregation .
- The active and backup ports may be of different types and have different rates. For example, an AP can be used as the active port and a physical port can be configured as the backup port.
- Ports configured with REUP are not involved in STP calculation.
- Each device can be configured with a maximum of 16 REUP pairs.
- Ports successfully configured with REUP cannot be configured as Layer 3 interfaces or added to a link aggregation.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 Ethernet interface configuration mode.

interface *interface-type interface-number*

(4) Configure a REUP dual-link backup port.

flexlink pairs interface *interface-type interface-number*

No REUP dual-link backup port is configured by default.

1.4 Configuring the Preemption Function of REUP

1.4.1 Overview

- If the active link needs to always forward packets or the link for forwarding packets needs to be determined based on the link bandwidth, the preemption mode and delay time must be configured.
- You can determine the link to be used by configuring the preemption mode. If the preemption mode is set to bandwidth mode, a link with a greater bandwidth is used. The value of the bandwidth attribute of a link aggregation is the actual bandwidth of the link aggregation, equal to the number of link up member ports of the link aggregation multiplied by the speed attribute value of a member port. You can also set the preemption mode to forced mode to select a stable and reliable link forcibly.
- To avoid frequent active/backup link switching caused by faults, REUP provides a preemption delay function. After the two links recover, link switching is delayed (**35** seconds by default) for the faulty link to become

stable. A smaller delay means more frequent preemption switching after the faulty link recovers. When STP is enabled on an uplink, the preemption delay time of REUP should be greater than 35 seconds.

- The preemption mode can be set to three modes: forced, bandwidth, and off. In the bandwidth mode, a port with a greater bandwidth is selected to transmit data; in the forced mode, the active port is selected to transmit data; in the off mode, no preemption is performed.

1.4.2 Restrictions and Guidelines

REUP dual-link backup must be configured.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 Ethernet interface configuration mode.

interface *interface-type interface-number*

(4) Configure the preemption mode of REUP.

flexlink pairs interface *interface-type interface-number* **preemption mode** { **bandwidth** | **forced** | **off** }

The default link preemption mode of REUP is off mode.

(5) Configure the preemption delay time of REUP.

flexlink pairs interface *interface-type interface-number* **preemption delay** *interval*

The default preemption delay time of REUP is **35** seconds.

1.5 Configuring the MAC Address Update Function

1.5.1 Overview

- The MAC address update function can rapidly delete and update MAC address information on interfaces during link switching to make packet convergence faster.
- To reduce data loss in downstream data stream during link switching, you need to enable the function of sending MAC address update messages on a device that performs switching.
- During link switching, REUP sends a certain number of MAC address update packets to an uplink device per second to recover the downstream data transmission of the uplink device.
- When dual-link backup switching occurs, downstream data streams may be lost because the MAC address table of the uplink device is not updated in time. In order to reduce loss of L2 data streams, you need to update the MAC address table of the uplink device. In this case, you need to enable the function of receiving MAC address update messages on the uplink device.
- To reduce massive flooding caused by MAC address update, which may affect normal data transmission of the device, you can configure a MAC address update group. Only after all ports in a switching path are added to the same MAC address update group, can transmission of downstream data be rapidly recovered.
- For a VLAN in which the function of processing MAC address update messages is disabled, MAC address update packets can be used to recover the downlink transmission of uplink devices; however, the

convergence performance will be decreased in the case of link faults.

- After the function of sending MAC address update messages is enabled, MAC address update messages are sent to uplink devices during link switching.

1.5.2 Restrictions and Guidelines

- REUP dual-link backup must be configured.
- A maximum of eight address update groups can be configured on each device. Each address update group can contain a maximum of eight member ports and a port can belong to multiple address update groups.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the function of sending MAC address update private multicast messages.

mac-addr-table move update transit

No MAC address update private multicast message is sent by default.

(4) (Optional) Configure the maximum rate for sending MAC address update broadcast packets.

mac-addr-table move update max-update-rate *packet-per-second*

The default maximum rate for sending MAC address update broadcast packets is **150** packets/second.

(5) On an uplink device, enable the function of receiving MAC address update private multicast messages.

mac-addr-table move update receive

No MAC address update private multicast message is received by default.

(6) (Optional) On an uplink device, configure the VLAN ID range, in which MAC address update private multicast messages are received.

mac-addr-table move update receive vlan *vlan-range*

The MAC address update private multicast messages in all the VLAN ID ranges are received by default.

(7) Enter the interface configuration mode.

interface *interface-type interface-number*

(8) Configure the MAC address update group ID for the interface.

mac-addr-table update group [*group-id*]

No MAC address update group is configured by default.

(9) (Optional) Configure the ID of a VLAN, in which a port sends MAC address update private multicast messages.

mac-addr-table move update transit vlan *vlan-id*

By default, the ID of the VLAN, in which a port sends MAC address update private multicast messages is the local default VLAN of the port.

1.6 Configuring VLAN Load Balancing

1.6.1 Overview

- This function is used to maximize the utilization of link bandwidth.
- You can modify the mapping between instances and VLANs by using the instance mapping function of MSTP.
- The VLAN load balancing function allows the two ports of a REUP pair to forward data packets of mutually exclusive VLANs so as to make full use of the link bandwidth. That is, it allows some VLAN data packets to be forwarded through the backup link, but the active link does not forward these VLAN data packets. In this way, the active link and backup link forward different VLAN data packets and achieve load sharing.
- The VLAN load balancing of REUP relies on the mappings between VLANs and MSTP instances, and the instance mappings are controlled by the MSTP module in a unified manner. For details about how to configure the instances, see the description in "Configuring MSTP."

1.6.2 Restrictions and Guidelines

- REUP dual-link backup must be configured.
- VLAN load balancing cannot be configured on an access port but can share a port with STP.
- For ports successfully configured with VLAN load balancing, you cannot modify the port attributes but can modify the VLAN attributes of the ports.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a group of VLANs.

vlan { *vlan-id* | **range** *vlan-range* }

Only VLAN 1 exists by default.

(4) Return to the global configuration mode.

exit

(5) Enter the MST configuration mode.

spanning-tree mst configuration

(6) Configure the mappings between instances and VLANs, and add the VLAN group to one MST instance.

instance *instance-id* **vlan** *vlan-range*

Only instance **0** exists and all VLANs belong to instance **0** by default.

(7) Enter the Layer 2 Ethernet interface configuration mode.

interface *interface-type interface-number*

(8) Configure the VLAN load balancing function.

flexlink pairs interface *interface-type interface-number* **prefer instance** *instance-range*

The VLAN load balancing function is not configured by default.

1.7 Configuring Link State Tracking

1.7.1 Overview

- After detecting that the uplink is disconnected, REUP forcibly disconnects the downlink so that link switching can be performed.
- You must create a link state tracking group first and then add a port to the tracking group.
- You must enable the delay function so that the downlink can be up after the delay.
- You must create a link tracking group first and then add a port to the tracking group.
- When the number of uplink ports whose link status is restored in a link state tracking group is lower than the uplink port quantity threshold, the link state tracking group is in the Down state, and the link status of all downlink ports in the link state tracking group is forcibly changed to Down. When the number of uplink ports whose link status is restored is greater than or equal to the uplink port quantity threshold, the link state tracking group is restored to the normal state and the link status of all downlink ports in the link state tracking group is restored. The default uplink port quantity threshold is **1**.

1.7.2 Restrictions and Guidelines

- REUP dual-link backup must be configured.
- For the link state tracking function, each port belongs to only one link state tracking group and each device can be configured with up to two link state tracking groups. Each link state tracking group can have 8 upstream ports and 256 downstream ports.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable a link state tracking group.

link state enable-track-group [*track-number*]

All the link state tracking groups are disabled by default.

(4) Configure the downlink recovery delay time for the link state tracking group.

link state enable-track-group *track-number* **up-delay** *interval*

The downlink of a link state tracking group recovers without delay by default.

(5) Configure the uplink port quantity threshold for triggering status switching of a link state tracking group.

link state enable-track-group *track-number* **uplink-port-threshold** port-*number*

The default uplink port quantity threshold for triggering status switching of a link state tracking group is **1**.

(6) Enter the Layer 2 Ethernet interface configuration mode.

interface *interface-type interface-number*

(7) Add a port to the link state tracking group, and specify it as the uplink port of the group.

link state trace-group *track-number* **upstream**

No port is added to a link state tracking group by default.

(8) Add a port to the link state tracking group and specify it as the downlink port of the group.

link state trace-group *track-number* **downstream**

No port is added to a link state tracking group by default.

1.8 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-3 REUP Monitoring

Command	Purpose
display interfaces [*interface-type interface-number*] **flexlink pairs** [**detail**]	Displays the REUP dual-link backup information.
display mac-addr-table update group [**detail**]	Displays the configurations of a MAC address update group.
display mac-addr-table move update	Displays the statistics on MAC address update messages sent by an REUP pair.
display link state trace-group	Displays the information about a link state tracking group.
debugging reup all	Debugs all functions of REUP.
debugging reup process	Debugs the normal running process of REUP.
debugging reup packet	Debugs MAC address update messages of REUP.
debugging reup macupdt	Debugs MAC address update packets of REUP.
debugging reup ha	Debugs hot backup.
debugging reup error	Debugs errors occurring in REUP running.
debugging reup event	Debugs received events.
debugging reup status	Debugs statistics displayed when a show command is executed.

1.9 Configuration Examples

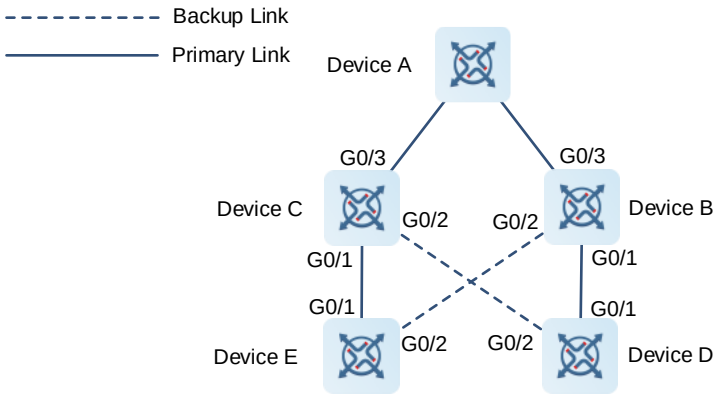
1.9.1 Configuring Basic Features of REUP

1. Requirements

An L3 device networking topology need to be built to simulate the scenario, in which there are two interconnection links between the access devices and the core devices.

2. Topology

Figure 1-6 Configuring Basic Features of REUP



3. Notes

- Configure REUP dual-link backup (GigabitEthernet 0/1 is the active port and GigabitEthernet 0/2 is the backup port) on access device D (device E).

4. Procedure

Perform the following configuration on device D.

```
DeviceD > privilege
DeviceD # config-system terminal
DeviceD (cfg)# interface gigabitethernet 0/1
DeviceD (cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceD (cfg-if-GigabitEthernet 0/1)# flexlink pairs interface gigabitethernet 0/2
DeviceD (cfg-if-GigabitEthernet 0/1)# exit
```

Perform the following configuration on device E.

```
DeviceE > privilege
DeviceE # config-system terminal
DeviceE (cfg)# interface gigabitethernet 0/1
DeviceE (cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceE (cfg-if-GigabitEthernet 0/1)# flexlink pairs interface gigabitethernet 0/2
DeviceD (cfg-if-GigabitEthernet 0/1)# exit
```

5. Verification

Run the **display interfaces flexlink pairs detail** command to verify that the configuration is correct.

```
DeviceD #display interfaces flexlink pairs detail
Device Backup Interface Pairs:
Active Interface            Backup Interface            State
--------------------------------------------------------------------------------
Gi0/1                       Gi0/2                       Active Up/Backup Standby

Interface Pair : Gi0/1, Gi0/2
Preemption Mode : off
Preemption Delay : 35 seconds
Bandwidth : Gi0/1(100000 Mbits), Gi0/2(100000 Mbits)
```

Run the **display interfaces flexlink pairs detail** command to check whether the configuration on device E is correct.

```
DeviceE #display interfaces flexlink pairs detail
Device Backup Interface Pairs:
Active Interface            Backup Interface            State
--------------------------------------------------------------------------------
Gi0/1                       Gi0/2                       Active Up/Backup Standby

Interface Pair : Gi0/1, Gi0/2
Preemption Mode : off
Preemption Delay : 35 seconds
Bandwidth : Gi0/1(100000 Mbits), Gi0/2(100000 Mbits)
```

6. Configuration Files

- Device D configuration file

```
hostname DeviceD
!
interface GigabitEthernet 0/1
 switch-port mode trunk
 flexlink pairs interface GigabitEthernet 0/2
!

end
```

- Device E configuration file

```
hostname DeviceE
!
interface GigabitEthernet 0/1
 switch-port mode trunk
 flexlink pairs interface GigabitEthernet 0/2
!

end
```

7. Common Errors

- When a port is added to an REUP pair, the port has been configured to belong to another REUP pair.

- A configured port is not a Layer 2 physical port or aggregate interface.

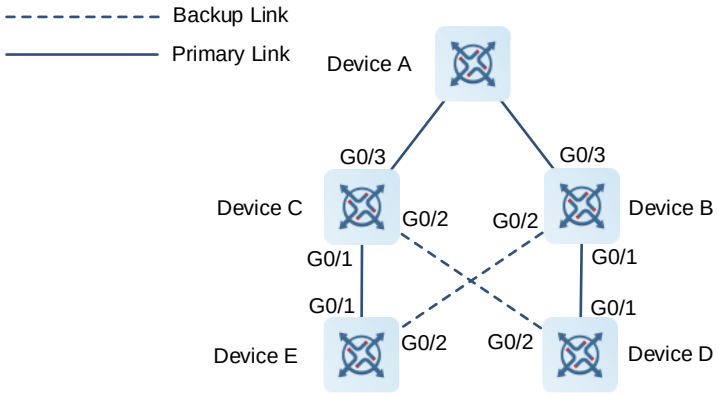
1.9.2 Configuring the MAC Address Update Function

1. Requirements

An L3 device networking topology need to be built to simulate the scenario, in which there are two interconnection links between the access devices and the core devices.

2. Topology

Figure 1-7 Configuring MAC Address Update



3. Notes

- Enable the function of sending MAC address update messages on access device D (device E).
- Enable the function of receiving MAC address update packets on device B (device C).
- Add all ports in the REUP switching path to the same MAC address update group.
- Gi0/1 and Gi0/3 of device B are the ports in the uplink switching path of device D, and Gi0/3 and Gi0/2 are the ports in the uplink switching path of device E. You can add ports Gi0/1, Gi0/2, and Gi0/3 to the same address update group. Similarly, you can perform the configuration on device C.
- Enable the function of receiving MAC address update packets on device A.
- Add all ports in the REUP switching path of device A to the same MAC address update group.

4. Procedure

Perform the following configuration on device D.

```
DeviceD > privilege
DeviceD # config-system terminal
DeviceD (cfg)# mac-addr-table move update transit
```

Perform the following configuration on device E.

```
DeviceE > privilege
DeviceE # config-system terminal
DeviceE ((cfg)# mac-addr-table move update transit
```

Perform the following configuration on device B.

```
DeviceB > privilege
```

```
DeviceB # config-system terminal
DeviceB (cfg)# mac-addr-table move update receive
DeviceB (cfg)# interface range gigabitethernet 0/1-3
DeviceB (cfg-if-list)# switch-port mode trunk
DeviceB (cfg-if-list)# mac-addr-table update trace-group 1
```

Perform the following configuration on device C.

```
DeviceC > privilege
DeviceC # config-system terminal
DeviceC (cfg)# mac-addr-table move update receive
DeviceC (cfg)# interface range gigabitethernet 0/1-3
DeviceC (cfg-if-list)# switch-port mode trunk
DeviceC (cfg-if-list)# mac-addr-table update trace-group 1
```

Perform the following configuration on device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# mac-addr-table move update receive
DeviceA(cfg)# interface range gigabitethernet 0/1-2
DeviceA(cfg-if-list)# switch-port mode trunk
DeviceA(cfg-if-list)# mac-addr-table update trace-group 1
```

5. Verification

Check the information about the MAC address update groups on device B.

```
DeviceD # display current-config | incl mac-ad
mac-addr-table move update transit
```

```
DeviceE # display current-config | incl mac-ad
mac-addr-table move update transit
```

```
DeviceB # display mac-addr-table update trace-group detail
display mac-addr-table update trace-group detailMac-address-table Update Group:1
Received mac-addr-table update message count:0
Group member            Receive Count      Last Receive Switch-ID          Receive Time
--------------------------------------------------------------------------------
Gi0/1                   0                  0000.0000.0000
Gi0/2                   0                  0000.0000.0000
Gi0/3                   0                  0000.0000.0000
```

Check the information about the MAC address update groups on device C.

```
DeviceC # display mac-addr-table update trace-group detail
Mac-address-table Update Group:1
Received mac-addr-table update message count:0
Group member            Receive Count      Last Receive Switch-ID          Receive Time
--------------------------------------------------------------------------------
Gi0/1                   0                  0000.0000.0000
```

```
Gi0/2                0                0000.0000.0000
Gi0/3                0                0000.0000.0000
```

Check the information about the MAC address update groups on device A.

```
DeviceA# display mac-addr-table update trace-group detail
Mac-address-table Update Group:1
Received mac-addr-table update message count:0
Group member          Receive Count      Last Receive Switch-ID          Receive Time
------------------------------------------------------------------------------
Gi0/1                0                0000.0000.0000
Gi0/2                0                0000.0000.0000
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 0/1
 switch-port mode trunk
 mac-addr-table update trace-group 1
!
interface GigabitEthernet 0/2
 switch-port mode trunk
 mac-addr-table update trace-group 1
!
mac-addr-table move update receive

end
```

- Device B configuration file

```
hostname DeviceB
!
interface GigabitEthernet 0/1
 switch-port mode trunk
 mac-addr-table update trace-group 1
!
interface GigabitEthernet 0/2
 flexlink mode trunk
 mac-addr-table update trace-group 1
!
interface GigabitEthernet 0/3
 switch-port mode trunk
 mac-addr-table update trace-group 1
!
mac-addr-table move update receive

end
```

- Device C configuration file

```
hostname DeviceC
!
interface GigabitEthernet 0/1
 switch-port mode trunk
 mac-addr-table update trace-group 1
!
interface GigabitEthernet 0/2
 switch-port mode trunk
 mac-addr-table update trace-group 1
!
interface GigabitEthernet 0/3
 switch-port mode trunk
 mac-addr-table update trace-group 1
!
mac-addr-table move update receive

end
```

- Device D configuration file

```
hostname DeviceD
!
mac-addr-table move update transit

end
```

- Device E configuration file

```
hostname DeviceE
!
mac-addr-table move update transit

end
```

7. Common Errors

- A configured port is not a Layer 2 physical port or aggregate interface.

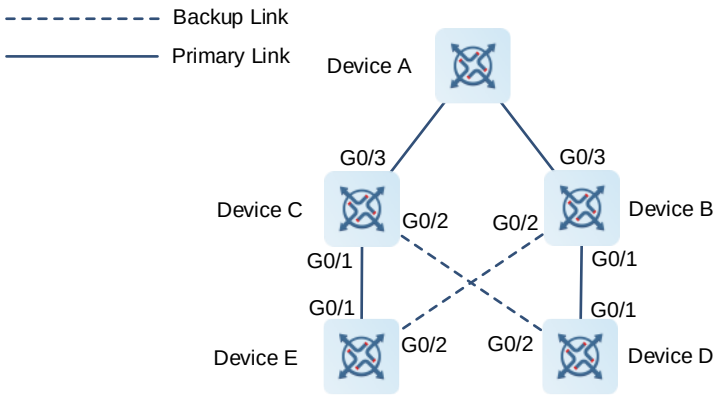
1.9.3 Configuring VLAN Load Balancing

1. Requirements

An L3 device networking topology need to be built to simulate the scenario, in which there are two interconnection links between the access devices and the core devices.

2. Topology

Figure 1-8 Configuring VLAN Loading Balancing



3. Notes

- Configure instance mappings on device D (device E) to map VLAN 1 to instance 1, VLAN 2 to instance 2, VLAN 3 to instance 3, and VLAN 4 to instance 4. For details about how to configure the mappings, see "Configuring MSTP" in *Ethernet Switching Configuration Guide*.
- Configure the VLAN load balancing function on device D (device E).

4. Procedure

Perform the following configuration on device D.

```
DeviceD > privilege
DeviceD # config-system terminal
DeviceD (cfg)# interface gigabitethernet 0/1
DeviceD (cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceD (cfg-if-GigabitEthernet 0/1)# flexlink pairs interface gigabitethernet 0/2 prefer instance 2
DeviceD (cfg-if-GigabitEthernet 0/1)# exit
```

Perform the following configuration on device E.

```
DeviceE > privilege
DeviceE # config-system terminal
DeviceE (cfg)# interface gigabitethernet 0/1
DeviceE (cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
DeviceD (cfg-if-GigabitEthernet 0/1)# flexlink pairs interface gigabitethernet 0/2 prefer instance 4
DeviceD (cfg-if-GigabitEthernet 0/1)# exit
```

5. Verification

Run the **display interfaces flexlink pairs detail** command to check the dual-link backup information configured on device D (device E).

```
DeviceD # display interfaces flexlink pairs detail
Device Backup Interface Pairs:
Active Interface            Backup Interface            State
--------------------------------------------------------------------------------
```

```
Gi0/1                    Gi0/2                    Active Up/Backup Up

Instances Preferred on Active Interface: Instance 0-1,3-64
        Mapping VLAN 1,3-4094

Instances Preferred on Backup Interface: Instance 2
        Mapping VLAN 2

Interface Pair : Gi0/1, Gi0/2
Preemption Mode : balance
Preemption Delay : 35 seconds
Bandwidth : Gi0/1(800 kbits), Gi0/2(100000 kbits)
```

Run the **display interfaces flexlink pairs detail** command to check the dual-link backup information configured on device E.

```
DeviceE # display interfaces flexlink pairs detail
Device Backup Interface Pairs:
Active Interface        Backup Interface        State
--------------------------------------------------------------------------------
Gi0/1                    Gi0/2                    Active Up/Backup Up

Instances Preferred on Active Interface: Instance 0-3,5-64
        Mapping VLAN 1-3,5-4094

Instances Preferred on Backup Interface: Instance 4
        Mapping VLAN 4

Interface Pair : Gi0/1, Gi0/2
Preemption Mode : balance
Preemption Delay : 35 seconds
Bandwidth : Gi0/1(800 kbits), Gi0/2(100000 kbits)
```

6. Configuration Files

- Device D configuration file

```
hostname DeviceD
!
interface GigabitEthernet 0/1
 switch-port mode trunk
 flexlink pairs interface gigabitethernet 0/2 prefer instance 2
!

end
```

- Device E configuration file

```
hostname DeviceE
!
interface GigabitEthernet 0/1
```

```
 switch-port mode trunk
 flexlink pairs interface gigabitethernet 0/2 prefer instance 2
!

end
```

7. Common Errors

- The mappings between VLAN IDs and instances are not configured.

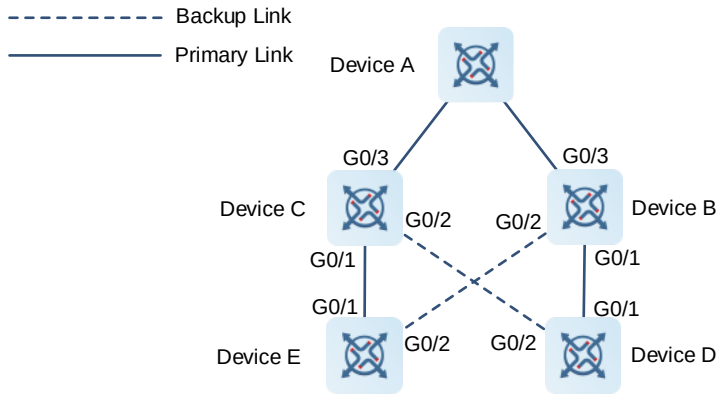
1.9.4 Configuring Link State Tracking

1. Requirements

An L3 device networking topology need to be built to simulate the scenario, in which there are two interconnection links between the access devices and the core devices.

2. Topology

Figure 1-9 Configuring Link State Tracking



3. Notes

- Create link state tracking group 1 on device B (device C).
- On device B (device C), add the ports GigabitEthernet 0/1 and GigabitEthernet 0/2 as downlink ports of the link state tracking group and add the port GigabitEthernet 0/3 as an uplink port of the link state tracking group.

4. Procedure

Perform the following configuration on device B.

```
DeviceB > privilege
DeviceB # config-system terminal
DeviceB (cfg)# link state enable-track-group 1
DeviceB (cfg)# interface GigabitEthernet 0/1
DeviceB (cfg-if-GigabitEthernet 0/1)# link state trace-group 1 downstream
DeviceB (cfg-if-GigabitEthernet 0/1)# exit
DeviceB (cfg)# interface GigabitEthernet 0/2
DeviceB (cfg-if-GigabitEthernet 0/2)# link state trace-group 1 downstream
```

```
DeviceB (cfg-if-GigabitEthernet 0/2)# exit
DeviceB (cfg)# interface GigabitEthernet 0/3
DeviceB (cfg-if-GigabitEthernet 0/3)# link state trace-group 1 upstream
```

Perform the following configuration on device C.

```
DeviceC > privilege
DeviceC # config-system terminal
DeviceC (cfg)# link state enable-track-group 1
DeviceC (cfg)# interface GigabitEthernet 0/1
DeviceC (cfg-if-GigabitEthernet 0/1)# link state trace-group 1 downstream
DeviceC (cfg-if-GigabitEthernet 0/1)# exit
DeviceC (cfg)# interface GigabitEthernet 0/2
DeviceC (cfg-if-GigabitEthernet 0/2)# link state trace-group 1 downstream
DeviceC (cfg-if-GigabitEthernet 0/2)# exit
DeviceC (cfg)# interface GigabitEthernet 0/3
DeviceC (cfg-if-GigabitEthernet 0/3)# link state trace-group 1 upstream
```

5. Verification

Run the **display link state trace-group** command to check whether the configuration on device B is correct.

```
DeviceB # display link state trace-group
Link State Group:1    Status: enabled, Down
Upstream Interfaces :Gi0/3(Down)
Downstream Interfaces : Gi0/2(Down)

Link State Group:2    Status: Disabled, Down
Upstream Interfaces :
Downstream Interfaces :

(Up):Interface up     (Down):Interface Down     (Dis):Interface disabled
```

Run the **display link state trace-group** command to check whether the configuration on device C is correct.

```
DeviceC # display link state trace-group
Link State Group:1    Status: enabled, Down
Upstream Interfaces :Gi0/3(Down)
Downstream Interfaces : Gi0/2(Down)

Link State Group:2    Status: Disabled, Down
Upstream Interfaces :
Downstream Interfaces :

(Up):Interface up     (Down):Interface Down     (Dis):Interface disabled
```

6. Configuration Files

- Device B configuration file

```
hostname DeviceB
!
link state enable-track-group 1
```

```
interface GigabitEthernet 0/1
 link state trace-group 1 downstream
!
interface GigabitEthernet 0/2
 link state trace-group 1 downstream
!
interface GigabitEthernet 0/3
 link state trace-group 1 upstream
!

end
```

- Device C configuration file

```
hostname DeviceC
!
link state enable-track-group 1
interface GigabitEthernet 0/1
 link state trace-group 1 downstream
!
interface GigabitEthernet 0/2
 link state trace-group 1 downstream
!
interface GigabitEthernet 0/3
 link state trace-group 1 upstream
!

end
```

7. Common Errors

- Ports are added to a link state tracking group before the link state tracking group is enabled.

Contents

1 Configuring RLDP 1

1.1 Introduction 1

1.1.1 Basic Concepts 1

1.1.2 Detection Principles 3

1.1.3 RLDP Detection 4

1.2 Configuring Basic Features of RLDP 5

1.2.1 Overview 5

1.2.2 Restrictions and Guidelines 5

1.2.3 Procedure 6

1.3 Monitoring 6

1.4 Configuration Examples 7

1.4.1 Configuring Basic Features of RLDP 7

1.4.2 Configuring Unidirectional Link Detection 10

1.4.3 Configuring Bidirectional Link Detection 11

1 Configuring RLDP

1.1 Introduction

The Rapid Link Detection Protocol (RLDP) is an Ethernet link failure detection protocol, which is used to rapidly detect unidirectional link failures, bidirectional link failures, and downlink loop failures. When a failure is found, RLDP automatically shuts down relevant ports or asks users to manually shut down the ports according to the configured failure handling methods, to avoid wrong forwarding of traffic or Ethernet L2 loops.

1.1.1 Basic Concepts

1. Unidirectional Link Failure

A unidirectional link failure occurs when one end of the receiving or transmission link in an optical fiber is disconnected or not connected, one line in a twisted-pair cable is disconnected, or unidirectional disconnection occurs on an intermediate device between two devices. A unidirectional link failure can lead to wrong forwarding of traffic or the failure of a loop protection protocol such as Spanning Tree Protocol (STP).

2. Bidirectional Link Failure

A bidirectional link failure occurs when two optical fibers are disconnected, two lines in a twisted-pair cable are disconnected, or bidirectional disconnection occurs on an intermediate device between two devices. A bidirectional link failure can lead to the wrong forwarding of traffic.

3. Loop Failure

A loop failure occurs when a downlink port of a device is wrongly connected to another device and a loop forms. A loop failure can result in a broadcast storm.

4. Protocol Packets

RLDP defines the following three types of packets:

- Probe packets: L2 multicast packets used for neighbor negotiation, and unidirectional or bidirectional link detection. The default encapsulation format of the packets is of the Subnetwork Access Protocol (SNAP) type, which automatically changes to EthernetII if a neighbor sends EthernetII packets.
- Echo packets: L2 unicast packets in response to probe packets and used for unidirectional or bidirectional link detection. The default encapsulation format is of the SNAP type, which automatically changes to EthernetII if a neighbor sends EthernetII packets.
- Loop packets: L2 multicast packets for downlink loop detection. They can only be received by senders. The default encapsulation format of the packets is of the SNAP type.

5. Detection Interval

A detection interval and the maximum detection count can be configured for RLDP. A detection interval determines the period of sending probe packets and loop packets. When a device receives a probe packet, it replies with an echo packet immediately. A detection interval and the maximum detection count determine the maximum detection time (Detection interval × Maximum detection count + 1) for unidirectional or bidirectional link detection. If neither a probe nor an echo packet from a neighbor is correctly received within the maximum

detection time, the handling of unidirectional or bidirectional failures is triggered. The timers involved are described in Table 1-1.

Table 1-1 Description of RLDP Timers

Timer Name	Timer Description
Send timer	After the link detection function is enabled, the send timer is started and the probe packets or loop packets are sent regularly. The time interval of the send timer is the configured detection interval.
Timeout timer	● If a unidirectional or bidirectional link failure is detected, the timeout timer is started after the probe packet is sent, and the timeout time is the maximum detection time (Detection interval × Maximum detection count + 1). ● In unidirectional link detection, when the detection timeout time expires and a port can only receive the probe packet or response packet from the neighbor, it is considered that a unidirectional link failure occurs. ● In bidirectional link detection, when the detection timeout time expires and a port fails to receive the probe packet and response packet from the neighbor, it is considered that a bidirectional link failure occurs.

6. Neighbor Negotiation

When configured with unidirectional or bidirectional link detection, a port can learn a peer device as its neighbor. One port can learn only one neighbor, which is changeable. If the negotiation function is enabled, unidirectional or bidirectional link detection starts after a port finds a neighbor through negotiation. Negotiation is considered successful if the port receives a probe packet from the neighbor. However, if RLDP is enabled on a port after a failure already occurs, the port fails to learn a neighbor and detection fails to be started. In this case, you are advised to rectify the link failure first.

7. Associated Actions

The associated actions and recovery methods after a RLDP port failure are described in Table 1-2.

Table 1-2 Description of RLDP Port Failure Handling

Type	Associated Action/ Recovery Method	Description
Associated actions after a failure	warning	Only the relevant system log is printed to indicate the failed port and the failure type.
	Shutdown SVI	A system log is printed. If the failed port is a Layer 2 Ethernet interface or member port of a Layer 2 link aggregation, a switch virtual interface (SVI) is queried based on the access VLAN or native VLAN of the port, and then the SVI is shut down.
	Port errdisable	A system log is printed, the failed port is set in errdisable state, and the port enters the linkdown state physically.
	Block	A system log is printed, the forwarding state of the failed port is set to blocked, and the port does not forward received packets.

Type	Associated Action/ Recovery Method	Description
Recovery methods after a failure	Manual reset	Manually reset all the failed ports to the initialized state to restart link detection.
	Manual or automatic errdisable recovery	Recover all the failed ports to the initialized state manually or regularly (**30** seconds by default and configurable) and restart link detection.
	Automatic recovery	Under unidirectional or bidirectional link detection, if the handling method specified for failed ports does not enable the ports to enter the errdisable state, the ports can automatically recover to the initialized state based on probe packets exchanged with the neighbor, and link detection is restarted.
	Manual or automatic execution of the **rldp error-recover interval** command	All the failed ports recover to the initialized state regularly (configurable) and link detection is restarted.

8. RLDP Port State

RLDP ports can be in two states:

- Normal: Indicates the state of a port, on which link detection is enabled.
- Error: Indicates the state of a port, on which a unidirectional or bidirectional link failure or a loop failure is detected.

1.1.2 Detection Principles

A typical Ethernet link detection mechanism detects physical link connectivity through auto-negotiation at the physical layer. However, this detection mechanism has limitations. In some cases, devices are connected and work normally at the physical layer but L2 link communication fails or is abnormal. RLDP recognizes a neighbor device and detects whether a link failure exists by exchanging probe packets, echo packets, or loop packets with the neighbor device. Figure 1-1 shows the protocol states.

Figure 1-1 RLDP Protocol State Diagram

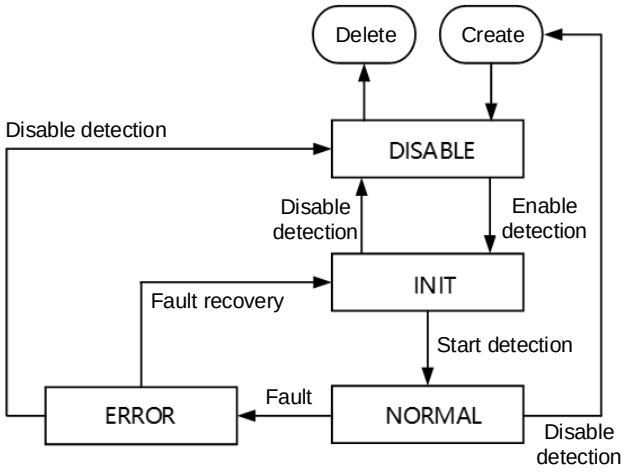


Figure 1-1 shows the detection logic of RLDP, which can be mainly in four states. The four states are described in Table 1-3.

Table 1-3 Description of RLDP States

State	Description
DISABLE	After the global RLDP function is enabled, all ports enter the disable state. Now, unidirectional link detection, bidirectional link detection, or downlink loop detection is not enabled on the ports. If the RLDP is disabled in any state, a port enters this state.
INIT	It refers to the initialized state after RLDP is just configured and enabled or a failure is rectified. Now, unidirectional link detection, bidirectional link detection, or downlink loop detection is enabled on ports, and the corresponding link failure handling methods are specified, or the links can recover from unidirectional failures, bidirectional failures, or loop failures automatically.
NORMAL	It refers to the state after detection is enabled for RLDP. For downlink loop detection, RLDP enters the state directly after enabled. For unidirectional or bidirectional link detection, RLDP enters the state directly or enters the state only after negotiation with a specified neighbor succeeds.
ERROR	It refers to the state after RLDP detects a failure. In this case, a unidirectional link failure or bidirectional link failure is detected on a port. The failure will be handled according to the specified failure handling method.

1.1.3 RLDP Detection

RLDP provides unidirectional link detection, bidirectional link detection, and downlink loop detection.

1. Unidirectional Link Detection

After unidirectional link detection is enabled, a port sends probe packets and receives echo packets from a neighbor regularly. It also receives probe packets from the neighbor and replies with echo packets to the

neighbor. If the port receives only probe packets but no echo packets, or none of them from the neighbor within the maximum detection time, handling of a unidirectional failure is triggered and the detection stops.

2. Bidirectional Link Detection

After bidirectional link detection is enabled, a port sends probe packets and receives echo packets from a neighbor regularly. It also receives probe packets from the neighbor and replies with echo packets to the neighbor. If the port receives neither probe packets nor echo packets from the neighbor within the maximum detection time, handling of a bidirectional failure is triggered and the detection stops.

3. Downlink Loop Detection

After downlink loop detection is enabled, a port sends loop packets regularly. After the same port or different ports on a device receive loop packets, a loop failure is triggered if any of the following cases occurs: The transmission and receiving ports of the packets are the same Layer 3 Ethernet interfaces or member ports of a Layer 3 link aggregation; the transmission and receiving ports of packets are Layer 2 interfaces or member ports of a Layer 2 link aggregation, the default VLANs of the ports are the same, and forwarding states of the ports are forward. The failure is handled according to the corresponding method and the detection stops.

1.2 Configuring Basic Features of RLDP

1.2.1 Overview

- You can enable RLDP unidirectional link detection, bidirectional link detection, or downlink loop detection to discover unidirectional, bidirectional, or loop failures.
- After global RLDP detection is enabled, the RLDP detection is started on each port normally.
- After the function is enabled in global configuration mode, RLDP detection on a port is started after successful neighbor negotiation.
- If loop detection latency interval is configured in global configuration mode, the downlink port loop cannot be detected in this period of time.
- In interface configuration mode, you can configure unidirectional RLDP detection, bidirectional RLDP detection, or downlink loop detection, and specify a failure handling method.
- You can restore all the failed ports and restart detection.

1.2.2 Restrictions and Guidelines

- Loop detection configured on an AP member port takes effect on all the member ports of the AP. Unidirectional link detection and bidirectional link detection take effect on an AP member port directly.
- The loop detection configuration of a physical port added to an AP must be the same as that of the other member ports of the AP. There are two cases. (1) If loop detection is not configured on the new member port but on the existing member ports, the new port adopts the configuration and detection results of the existing ports. (2) If the new member port and the existing member ports have different loop detection configurations, the new port adopts the configuration and detection results of the existing ports.
- When configuring the RLDP on an AP member port, you can configure only the "shutdown-port" and "warning" handling methods. If the failure handling method is not set to "shutdown-port" or "warning", it is modified to "shutdown-port" and takes effect.
- After the "shutdown-port" handling method is configured on a port, the port cannot resume RLDP detection

in the case of a failure. After confirming that the failure is rectified, you can run the **rldp reset** or **errdisable recovery** command to restore the port and restart detection. For the description and configuration of **errdisable recovery**, see *Configuring Ethernet Interface*.

1.2.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the global RLDP detection function.

rldp enable

The global RLDP detection function is disabled by default.

(4) (Optional) Configure global neighbor negotiation.

rldp neighbor-negotiation

The neighbor negotiation function is disabled by default.

(5) (Optional) Configure the global RLDP detection interval.

rldp detect-interval *interval*

The default interval for sending RLDP packets by a port is **3** seconds.

(6) (Optional) Configure a time interval for RLDP to recover failed ports.

rldp error-recover interval *interval*

No time interval for RLDP to recover failed ports regularly is configured by default.

(7) (Optional) Configure the maximum detection count of RLDP.

rldp detect-max *max-detect-num*ber

The default maximum detection count of RLDP is **2**.

(8) (Optional) Configure the global loop detection latency interval.

rldp detect-latency *interval*

The default latency interval for global RLDP loop detection is **0** seconds.

(9) Enter the Layer 2 Ethernet interface configuration mode.

interface *interface-type interface-number*

(10) Configure RLDP detection on the port.

rldp port { **bidirection-detect** | **loop-detect** | **unidirection-detect** } { **block** | **shutdown-port** | **shutdown-svi** | **warning** }

RLDP detection is not configured on a port by default.

1.3 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-4 RLDP Monitoring

Command	Purpose
display rldp [**interface** *interface-type interface-number*]	Displays the RLDP running state.

1.4 Configuration Examples

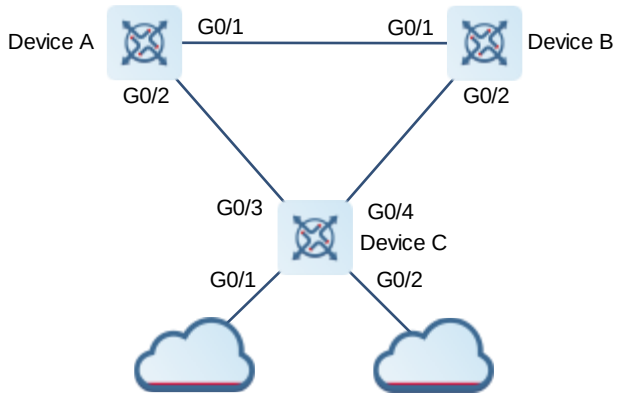
1.4.1 Configuring Basic Features of RLDP

1. Requirements

The aggregation and access devices form a ring topology. STP is enabled on all devices to prevent loops and provide redundancy protection. To avoid a unidirectional or bidirectional link failure resulting in an STP failure, RLDP unidirectional link detection and bidirectional link detection are enabled between aggregation devices as well as between the aggregation devices and the access device. To avoid loops due to wrong connection in the downlink direction of the aggregation devices, RLDP downlink loop detection is enabled on the downlink ports of the aggregation devices and the access device. To avoid loops due to wrong connection in the downlink direction of the access device, RLDP downlink loop detection is enabled on all the downlink ports of the access device.

2. Topology

Figure 1-2 Configuring Basic Feature of RLDP



3. Notes

- Devices A and B are aggregation devices, and device C is an access device. User devices can be connected to downlink ports of device C. Devices A, B, and C are structured in a ring topology, and STP is enabled on each of them. For STP configuration, refer to the relevant configuration guide.
- Enable RLDP on device A, enable unidirectional and bidirectional link detection on the two ports, and enable loop detection on the downlink port.
- Enable RLDP on device B, enable unidirectional and bidirectional link detection on the two ports, and enable loop detection on the downlink port.
- Enable RLDP on device C, enable unidirectional and bidirectional link detection on the two uplink ports, and

enable loop detection on the two downlink ports.

4. Procedure

Perform the following configuration on device A.

```
Device A# config-system terminal
Device A(cfg)# rldp enable
Device A(cfg)# interface GigabitEthernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# rldp port unidirection-detect shutdown-port
Device A(cfg-if-GigabitEthernet 0/1)# rldp port bidirection-detect shutdown-port
Device A(cfg-if-GigabitEthernet 0/1)# exit
Device A(cfg)# interface GigabitEthernet 0/2
Device A(cfg-if-GigabitEthernet 0/1)# rldp port unidirection-detect shutdown-port
Device A(cfg-if-GigabitEthernet 0/1)# rldp port bidirection-detect shutdown-port
Device A(cfg-if-GigabitEthernet 0/1)# rldp port loop-detect shutdown-port
```

Perform the following configuration on device B.

```
Device B# config-system terminal
Device B(cfg)# rldp enable
Device B(cfg)# interface GigabitEthernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# rldp port unidirection-detect shutdown-port
Device B(cfg-if-GigabitEthernet 0/1)# rldp port bidirection-detect shutdown-port
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface GigabitEthernet 0/2
Device B(cfg-if-GigabitEthernet 0/1)# rldp port unidirection-detect shutdown-port
Device B(cfg-if-GigabitEthernet 0/1)# rldp port bidirection-detect shutdown-port
Device B(cfg-if-GigabitEthernet 0/1)# rldp port loop-detect shutdown-port
```

Perform the following configuration on device C.

```
Device C# config-system terminal
Device C(cfg)# rldp enable
Device C(cfg)# interface GigabitEthernet 0/3
Device C(cfg-if-GigabitEthernet 0/3)# rldp port unidirection-detect shutdown-port
Device C(cfg-if-GigabitEthernet 0/3)# rldp port bidirection-detect shutdown-port
Device C(cfg-if-GigabitEthernet 0/3)# exit
Device C(cfg)# interface GigabitEthernet 0/4
Device C(cfg-if-GigabitEthernet 0/4)# rldp port unidirection-detect shutdown-port
Device C(cfg-if-GigabitEthernet 0/4)# rldp port bidirection-detect shutdown-port
Device C(cfg-if-GigabitEthernet 0/4)# exit
Device C(cfg)# interface GigabitEthernet 0/1
Device C(cfg-if-GigabitEthernet 0/1)# rldp port loop-detect shutdown-port
Device C(cfg-if-GigabitEthernet 0/1)# exit
Device C(cfg)# interface GigabitEthernet 0/2
Device C(cfg-if-GigabitEthernet 0/2)# rldp port loop-detect shutdown-port
```

5. Verification

Check the RLDP state information on devices, A, B, and C. Device A is taken as an example. Run the **display rldp** command to check whether the configuration is correct.

```
Device A# display rldp
rldp state                    : enable
rldp hello interval: 3
rldp max hello           : 2
rldp local bridge    : 00d0.f822.33aa
-----------------------------------
Interface GigabitEthernet 0/1
port state             : normal
neighbor bridge : 00d0.f800.51b1
neighbor port       : GigabitEthernet 2/0/1
unidirection detect information:
        action: shutdown-port
        state : normal
bidirection detect information:
        action: shutdown-port
        state : normal

Interface GigabitEthernet 0/2
port state             : normal
neighbor bridge : 00d0.f800.41b0
neighbor port       : GigabitEthernet 0/3
unidirection detect information:
        action: shutdown-port
        state : normal
bidirection detect information:
        action: shutdown-port
        state : normal
loop detect information:
        action: shutdown-port
        state : normal
```

6. Common Errors

- RLDP is enabled together with private multicast address authentication or TPP.
- Neighbor negotiation is not enabled when unidirectional or bidirectional link detection is configured. RLDP must be enabled on a neighbor device in global or interface configuration mode; otherwise a unidirectional or bidirectional failure is detected.
- If neighbor negotiation prior to RLDP detection is specified when unidirectional or bidirectional link detection is configured, normal detection fails because no neighbor is learned in the case of an existing link failure. In this situation, you are advised to rectify the link failure first.
- You are not advised to specify Shutdown SVI as the failure handling method on a Layer 3 interface.
- You are not advised to specify Block as the failure handling method for a port, on which a loop protection protocol such as STP is enabled.

1.4.2 Configuring Unidirectional Link Detection

1. Requirements

As shown in Figure 1-3, device A is connected to device B through an optical fiber. The two lines are the Tx and Rx lines of the optical fiber. Unidirectional link detection is enabled on devices A and B. If a fault occurs on either the Tx end of device A or Rx end of device B or on either the Rx end of device A or Tx end of device B, RLDP can detect the unidirectional failure and handle it. If the failure is eliminated, the administrator can manually restore the RLDP state on devices A and B and restart detection.

2. Topology

Figure 1-3 Configuring Unidirectional Link Detection



3. Notes

- Enable global RLDP.
- Configure unidirectional link detection of RLDP on ports and define a method for handling unidirectional failures.

4. Procedure

Perform the following configuration on device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# rldp enable
DeviceA(cfg)# interface GigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# rldp port unidirection-detect shutdown-port
```

Perform the following configuration on device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# rldp enable
DeviceB(cfg)# interface GigabitEthernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# rldp port unidirection-detect shutdown-port
```

5. Verification

Check the RLDP state information on device A.

```
DeviceA> privilege
DeviceA# display rldp
rldp state              : enable
rldp hello interval: 3
rldp max hello          : 2
rldp local bridge    : 00d0.f822.33aa
```

```
-----------------------------------
Interface GigabitEthernet 0/1
port state          : normal
neighbor bridge : 00d0.f800.51b1
neighbor port     : GigabitEthernet 0/1
unidirection detect information:
        action: shutdown-port
        state : normal
```

Check the RLDP state information on device B.

```
DeviceB> privilege
DeviceB# display rldp
rldp state               : enable
rldp hello interval: 3
rldp max hello         : 2
rldp local bridge    : 00d0.f800.51b1
-----------------------------------
Interface GigabitEthernet 0/1
port state          : normal
neighbor bridge : 00d0.f822.33aa
neighbor port     : GigabitEthernet 0/1
unidirection detect information:
        action: shutdown-port
        state : normal
```

6. Common Errors

- Global RLDP is not enabled at the local end or peer end.
- The RLDP detection method configured at the local end is inconsistent with that at the peer end.
- RLDP negotiation is not enabled, so a link failure is judged by mistake when unidirectional link detection of RLDP is configured at a single end.

1.4.3 Configuring Bidirectional Link Detection

1. Requirements

As shown in Figure 1-4, device A is connected to device B through an optical fiber. The two lines are the Tx and Rx lines of the optical fiber. Bidirectional link detection is enabled on devices A and B. If a fault occurs on the Tx end of device A and Rx end of device B or on the Rx end of device A and Tx end of device B, RLDP can detect a bidirectional failure and handle it. If the failure is eliminated, the administrator can manually restore the RLDP state on devices A and B and restart detection.

2. Topology

Figure 1-4 Configuring Bidirectional Link Detection



3. Notes

- Enable global RLDP.
- Configure bidirectional link detection of RLDP on ports and define a method for handling bidirectional failures.

4. Procedure

Perform the following configuration on device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# rldp enable
DeviceA(cfg)# interface GigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# rldp port bidirection-detect shutdown-port
```

Perform the following configuration on device B.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# rldp enable
DeviceB(cfg)# interface GigabitEthernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# rldp port bidirection-detect shutdown-port
```

5. Verification

Check the RLDP state information on device A.

```
DeviceA> privilege
DeviceA# display rldp
rldp state               : enable
rldp hello interval: 3
rldp max hello         : 2
rldp local bridge    : 00d0.f822.33aa
-----------------------------------
Interface GigabitEthernet 0/1
port state           : normal
neighbor bridge : 00d0.f800.51b1
neighbor port      : GigabitEthernet 0/1
bidirection detect information:
        action: shutdown-port
        state : normal
```

Check the RLDP state information on device B.

```
DeviceB> privilege
DeviceB# display rldp
rldp state               : enable
rldp hello interval: 3
rldp max hello         : 2
rldp local bridge    : 00d0.f800.51b1
```

```
-----------------------------------
Interface GigabitEthernet 0/1
port state          : normal
neighbor bridge : 00d0.f822.33aa
neighbor port     : GigabitEthernet 0/1
bidirection detect information:
        action: shutdown-port
        state : normal
```

6. Common Errors

- Global RLDP is not enabled at the local end or peer end.
- The RLDP detection method configured at the local end is inconsistent with that at the peer end.
- If bidirectional link detection needs to be configured but RLDP negotiation is not enabled, a link failure will be detected at one end.

Contents

1 Configuring DLDP 1

1.1 Introduction 1

1.1.1 Basic Concepts 1

1.1.2 DLDP Detection 2

1.1.3 MAC Address Binding 2

1.2 Configuring DLDP Detection 2

1.2.1 Overview 2

1.2.2 Restrictions and Guidelines 2

1.2.3 Procedure 3

1.3 Monitoring 3

1.4 Configuration Examples 4

1.4.1 Configuring DLDP Detection 4

1 Configuring DLDP

The Data Link Detection Protocol (DLDP) achieves rapid detection of Ethernet link failures. A typical Ethernet link detection mechanism detects physical link connectivity through auto-negotiation at the physical layer. However, such a mechanism has limitations in L3 data connectivity detection, for example, the physical connection is normal but L3 data communication is abnormal. DLDP provides reliable L3 link detection information. After detecting a faulty link, DLDP disables the logical function of L3 ports to realize fast L3 protocol convergence.

1.1 Introduction

1.1.1 Basic Concepts

1. Detection Mode

Active mode and passive mode are two DLDP detection modes.

- Active mode (default): In this mode, Internet Control Message Protocol (ICMP) detection packets are sent actively.
- Passive mode: In this mode, ICMP detection packets are received passively. Namely, DLDP returns an ICMP reply packet upon receiving an ICMP echo packet, instead of actively sending ICMP echo packets to the peer end. It judges whether an ICMP echo packet is received within a specified period to determine whether the link to the peer port is faulty. This mode not only implements the function of detecting link connectivity for the both devices, but also saves the bandwidth resources and device CPU resources.

2. Inter-Network Segment Detection

If DLDP needs to detect reachability to an IP address in a non-directly connected network segment, you need to configure the next-hop IP address for the local port so that DLDP can obtain the next-hop MAC address through an Address Resolution Protocol (ARP) packet, encapsulate an ICMP packet correctly, and send it out. In this situation, however, you need to avoid receiving an ICMP reply packet from another link; otherwise, DLDP misjudges that the port does not receive an ICMP reply.

3. Detection Time

When a network device does not receive a reply packet from the peer end within the period of the detection interval multiplied by the number of retransmission times, the device determines that an L link failure occurs and actively shuts down the logical function of its L3 port (despite the physical link reachable). Once the L3 link connectivity recovers, the device restores the logical function of the L3 port.

- Detection interval: Indicates the interval at which DLDP detection packets (ICMP echo) are transmitted.
- Retransmission times: Indicates the maximum number of times that DLDP detection packets can be retransmitted after a DLDP detection failure occurs.

4. Recovery Times

A detected link may be unstable, for example, the link can be pinged through only intermittently. The results of single DLDP detection are that the link is up/ down multiple times, which further destabilizes the ring network.

The term recovery times indicates the number of times that DLDP needs to receive reply packets consecutively before a link switches from the down state to up state. The default number of recovery times is 3, that is, a link can be pinged successfully three times before it is set to up. The recovery times configuration reduces link detection sensitivity but increases stability. Related parameters can be adjusted according to the network condition.

1.1.2 DLDP Detection

When detecting that an L3 link is abnormal, DLDP actively shuts down the Layer 3 interfaces. After the DLDP function is enabled, DLDP sends an ARP packet to obtain the MAC address and outbound port of the detected device or the next-hop device reachable to the detected device. Then DLDP periodically sends IPv4 ICMP echo packets to the MAC address and outbound port to detect link connectivity. If DLDP does not receive an IPv4 ICMP reply packet from the detected device within a specified period, DLDP determines that the link is abnormal and sets the L3 port to down. You can configure the next-hop IP address, MAC address of the detected device, transmission interval, number of retransmission times, and number of recovery times based on the actual environment to enable the DLDP detection function.

1.1.3 MAC Address Binding

In a complex network environment, DLDP may obtain an invalid MAC address if abnormal ARP packets are transmitted (due to ARP spoofing) in the detected link. As a result, DLDP fails to conduct detection normally. To address this problem, you can bind the to-be-detected IP address (or the next-hop IP address) to the MAC address of the device to avoid a DLDP failure caused by ARP spoofing. After binding, DLDP sends ARP packets and ICMP packets with a fixed destination IP address and a fixed destination MAC address during detection. If the source IP address and MAC address in a received packet do not match the bound IP address and MAC address, DLDP does not process the packet.

1.2 Configuring DLDP Detection

1.2.1 Overview

DLDP detects L3 link connectivity. When an L3 link is abnormal, DLDP shuts down the related Layer 3 interfaces.

1.2.2 Restrictions and Guidelines

- To configure the DLDP function, configure the next-hop IP address, MAC address, transmission interval, retransmission times, and recovery times based on the actual environment. You are not advised to set the number of retransmission times and the number of recovery times to 1. You are advised to use the default value for the transmission interval. The number of DLDP sessions and the configured CPU protect policy (CPP) bandwidth must be considered when you adjust the value. Otherwise, the device performance consumption increases and misjudgment and flapping can be caused easily. In the case of more than 10 sessions, the recommended interval is not smaller than 100 ticks.
- DLDP can detect multiple IP addresses configured on an L3 port. DLDP sets the port to down when none of the IP addresses returns an ICMP reply. If one of the IP addresses resumes communication, DLDP sets the port to up.
- DLDP uses the first IP address of the L3 port as the source IP address of detection packets.

1.2.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 3 Ethernet interface configuration mode.

interface *interface-type interface-number*

(4) Enable DLDP detection.

dldp *ipv4-address* [*next-hop-ipv4-address*] [**mac-address** *mac-address*] [**interval** *tick-interval* | **resume** *resume-number* | **retry** *retry-number*]

By default, the number of retransmission times of DLDP detection is **4**, the number of recovery times is **3**, and the detection interval is **1** second.

(5) (Optional) Configure a DLDP detection mode.

dldp passive

The default DLDP detection mode is active mode.

(6) (Optional) Configure a global detection interval for DLDP detection.

dldp interval *tick-interval*

The default global detection interval of DLDP detection is **1** second.

The configuration of this command takes effect immediately for all DLDP detection operations.

(7) (Optional) Configure the number of global retransmission times for DLDP detection.

dldp retry *retry-number*

The default number of global retransmission times of DLDP detection is **4**.

The configuration of this command takes effect immediately for all DLDP detection operations.

(8) (Optional) Configure the number of link recovery times for all DLDP detection operations.

dldp resume *resume-number*

The default number of link recovery times of all DLDP detection operations is **3**.

The configuration of this command takes effect immediately for all DLDP detection operations.

1.3 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.



Caution

Running the **cleanup** command during operation of the device may lose vital information and interrupt services.

Table 1-1 DLDP Monitoring

Command	Purpose
cleanup dldp [**interface** *interface-type interface-number* [*ipv4-address*]]	Clears the statistics on link up/down counts of a DLDP monitoring point.
display dldp [**interface** *interface-type interface-number*]	Displays the configuration of a DLDP monitoring point.
display dldp [**interface** *interface-type interface-number*] [**statistic**]	Displays the statistics of a DLDP monitoring point.

1.4 Configuration Examples

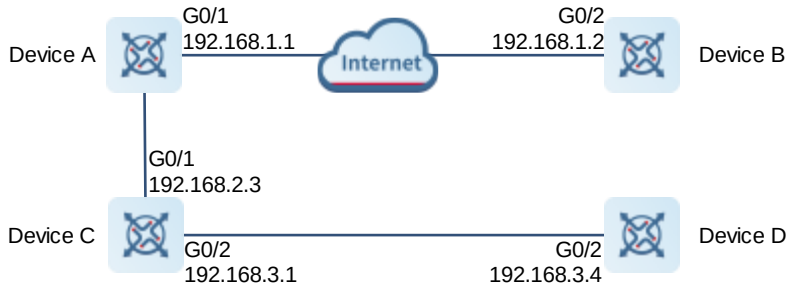
1.4.1 Configuring DLDP Detection

1. Requirements

DLDP detection needs to be enabled on an L3 network to control the Layer 3 interfaces of devices A and B.

2. Topology

Figure 1-1 Configuring DLDP Detection



3. Notes

- Enable DLDP detection on the Layer 3 interfaces (GigabitEthernet 0/1 and GigabitEthernet 0/2) on device A to detect the L3 network connectivity from device A to device B and from device A to device D.
- To control the Layer 3 interfaces (GigabitEthernet 0/2) of device B, enable DLDP detection and configure passive mode on the port.

4. Procedure

Perform the following configuration on device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# dldp 192.168.1.2
Device A(cfg-if-GigabitEthernet 0/1)# exit
```

```
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# dldp 192.168.3.4 192.168.2.3
```

Perform the following configuration on device B.

```
Device B> privilege
Device B# config-system terminal
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# dldp 192.168.1.1
Device B(cfg-if-GigabitEthernet 0/2)# dldp passive
```

5. Verification

Run the **display dldp** command to check the DLDP state information on devices A and B and check whether DLDP detection is enabled and works normally.

```
Device A# display dldp
Interface   Type            Ip           Next-hop      Interval   Retry   Resume   State
---------   -------   -----------   -----------   --------   -----   ------   ------
Gi0/1         Active    192.168.1.2                    100          4         3          Up
Gi0/1         Active    192.168.3.4   192.168.2.3   100          4         3          Up
```

```
Device B# display dldp
Interface   Type            Ip           Next-hop      Interval   Retry   Resume   State
---------   -------   -----------   -----------   --------   -----   ------   ------
Gi0/2         Passive   192.168.1.1                    100          4         3          Up
```

6. Common Errors

- An unreachable IPv4 unicast route is misjudged as a DLDP detection failure.
- DLDP detection fails because the peer device does not support ARP/ICMP replies.
- No next-hop IP address is configured for inter-network segment DLDP detection.

Contents

1 Configuring M-LAG 1

1.1 Introduction 1

1.1.1 Basic Concepts 1

1.1.2 M-LAG System Negotiation 3

1.1.3 Anti-Loop Mechanism 4

1.1.4 Local Port Preferential Forwarding 5

1.1.5 Configuration Consistency Check 6

1.2 Restrictions and Guidelines 7

1.3 Configuration Task Summary 7

1.4 Configuring the M-LAG Function 7

1.4.1 Overview 7

1.4.2 Restrictions and Guidelines 8

1.4.3 Configuration Tasks 9

1.4.4 Configuring Local and Peer IP Addresses 10

1.4.5 Configuring Peer Link Ports 10

1.4.6 Configuring M-LAG Attributes 11

1.4.7 Configuring Member Ports of a Virtual Aggregate Interface 12

1.4.8 Configuring Exceptional Interfaces 12

1.5 Monitoring 13

1.6 Configuration Examples 14

1.6.1 Connecting to an IP Network in Dual-Homing Mode 14

1 Configuring M-LAG

Note

M-LAG is not supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.1 Introduction

The multi-chassis link aggregation group (M-LAG) technology enables two devices supporting link aggregation to form a dual-active system, which is equivalent to one device. The equivalent device and the downlink device are interconnected through an aggregate port (AP). Deploying M-LAG improves the link reliability to the device level to meet the requirements of high availability scenarios.

1.1.1 Basic Concepts

Figure 1-1 M-LAG Topology

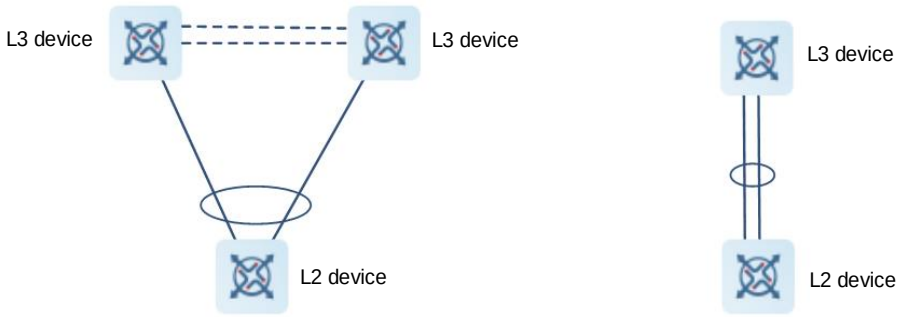


Figure 1-1 shows the M-LAG topology. Another similar technology is virtual switching unit (VSU), which also virtualizes multiple devices into one. M-LAG features the following advantages compared with the VSU topology.

- Stacking is discarded and only L2 virtualization is required. The two devices are still independent of each other. M-LAG reduces deployment difficulties and eliminates shortcomings brought by device stacking (for example, the software failure of the master device may make both devices unavailable).
- Devices can be upgraded independently without affecting the normal operation of the other device.

1. Domain ID

A domain ID is the unique identifier of an M-LAG system and distinguishes different M-LAG systems. Only two devices that share the same domain ID can compose an M-LAG system. At present, only one domain can be configured for one device, and only two devices can be configured in one domain.

2. VAP

On the two devices of M-LAG, a pair of link aggregations connected to the same access device form a virtual AP group (also known as VAP), which allows an access device to connect in dual-homing mode. Different access devices connect to different VAPs. Only Layer 2 link aggregations are supported in a VAP.

When the Link Aggregation Control Protocol (LACP) is enabled on the two link aggregations of a VAP, the same LACP system ID and system priority are used in LACP negotiation. For the uplink devices of an access device, the M-LAG dual devices are logically one device.

Note

Only Layer 2 link aggregations are supported in a VAP.

3. Peer Link

The link between two devices of M-LAG is called a peer link, which is used to synchronize data and transmit some traffic. The directly connected ports at both ends of a peer link are called peer link ports, which must be configured as Layer 2 link aggregations. To improve the reliability of the peer link, you are advised to deploy multiple physical links for the peer link. For chassis devices, deploy multiple physical ports on different cards to reduce the impact of a card failure.

To enhance the reliability of data transmission between M-LAG devices, use Transmission Control Protocol (TCP) for synchronous data transmission. You need to configure a switch virtual interface (SVI) on the peer link and configure L3 reachability between the master and slave devices of M-LAG to achieve data synchronization between the devices.

4. Device Role

During normal operation, the two M-LAG devices negotiate to decide the master and slave roles. Normally, both the master and slave devices participate in forwarding and there is no difference between them. When a fault occurs, the behaviors of the master and slave devices are different. For example, if the peer link gets faulty and the VAP of the two devices fails, service ports on the slave device need to be shut down and traffic needs to be forcibly switched to the master device, to prevent abnormal forwarding of the access device.

5. VAP Member Port

An AP joining a VAP is called a VAP member port, which is used to connect to an AP of an access device. You are advised to use LACP to enhance reliability.

6. Dual-Active Detection Link

A dual-active detection link is an L3 path, which is used to detect the dual-active state of M-LAG devices in case of a peer link failure. To ensure real-time detection, the dual-active detection link does not participate in any forwarding. You are advised to configure a separate L3 reachable link as a dual-active detection link (for example, configure the directly connected link between the management ports of two devices as a dual-active detection link).

7. Heartbeat Hold Time

When the peer link fails, both devices start the heartbeat hold timer, and waits for the packet of the keepalive mechanism on the dual-active detection link to be ready. After the timer expires, the devices start the dual-active detection to prevent error detection due to a packet delay.

8. Domain Negotiation Timeout Time

The VAPs on the two M-LAG devices use the same LACP system ID and system priority. When the devices are just started, VAP member ports start LACP negotiation only after M-LAG negotiation succeeds and the LACP

system ID and system priority are obtained. However, if the peer device is powered off, the negotiation fails all the time, causing the local VAP member port to fail. Therefore, you need to set negotiation timeout time. After the timeout time expires, the local device is deemed to work in the standalone state, and the LACP system ID and system priority of the local device are directly used to perform LACP negotiation.

9. Fast Convergence Mode

Fast convergence mode refers to unidirectional isolation without blocking unicast packets. When a VAP member port recovers from a failure, the MAC entry egress of the uplink port switches from the peer link port to the VAP member port. However, when there are many MAC entries, some traffic sent to the access device continues to use the peer link port during the switching. Due to the anti-loop mechanism of the M-LAG system, the peer device enables unidirectional isolation (the peer link port does not forward packets to the downlink port), resulting in traffic interruption. To implement fast convergence, you can configure not to block unicast packets during unidirectional isolation, but this may cause the downlink access device to receive duplicate packets. For example, device 1 and device 2 form an M-LAG system according to M-LAG configuration. The request packet of a downlink device is forwarded by device 1. At this moment, device 1 learns the MAC entry of the downlink device, but has not synchronized it to device 2. If the reply packet is forwarded by device 2, device 2 floods the packet in the VAP member port and peer link port. After receiving the packet from the peer link port, device 1 searches for the MAC address of the downlink device in the MAC address table, and also forwards the packet through the VAP member port. Now, the downlink device receives two duplicate reply packets.

Enabling the fast convergence mode accelerates the convergence for failure recovery, but there may be instantaneous duplicate packets. If you do not want duplicate packets in some scenarios, disable the fast convergence mode.

1.1.2 M-LAG System Negotiation

The basis of M-LAG application is that two devices pair with each other to form a system and provide the inter-device aggregation capability. The process is as follows:

1. System Pairing

After the domain ID, peer link, and peer-link-based L3 channel are configured for the two M-LAG devices, the devices synchronize the domain ID with each other through the L3 channel. After receiving a synchronization message, a device checks whether the domain ID is consistent with its local domain ID. If yes, the two devices pair with each other successfully.

2. Master/Slave Negotiation

After successful pairing, the two devices synchronize the system MAC address and M-LAG priority with each other, and compare their M-LAG priorities and system MAC addresses to elect the master and slave roles. The negotiation rules are as follows: The M-LAG priority is compared first, and a device with a higher M-LAG priority is elected as the master device. If the priorities are the same, the system MAC address is compared, and a device with a smaller MAC address is elected as the master device.

The LACP system ID and LACP system priority of the master device are selected as the LACP system ID and system priority of the M-LAG VAP. The LACP system ID and system priority are used when Layer 2 link aggregations in a VAP perform LACP negotiation.

Caution

When both the master and slave devices operate properly, packet forwarding is the same on them. However, packet forwarding is different on them in failure scenarios. For example, when the peer link is faulty, the VAP between the two devices fails. To avoid abnormal forwarding of traffic from the access device, the service ports on the slave device are shut down to switch traffic to the master device.

3. Forwarding Entry Synchronization

After successful M-LAG negotiation, the two devices synchronize forwarding entries mutually such as MAC entries and ARP entries, to ensure that the forwarding function of the dual-active system is consistent.

4. Dual-Active Detection

After M-LAG runs properly, heartbeat packets are sent periodically between the two devices to detect availability of the dual-active link. When a dual-active link failure is detected, if the peer link works properly, a prompt is displayed, reminding you to check the dual-active link.

When the peer link is faulty, the devices send dual-active detection packets to each other.

- If a device receives a detection response packet, dual master devices exist, and the service ports (the ports other than MGMT ports, peer link ports, and stacked ports) on the slave device are forced to enter the errdisable down state. To exempt a port from the errdisable down state, you can configure it as an excluded port, for example, the port used for dual-active detection.
- If no detection response packet is received, the peer device is faulty. If the local device is the slave device, it becomes the master device. If the local device is the master device, it does not take any actions.
- When the peer link recovers from the failure, the ports in the errdisable down state on the slave device recover immediately. However, because there may be a lot of protocols, it takes a certain period of time to complete convergence. During this period of time, the VAP member port is set to the up state in advance, which leads to a traffic forwarding failure and traffic interruption. Therefore, you can configure a delay for the port recovery so that the VAP member port recovers after the delay time expires.

1.1.3 Anti-Loop Mechanism

When an access device connects to a network through M-LAG in dual-homing mode, the two APs of the VAP are distributed on two independent devices, and independent forwarding by the two APs may result in a loop or receiving of two copies of packets. M-LAG anti-loop rules are as follows:

- When the peer VAP member port works properly, traffic from the peer link is not forwarded to the local VAP member port, as shown in Figure 1-2.
- When the peer VAP member port fails, traffic from the peer link needs to be forwarded to the local VAP member port, as shown in Figure 1-3.

Figure 1-2 Peer Link Loop Prevention

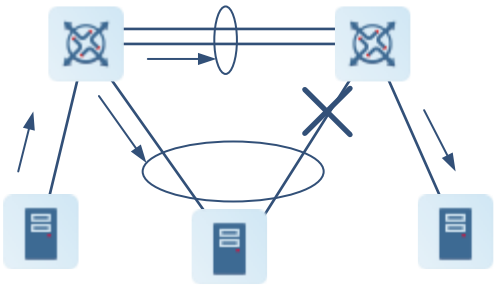
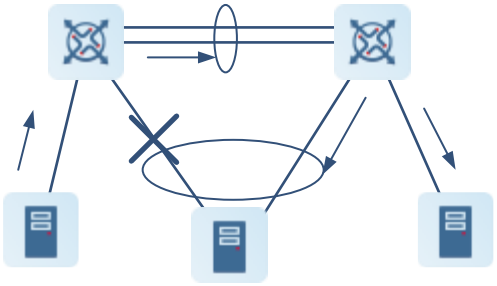


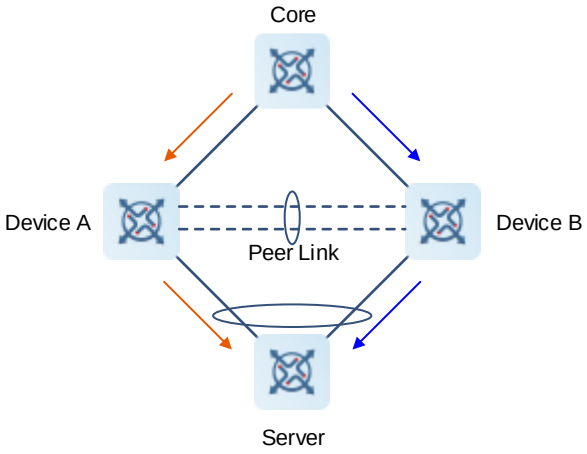
Figure 1-3 Peer Link Loop Removal



1.1.4 Local Port Preferential Forwarding

In a network, a peer link is regarded as a backup link. Service traffic is preferentially forwarded by local VAP member ports instead of peer links.

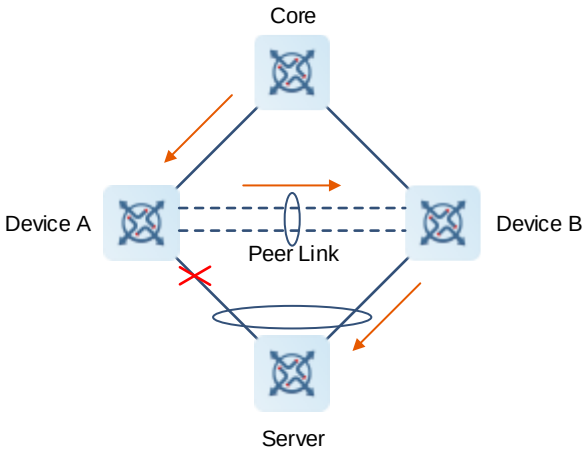
Figure 1-4 Forwarding Path of Service Traffic on a Normal Network



Through the following mechanism, M-LAG ensures that the device forwards service traffic by local VAP member ports preferentially on a normal network.

- The interfaces of a peer link keeps learning MAC addresses to ensure that the egress MAC addresses learned by the local ports are downlink ports.
- If a MAC table entry is synchronized from the M-LAG peer side, the device replaces the egress port of the MAC table entry with a local downlink port.
- Once the local VAP member ports of the device fail, service traffic will be forwarded through a peer link.

Figure 1-5 Forwarding Path of Service Traffic When the VAP Member Ports of Device A Fail



1.1.5 Configuration Consistency Check

In the M-LAG scenario, some service configurations need to be consistent on the two devices composing the M-LAG system, to ensure normal operation of the M-LAG system functions. Configuration information sharing of the M-LAG system realizes automatic configuration consistency check for the two devices in the M-LAG system. The configuration change at one end automatically triggers the configuration consistency check. When the configurations of the M-LAG master and slave devices are inconsistent, an alarm is prompted. If the key configurations on the master and slave devices of the M-LAG system are inconsistent, the service ports on the slave device are set to the errdisable down state; when the configuration consistency is restored, M-LAG recovers the ports in the errdisable down state on the slave device. According to the impact caused by configuration inconsistency, the configurations that need to be kept consistent are classified into two categories: key configuration (Type-1) and general configuration (Type-2).

- Key configuration: If the configurations are inconsistent between the devices at both ends of the M-LAG system, it causes the problems of a loop and a normal state with long-term packet loss.
- General configuration: If the configurations are inconsistent between the devices at both ends of the M-LAG system, the M-LAG system may operate abnormally. Compared with the key configuration, the problem of general configuration is easier to be found, and its impact on the networking environment is relatively minor.

Table 1-1 Configuration Consistency Check Table

Command Mode	Consistency Check Configuration Item	Type
Global configuration mode	Whether the Spanning Tree Protocol (STP) function is enabled	Type-1
	Configuration of the STP working mode	
	Whether the bridge protocol data unit (BPDU) protection function is enabled	
	Mappings between the Multiple Spanning Tree Protocol (MSTP) instances and virtual local area networks (VLANs)	
	SVI	Type-2

Command Mode	Consistency Check Configuration Item	Type
	VLAN and VXLAN network identifier (VNI) mapping configuration	
	OverlayRouter (OR) port configuration	
VAP member port	Whether the root protection function of an STP port is enabled	Type-1
	LACP mode configuration	
	STP priority	Type-2
	VLAN, to which a port is added	

Based on the handling method of inconsistency found in key configuration check, M-LAG consistency check can be performed in strict mode or loose mode.

- In strict mode, if the key configurations are inconsistent, the service ports on the M-LAG slave device are set to the errdisable state.
- When the general configurations are inconsistent, only an alarm is prompted regardless of the check mode.

1.2 Restrictions and Guidelines

- The peer link between the two devices must be an aggregate link. You are advised to use LACP aggregation and configure the peer link as a trunk port in L2 mode. Do not configure other services on the port.
- You are advised to configure multiple physical links for the peer link. For chassis devices, deploy physical links on different cards to avoid the impact of a card failure on the network.
- The dual-active detection link needs to be independent of the peer link to avoid a dual-active detection failure in case of a peer link failure.
- The M-LAG devices only synchronize the dynamic MAC/ARP entries, but not the static MAC/ARP configuration. The same static MAC/ARP entries need to be configured on the M-LAG devices.
- Only L2 APs can be configured as peer link ports or VAP member ports.

1.3 Configuration Task Summary

M-LAG configuration includes the following tasks:

(1) Configuring the M-LAG Function

1.4 Configuring the M-LAG Function

1.4.1 Overview

You can configure device link aggregation to enable an access device to connect to an M-LAG system in dual-homing mode. The traffic from an M-LAG device to the access device is first forwarded through the local VAP member port. When the local VAP member port fails, the traffic is forwarded to the M-LAG peer device. The dual-homing access meets requirements for balanced forwarding of traffic. When a single device or a single link fails, the device traffic is not affected and can be forwarded continuously.

1.4.2 Restrictions and Guidelines

- On M-LAG device, the two ports connected to the dual-homed access device need to join the same VAP.
- You are advised to use consistent configuration for the member ports of the same VAP on the two M-LAG devices to ensure normal traffic forwarding.
- The data synchronization channel needs to be established on the peer link.
- Due to the restriction of the M-LAG master and slave role election rules, the M-LAG master and slave roles are not changed under the following cases:
  - When the role of one device in the M-LAG system has been selected and the role of the other device is not selected, the priority change will not lead to role election.
  - When one device in the M-LAG system is in the dual-active errdisable state, the priority change will not lead to role election.
- Devices in an M-LAG system can be restarted only after you power down the system or run the **reboot** command. Using other methods to restart the device may cause the shutdown time of the keepalive port to be later than that of the peer link and the keepalive packet transmission interval out of the heartbeat hold time, and, consequently, the M-LAG slave device is wrongly set to the errdisable state by the dual-active detection mechanism. In this case, you are advised to configure a longer heartbeat hold time.

Table 1-2 Table of Special Constraints with Other Services

Feature	Description
Stacking	Devices can be used to build a stack first, and then the stack system can be used as an independent unit to build an M-LAG system.
CFM	The M-LAG and connectivity fault management (CFM) functions are mutually exclusive. They cannot be configured at the same time.
GVRP	The M-LAG and GARP VLAN Registration Protocol (GVRP) functions are mutually exclusive. They cannot be configured at the same time.
ERPS	The M-LAG and Ethernet Ring Protection Switching (ERPS) functions are mutually exclusive. They cannot be configured at the same time.
RLDP	The M-LAG and Rapid Link Detection Protocol (RLDP) functions are mutually exclusive. They cannot be configured at the same time.
Web authentication	The M-LAG and Web authentication functions are mutually exclusive. They cannot be configured at the same time.
dot1x	The M-LAG and dot1x functions are mutually exclusive. They cannot be configured at the same time.
Dynamic ARP detection	The M-LAG and dynamic ARP detection functions are mutually exclusive. They cannot be configured at the same time.

Feature	Description
DHCP	The two devices composing an M-LAG system do not support Dynamic Host Configuration Protocol (DHCP) snooping DHCP relay needs to be configured on both devices composing an M-LAG system. The DHCP server needs to be configured on both devices composing an M-LAG system and the addresses in the address pools of the two devices cannot be overlapped.
IP unicast routing	The M-LAG devices and connected devices cannot establish routing neighbor relationships through VAP member ports.
Multicast	M-LAG does not support IPv4 multicast or IPv6 multicast services.
IPSG	The two devices composing an M-LAG system do not support the IP source guard or IPv6 source guard.
VXLAN	● On the two devices composing an M-LAG system, L2 VNI mapping or outside center cannot be configured to implement the L2 data center interconnect (DCI). ● In scenarios where both M-LAG and VXLAN are configured, you are advised to configure one-to-one mappings between VNIs and VLANs on the peer link ports. The one-to-many mappings cause flooding of broadcast packets on the peer link ports. ● In scenarios where both M-LAG and VXLAN Layer 2 Ethernet subinterfaces are configured, peer link ports need to be added to all VNIs. ● In scenarios where both M-LAG and VXLAN Layer 2 Ethernet subinterfaces are configured, QinQ access is not supported on VXLAN L2 sub-ports.
Storm control	You are not advised to configure multicast packet storm control on physical member ports of a peer link. Otherwise, M-LAG synchronization packets are suppressed, resulting in data packet forwarding exceptions in the M-LAG system.
VLAN-mapping	The M-LAG and VLAN-mapping functions are mutually exclusive. They cannot be configured at the same time.

1.4.3 Configuration Tasks

M-LAG configuration includes the following tasks:

(1) Configuring Local and Peer IP Addresses

(2) Configuring Peer Link Ports

(3) Configuring M-LAG Attributes

(4) Configuring Member Ports of a Virtual Aggregate Interface

(5) Configuring Exceptional Interfaces

1.4.4 Configuring Local and Peer IP Addresses

1. Overview

To ensure normal negotiation between two devices that form an M-LAG system, you need to specify the local and peer IP addresses for entry synchronization and dual-active detection channels.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the dual-active link unreachability logging.

logging vap keepalive-det

Note

This command is supported by only the S6510-32C.

(4) Configure a VAP domain.

vap domain *domain-id*

No VAP domain ID is configured by default.

The same VAP domain ID needs to be configured for the two devices used for dual-homing access.

(5) Configure the local and peer IP addresses for the Layer 3 channel used for entry synchronization.

data-sync { *local-ipv4-address* **peer** *peer-ipv4-address* | *local-ipv6-address* **peer** *peer-ipv6-address* }

No local and peer IP addresses are configured for a Layer 3 channel used for entry synchronization by default.

(6) Configure the local and peer IP addresses for the Layer 3 channel used for dual-active detection.

peer-keepalive local { *ipv4-address* | *ipv6-address* } **peer** { *ipv4-address* | *ipv6-address* } [*interface-type interface-number*] (Optional) Configure the domain negotiation timeout time.

1.4.5 Configuring Peer Link Ports

Note

This function is supported by only the S6980-64QC. S6920-4C, S6510-4C, S6520-64CQ, S6510-32CQ, and S6510-48VS8CQ.

2. Overview

The peer link is used to synchronize data and transmit some traffic. The directly connected ports at both ends of a peer link are called peer link ports, which must be configured as a Layer 2 aggregate interface. To improve the reliability of the peer link, you are advised to deploy multiple physical links for the peer link.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 aggregate interface configuration mode.

interface aggregateport *interface-number*

(4) Configure the Layer 2 aggregate interface as a peer link interface.

peer-link

When an interface is configured as a peer link port, other services cannot be configured on the interface, for example, it cannot be configured as a protected port. Moreover, MAC address learning is disabled on the interface, that is, M-LAG devices communicate with each other through flooding. You are advised to configure a peer link port with the trunk mode.

No peer link port is configured by default.

1.4.6 Configuring M-LAG Attributes

1. Overview

The values of the M-LAG attributes can be adjusted based on networking and service requirements.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a VAP domain.

vap domain *domain-id*

No VAP domain ID is configured by default.

(4) (Optional) Configure the domain negotiation timeout time.

domain-match delay *interval*

The default delay time is **120** seconds.

(5) (Optional) Configure the fast convergence mode.

fast-convergence

The fast convergence mode is enabled by default.

(6) (Optional) Configure a priority.

priority *priority*

The default priority is **4**.

(7) (Optional) Configure an interface recovery delay.

recover up-delay *interval* [**none-vap** *none-vap-interval*]

The default recovery delay is 120 seconds for VAP interfaces and 0 seconds for other interfaces.

(8) (Optional) Configure the heartbeat hold time.

peer-keepalive hold-time *interval*

The default heartbeat hold time for dual-active detection is **3** seconds.

(9) (Optional) Configure dual-active automatic recovery.

dual-active auto recovery

The dual-active automatic recovery function is disabled by default.

(10) (Optional) Configure consistency check.

vap consistency-check [**disable** | **loose** | **strict**]

Consistency check is not configured by default.

1.4.7 Configuring Member Ports of a Virtual Aggregate Interface

1. Overview

You can add aggregate interfaces into a virtual aggregate interface. You are advised to configure LACP to enhance reliability.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 aggregate interface configuration mode.

interface aggregateport *interface-number*

(4) Add an aggregate interface to a virtual aggregate interface group.

vap *vap-id*

1.4.8 Configuring Exceptional Interfaces

1. Overview

If M-LAG detects two active devices, it will configure the service interfaces (except for management interfaces, peer link ports, and stack interfaces) on the standby device as error down interfaces. To prevent an interface from being configured as an error down interface, you can configure this interface as an exceptional interface.

2. Procedure

(1) Enter the privileged EXEC mode.
privilege

(2) Enter the global configuration mode.
config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 aggregate interface configuration mode.

  interface *aggregateport interface-number*

- Enter the Layer 2 Ethernet sub-interface configuration mode.

interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregate sub-interface configuration mode.

  interface *aggregateport interface-number.subnumber*

(4) (Optional) Configure an exceptional interface.
vap error-down except

No exceptional interface is configured by default.

1.5 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the **cleanup** command to clear information.

Caution

Running the **cleanup** commands may interrupt services due to loss of important information.

Table 1-3 M-LAG Monitoring

Command	Purpose
display vap	Displays the information about a VAP.
display vap data-sync	Displays the information about a data synchronization channel.
display vap keepalive	Displays the information about an M-LAG dual-active detection link.
display vap peer-link	Displays the information about the peer link.
display vap summary	Displays the brief information about M-LAG.
display vap peer-keepalive	Displays the information about a dual-active keepalive link.
display vap role	Displays the role information of M-LAG devices.
display vap error-down	Displays the errdisable state information of an M-LAG port.
display vap error-down except	Displays the configuration of the excluded port exempt from the errdisable state.
debugging mlag event	Debugs M-LAG events.

Command	Purpose
debugging mlag dad	Debugs M-LAG dual-active detection events.
debugging mlag dad-pkt	Debugs M-LAG keepalive packets.
debugging mlag mom	Debugs the M-LAG database.
debugging mlag cli	Debugs the M-LAG CLI configuration.
debugging mlag role	Debugs the M-LAG role negotiation.
debugging mlag role-pkt	Debugs M-LAG role negotiation packets.
debugging mlag error	Debugs M-LAG error events.
debugging mlag keypnt	Debugs M-LAG key events.
debugging mlag vap	Debugs the M-LAG group management.
debugging mlag log-file	Debugs the function of writing M-LAG logs into files.
debugging mlag [**cons-check** | **cons-mom**]	Debugs M-LAG configuration consistency check.
debugging mlag snmp	Debugs M-LAG trap.
cleanup vap error-down	Clear the ports from the errdisable state caused by M-LAG.
cleanup vap data-sync	Clear data records of the data synchronization channel.
cleanup vap keepalive	Clear data records of the heartbeat channel.

1.6 Configuration Examples

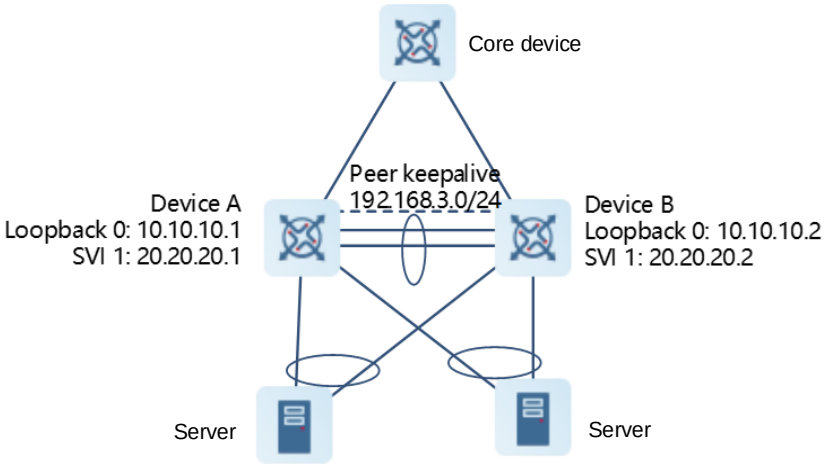
1.6.1 Connecting to an IP Network in Dual-Homing Mode

1. Requirements

A device needs to be configured to connect to an IP network in dual-homing mode.

2. Topology

Figure 1-6 Connecting to an IP Network in Dual-Homing Mode



3. Notes

- Configure interface IP addresses for all the devices.
- Configure a dynamic routing protocol (such as OSPF) on devices A and B and the core device.

4. Procedure

Configure a VAP domain on device A.

```
Device A> privilege
Device A# config-system terminal
Device A(cfg)# vap domain 1
Device A(cfg-vap-domain)# exit
```

Configure a data backup channel and a dual-active detection channel.

```
Device A(cfg)# interface loopback 0
Device A(cfg-if-Loopback 0)# ip address 10.10.10.1/32
Device A(cfg-if-Loopback 0)# exit
Device A(cfg)# vap domain 1
Device A(cfg-vap-domain)# data-sync local 20.20.20.1 peer 20.20.20.2
Device A(cfg-vap-domain)# peer-keepalive local 192.168.3.1 peer 192.168.3.2
Device A(cfg-vap-domain)# exit
Device A(cfg)# int vlan 1
Device A(cfg-if-VLAN 1)# ip address 20.20.20.1/24
Device A(cfg-if-VLAN 1)# exit
Device A(cfg)# ip route-config 10.10.10.2 255.255.255.255 20.20.20.2
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device A(cfg-if-GigabitEthernet 0/1)# ip address 192.168.3.1/24
```

Configure the peer link AP: aggregate interface 1. Add member port GigabitEthernet 0/4 and member port GigabitEthernet 0/5.

```
Device A(cfg)# interface aggregateport 1
```

```
Device A(cfg-if-AggregatePort 1)# switch-port mode trunk
Device A(cfg-if-AggregatePort 1)# peer-link
Device A(cfg-if-AggregatePort 1)# exit
Device A(cfg)# interface gigabitethernet 0/4
Device A(cfg-if-GigabitEthernet 0/4)# port-group 1 mode active
Device A(cfg-if-GigabitEthernet 0/4)# exit
Device A(cfg)# interface gigabitethernet 0/5
Device A(cfg-if-GigabitEthernet 0/5)# port-group 1 mode active
```

Configure VAP 2. Add the downlink port GigabitEthernet 0/2 to aggregate interface 2 and aggregate interface 2 to VAP 2.

```
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# port-group 2 mode active
Device A(cfg-if-GigabitEthernet 0/2)# exit
Device A(cfg)# interface aggregateport 2
Device A(cfg-if-AggregatePort 2)# switch-port access vlan 2
Device A(cfg-if-AggregatePort 2)# vap 2
```

Configure VAP 3. Add the downlink port GigabitEthernet 0/3 to aggregate interface 3 and aggregate interface 3 to VAP 3.

```
Device A(cfg)# interface gigabitethernet 0/3
Device A(cfg-if-GigabitEthernet 0/3)# port-group 3 mode active
Device A(cfg-if-GigabitEthernet 0/3)# exit
Device A(cfg)# interface aggregateport 3
Device A(cfg-if-AggregatePort 3)# switch-port access vlan 3
Device A(cfg-if-AggregatePort 3)# vap 3
```

Configure a VRRP dual-active gateway.

```
Device A(cfg)# vlan 2
Device A(cfg-vlan)# exit
Device A(cfg)# interface vlan 2
Device A(cfg-if-VLAN 2)# ip address 30.30.2.1/24
Device A(cfg-if-VLAN 2)# vrrp-config 1 ip 30.30.2.1
Device A(cfg-if-VLAN 2)# vrrp-config mode dual-active
Device A(cfg-if-VLAN 2)# exit
Device A(cfg)# vlan 3
Device A(cfg-vlan)# exit
Device A(cfg)# interface vlan 3
Device A(cfg-if-VLAN 2)# ip address 30.30.3.1/24
Device A(cfg-if-VLAN 2)# vrrp-config 1 ip 30.30.3.1
Device A(cfg-if-VLAN 2)# vrrp-config mode dual-active
```

Configure a monitor link, with the uplink port GigabitEthernet 0/1 and downlink ports GigabitEthernet 0/2 and GigabitEthernet 0/3.

```
Device A(cfg)# link state enable-track-group 1 up-delay 60
Device A(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# link state trace-group 1 upstream
Device A(cfg-if-GigabitEthernet 0/1)# exit
```

```
Device A(cfg)# interface gigabitethernet 0/2
Device A(cfg-if-GigabitEthernet 0/2)# link state trace-group 1 downstream
Device A(cfg-if-GigabitEthernet 0/2)# exit
Device A(cfg)# interface gigabitethernet 0/3
Device A(cfg-if-GigabitEthernet 0/3)# link state trace-group 1 downstream
```

Perform the following configuration on device B.

```
Device B> privilege
Device B# config-system terminal
```

Configure a VAP domain.

```
Device B(cfg)# vap domain 1
Device B(cfg-vap-domain)# exit
```

Configure a data backup channel and a dual-active detection channel.

```
Device B(cfg)# interface loopback 0
Device B(cfg-if-Loopback 0)# ip address 10.10.10.2/32
Device B(cfg-if-Loopback 0)# exit
Device B(cfg)# vap domain 1
Device B(cfg-vap-domain)# data-sync local 20.20.20.2 peer 20.20.20.1
Device B(cfg-vap-domain)# peer-keepalive local 192.168.3.2 peer 192.168.3.1
Device B(cfg-vap-domain)# exit
Device B(cfg)# int vlan 1
Device B(cfg-if-VLAN 1)# ip address 20.20.20.2/24
Device B(cfg-if-VLAN 1)# exit
Device B(cfg)# ip route-config 10.10.10.1 255.255.255.255 20.20.20.1
Device B(cfg)# interface gigabitethernet 0/1
Device A(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device B(cfg-if-GigabitEthernet 0/1)# ip address 192.168.3.2/24
```

Configure a peer link AP: aggregate interface1. Add member ports GigabitEthernet0/4 and GigabitEthernet 0/5.

```
Device B(cfg)# interface aggregateport 1
Device B(cfg-if-AggregatePort 1)# switch-port mode trunk
Device B(cfg-if-AggregatePort 1)# peer-link
Device B(cfg-if-AggregatePort 1)# exit
Device B(cfg)# interface gigabitethernet 0/4
Device B(cfg-if-GigabitEthernet 0/4)# port-group 1 mode active
Device B(cfg-if-GigabitEthernet 0/4)# exit
Device B(cfg)# interface gigabitethernet 0/5
Device B(cfg-if-GigabitEthernet 0/5)# port-group 1 mode active
```

Configure VAP 2. Add the downlink port GigabitEthernet 0/2 to aggregate interface 2 and aggregate interface 2 to VAP 2.

```
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# port-group 2 mode active
Device B(cfg-if-GigabitEthernet 0/2)# exit
Device B(cfg)# interface aggregateport 2
Device B(cfg-if-AggregatePort 2)# switch-port access vlan 2
Device B(cfg-if-AggregatePort 2)# vap 2
```

Configure VAP 3. Add the downlink port GigabitEthernet 0/3 to aggregate interface 3 and aggregate interface 3 to VAP 3.

```
Device B(cfg)# interface gigabitethernet 0/3
Device B(cfg-if-GigabitEthernet 0/3)# port-group 3 mode active
Device B(cfg-if-GigabitEthernet 0/3)# exit
Device B(cfg)# interface aggregateport 3
Device B(cfg-if-AggregatePort 3)# switch-port access vlan 3
Device B(cfg-if-AggregatePort 3)# vap 3
```

Configure a VRRP dual-active gateway.

```
Device B(cfg)# vlan 2
Device B(cfg-vlan)# exit
Device B(cfg)# interface vlan 2
Device B(cfg-if-VLAN 2)# ip address 30.30.2.2/24
Device B(cfg-if-VLAN 2)# vrrp-config 1 ip 30.30.2.1
Device B(cfg-if-VLAN 2)# vrrp-config mode dual-active
Device B(cfg-if-VLAN 2)# exit
Device B(cfg)# vlan 3
Device B(cfg-vlan)# exit
Device B(cfg)# interface vlan 3
Device B(cfg-if-VLAN 2)# ip address 30.30.3.2/24
Device B(cfg-if-VLAN 2)# vrrp-config 1 ip 30.30.3.1
Device B(cfg-if-VLAN 2)# vrrp-config mode dual-active
```

Configure a monitor link, with the uplink port GigabitEthernet 0/1 and downlink ports GigabitEthernet 0/2 and GigabitEthernet 0/3.

```
Device B(cfg)# link state enable-track-group 1 up-delay 60
Device B(cfg)# interface gigabitethernet 0/1
Device B(cfg-if-GigabitEthernet 0/1)# link state trace-group 1 upstream
Device B(cfg-if-GigabitEthernet 0/1)# exit
Device B(cfg)# interface gigabitethernet 0/2
Device B(cfg-if-GigabitEthernet 0/2)# link state trace-group 1 downstream
Device B(cfg-if-GigabitEthernet 0/2)# exit
Device B(cfg)# interface gigabitethernet 0/3
Device B(cfg-if-GigabitEthernet 0/3)# link state trace-group 1 downstream
```

5. Verification

Check whether the VAP member port of device A and its state are correct.

```
Device A# display vap summary
M-LAG domain              : 1
M-LAG role                : Master
Local address             : 20.20.20.1
Peer address              : 20.20.20.2
Negotiation status        : ok
Peer keepalive status : alive

Peer-link             Status
```

```
-----------------------------
Ag 1                    UP

Group ID    VAP Port    Local-Status    Peer-Status
-----------------------------------------------------
2           Ag 2        UP              UP
3           Ag 3        UP              UP
```

Check whether the VAP member port of device B and its state are correct.

```
Device B# display vap summary
M-LAG domain              : 1
M-LAG role                : Slave
Local address             : 20.20.20.2
Peer address              : 20.20.20.1
Negotiation status        : ok
Peer keepalive status : alive

Peer-link             Status
-----------------------------
Ag 1                    UP

Group ID    VAP Port    Local-Status    Peer-Status
-----------------------------------------------------
2           Ag 2        UP              UP
3           Ag 3        UP              UP
```

6. Configuration Files

- Device A configuration file

```
!
link state enable-track-group 1 up-delay 60
!
interface GigabitEthernet 0/1
 unset switch-port
 link state trace-group 1 upstream
 ip address 192.168.3.1 255.255.255.0
!
interface GigabitEthernet 0/2
 port-group 2 mode active
 link state trace-group 1 downstream
!
interface GigabitEthernet 0/3
 port-group 3 mode active
 link state trace-group 1 downstream
!
interface GigabitEthernet 0/4
 port-group 1 mode active
```

```
!
interface GigabitEthernet 0/5
 port-group 1 mode active
!
interface AggregatePort 1
 unset mac-address-learning
 switch-port mode trunk
 peer-link
!
interface AggregatePort 2
 switch-port access vlan 2
 vap 2
!
interface AggregatePort 3
 switch-port access vlan 3
 vap 3
!
interface Loopback 0
 ip address 10.10.10.1 255.255.255.255
!
interface VLAN 1
 ip address 20.20.20.1 255.255.255.0
!
interface VLAN 2
 ip address 30.30.2.1 255.255.255.0
 vrrp-config 1 ip 30.30.2.1
 vrrp-config mode dual-active
!
interface VLAN 3
 ip address 30.30.3.1 255.255.255.0
 vrrp-config 1 ip 30.30.3.1
 vrrp-config mode dual-active
!
vap domain 1
 data-sync local 20.20.20.1 peer 20.20.20.2
 peer-keepalive local 192.168.3.1 peer 192.168.3.2
!
```

- Device B configuration file

```
!
link state enable-track-group 1 up-delay 60
!
interface GigabitEthernet 1/0/1
 unset switch-port
 link state trace-group 1 upstream
 ip address 192.168.3.2 255.255.255.0
!
```

```
interface GigabitEthernet 1/0/2
 port-group 2 mode active
 link state trace-group 1 downstream
!
interface GigabitEthernet 1/0/3
 port-group 3 mode active
 link state trace-group 1 downstream
!
interface GigabitEthernet 1/0/4
 port-group 1 mode active
!
interface GigabitEthernet 1/0/5
 port-group 1 mode active
!
interface AggregatePort 1
 unset mac-address-learning
 switch-port mode trunk
 peer-link
!
interface AggregatePort 2
 switch-port access vlan 2
 vap 2
!
interface AggregatePort 3
 switch-port access vlan 3
 vap 3
!
interface Loopback 0
 ip address 10.10.10.2 255.255.255.255
!
!
interface VLAN 1
 ip address 20.20.20.2 255.255.255.0
!
interface VLAN 2
 ip address 30.30.2.2 255.255.255.0
 vrrp-config 1 ip 30.30.2.1
 vrrp-config mode dual-active
!
interface VLAN 3
 ip address 30.30.3.2 255.255.255.0
 vrrp-config 1 ip 30.30.3.1
 vrrp-config mode dual-active
!
vap domain 1
 data-sync local 20.20.20.2 peer 20.20.20.1
```

```
 peer-keepalive local 192.168.3.2 peer 192.168.3.1
!
```

Contents

1 Configuring VRRP 1
1.1 Introduction 1
1.1.1 Basic Concepts 1
1.1.2 VRRP Application 3
1.1.3 Packet Structure 5
1.1.4 VRRP State 8
1.1.5 VRRP Election 10
1.1.6 VRRP Timer 10
1.1.7 VRRP Tracking 11
1.1.8 Protocols and Standards 11
1.2 Restrictions and Guidelines 11
1.3 Configuration Task Summary 11
1.4 Configuring IPv4 VRRP 12
1.4.1 Overview 12
1.4.2 Restrictions and Guidelines 12
1.4.3 Procedure 12
1.5 Configuring IPv6 VRRP 13
1.5.1 Overview 13
1.5.2 Restrictions and Guidelines 13
1.5.3 Procedure 14
1.6 Configuring IPv4 VRRP Tracking 15
1.6.1 Overview 15

1.6.2 Restrictions and Guidelines 15
1.6.3 Procedure 15
1.7 Configuring IPv6 VRRP Tracking 16
1.7.1 Overview 16
1.7.2 Restrictions and Guidelines 16
1.7.3 Procedure 17
1.8 Configuring VRRP Attributes 18
1.8.1 Overview 18
1.8.2 Configuration Tasks 18
1.8.3 Notes 18
1.8.4 Configuring VRRP Basic Attributes 18
1.8.5 Configuring a Method of Sending IPv4 VRRP Packets on a Super VLAN Port 19
1.8.6 Configuring the Dual-Active Mode for an IPV4 VRRP Group 20
1.8.7 Configuring the IPv4 VRRP Packet Version 21
1.9 Monitoring 21
1.10 Configuration Examples 23
1.10.1 Configuring IPv4 VRRP 23
1.10.2 Configuring VRRP + MSTP 27

1 Configuring VRRP

1.1 Introduction

Virtual Router Redundancy Protocol (VRRP) is a fault-tolerant routing protocol. When a device used for routing and forwarding on a local area network (LAN) fails, VRRP enables another device to automatically take over services and data of the device, which helps achieve hot backup and fault tolerance of IP routing. In addition, VRRP ensures communication continuity and reliability for the hosts on the LAN. VRRP is applicable to LAN scenarios which require redundant backup of routing egresses.

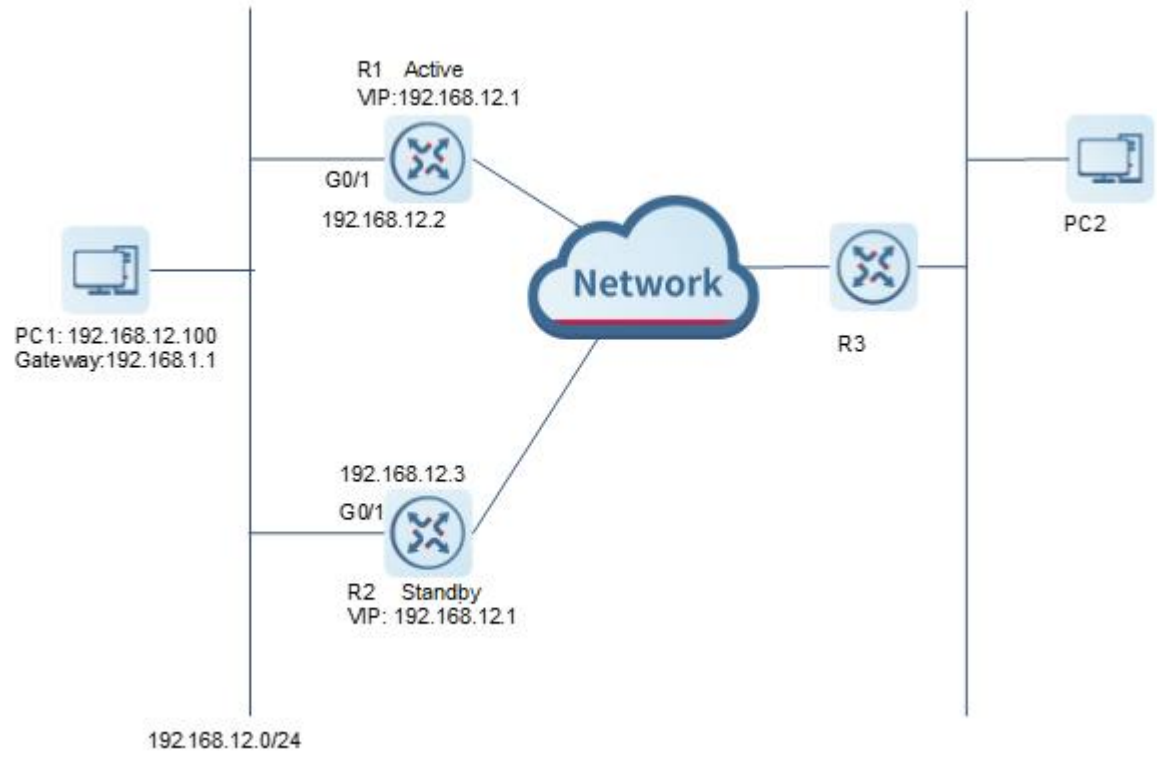
Note

In this chapter, routers refer to the devices supporting routing, such as Layer 3 switches, routers and firewalls.

1.1.1 Basic Concepts

As shown in Figure 1-1, router 1 and router 2 form a VRRP backup group, and the gateway IP address configured for PC 1 is the virtual IP address of the VRRP backup group.

Figure 1-1 Diagram of Configuring VRRP



1. VRRP Router

A VRRP router refers to a router running VRRP. It is composed of one or more virtual routers.

2. Virtual Router

A virtual router, also called a VRRP standby group, is usually regarded as the default gateway of hosts on a shared LAN. A VRRP standby group contains a virtual router identifier (VRID) and a set of virtual IP addresses.

3. Active Router

In a VRRP standby group, only the active router responds to Address Resolution Protocol (ARP) requests and forwards IP packets. If a device is the IP address owner, it becomes the active router.

4. Standby Router

In a VRRP standby group, a standby router only monitors the state of the active router but does not respond to ARP requests or forward IP packets. When the active router fails, the standby router takes the chance to become the active router via election.

5. VRID

A virtual router identifier (VRID) is the unique identifier of a virtual router and distinguishes different VRRP back groups. Only a group of routers with the same VRID can form a VRRP standby group.

6. Priority

A VRRP standby group determines the position of each router in the virtual router based on its priority. The configurable priority range is from 1 to 254, and the priority of the VRRP router with the virtual router IP address must be 255. When the priority is 0, the current active node no longer joins the VRRP standby group, and the standby router is triggered to become the active node rapidly without waiting for the current active node to time out.

7. Virtual IP Address

A virtual IP address is the IP address of a virtual router. A virtual router can be configured with one or multiple virtual IP addresses.

8. Virtual MAC Address

A virtual MAC address of VRRP complies with the Request for Comments (RFC) protocol standard. The virtual MAC address of an IPv4 VRRP standby group is "00-00-5E-00-01-{VRID}" (the virtual MAC address of an IPv6 VRRP standby group is “00-00-5E-00-02-{VRID}”), with the first five bytes fixed and the last byte of a VRRP standby group ID. A VRRP standby group responds to an ARP request by using its virtual MAC address instead of the real MAC address of an interface.

9. IP Address Owner

If the virtual IPv4/IPv6 address of a VRRP standby group is consistent with the IPv4/IPv6 address of the Ethernet port on which the VRRP standby group is configured, that is, the VRRP standby group has the real IP address of the Ethernet port, the VRRP standby group is collectively called an IP address owner. In such a case, the priority of the VRRP standby group is 255. If the Ethernet port is available, the VRRP standby group enters the active state automatically. The IP address owner receives and processes the packets whose destination IP address is the IP address of the virtual router.

10. Preemption Mode

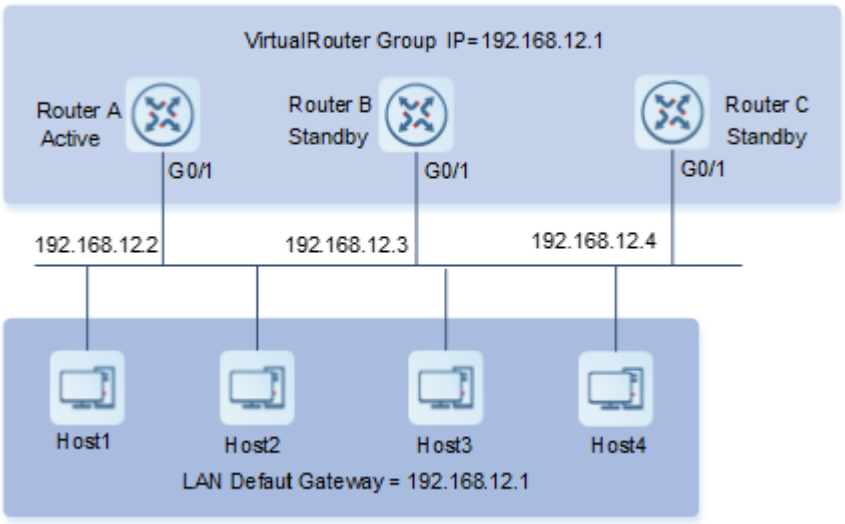
If a VRRP standby group runs in preemption mode, a standby router with a higher priority replaces the active router with a lower priority to become the active router of the VRRP standby group. If the VRRP standby group

runs in the non-preemption mode, as long as the active router runs properly, the standby router does not become the active router even if it is configured with a higher priority later. In some cases, even if the non-preemption mode is configured, a router that is just started in a VRRP standby group will preempt the position of the VRRP active router. The reason is that, when a device is started or the port just becomes active, the VRRP standby group on the port fails to receive the VRRP packet sent by the active device of the same standby group in time. In this case, the above problem can be avoided by delaying the start of the VRRP standby group.

1.1.2 VRRP Application

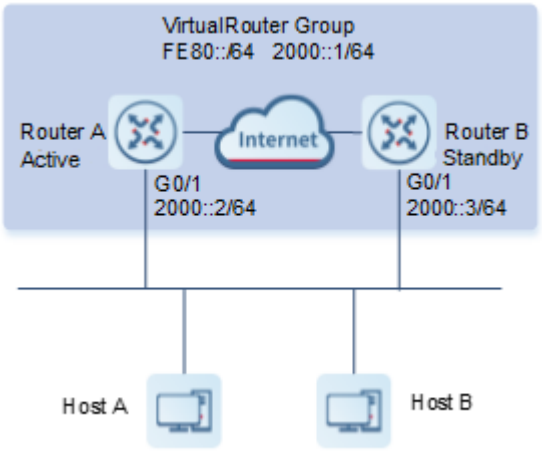
1. Active/Standby Redundancy

Figure 1-2 IPv4-based VRRP Active/Standby Redundancy



As shown in Figure 1-2, routers A, B, and C are all connected to the LAN through Ethernet ports, and VRRP is configured on the Ethernet ports connected to the LAN. The routers are in the same VRRP standby group, and the virtual IP address of the VRRP standby group is 192.168.12.1. Router A is elected as the active router of the VRRP standby group, and routers B and C are used as standby routers. Hosts 1, 2 and 3 use the IP address of the virtual router 192.168.12.1 as the gateway address. Packets from a host on the LAN to other networks are forwarded by the active router (router A). If router A fails, a active router is reelected between router B and router C to forward packets, to achieve simple routing redundancy.

Figure 1-3 IPv6-based VRRP Active/Standby Redundancy

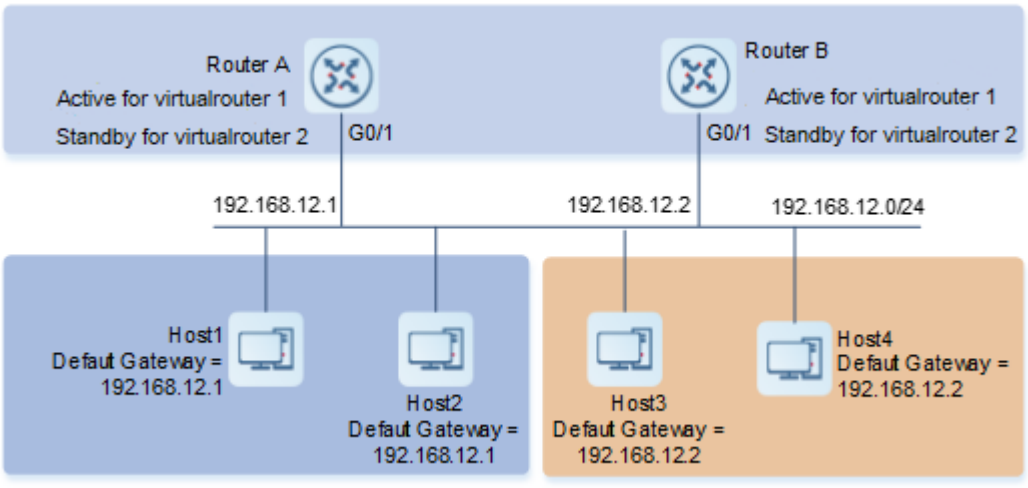


As shown in Figure 1-3, both routers A and B are connected to the LAN through Ethernet ports, and IPv6 VRRP is configured on the Ethernet ports connected to the LAN. The routers are in the same IPv6 VRRP standby group, and the virtual IPv6 addresses of the IPv6 VRRP standby group are FE80::1/64 and 2000::1/64. Router A is elected as the active router of VRRP, and router B functions as a standby router. The hosts on the LAN take the IPv6 link local address FE80::1/64 of the virtual router as the gateway address. Packets from the hosts on the LAN to other networks are forwarded by the active router. If router A fails, router B forwards packets on behalf of router A, achieving simple routing redundancy.

- Before implementation: If the uplink or downlink port of router A fails, the data flow is interrupted, that is, a single point of failure occurs.
- After implementation: If router A fails, VRRP implements switching within seconds to ensure that services are not interrupted. This effectively avoids the problem of network interruption after a single link fails. In addition, you do not need to modify the configuration such as the dynamic routing protocol and routing discovery protocol. You do not need to modify the default gateway configuration for the host node of local link.

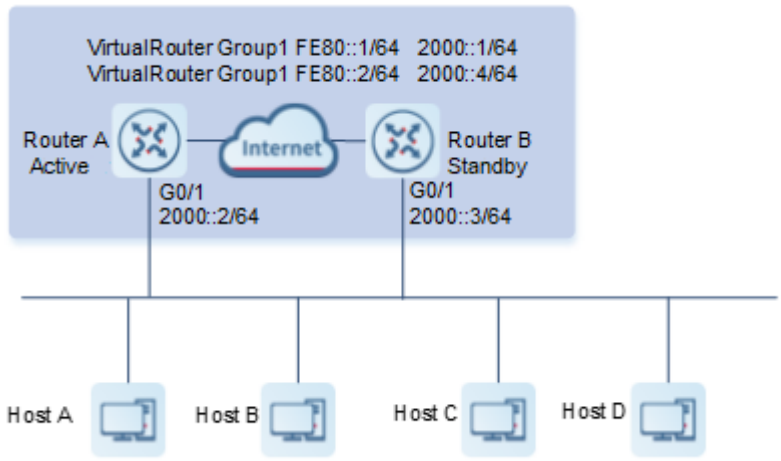
2. Load Balancing

Figure 1-4 IPv4-based VRRP Load Balancing



As shown in Figure 1-4, two virtual routers are configured. For virtual router 1, router A uses the IP address 192.168.12.1 of the Ethernet port GigabitEthernet 0/0 as the IP address of the virtual router, so router A becomes the active router and router B becomes the standby router. For virtual router 2, router B takes the IP address 192.168.12.2 of Ethernet port GigabitEthernet 0/0 as the IP address of the virtual router, so router B becomes the active router and router A becomes the standby router. In the LAN, hosts 1 and 2 take the IP address 192.168.12.1 of virtual router 1 as the default gateway address, while hosts 3 and 4 take the IP address 192.168.12.2 of virtual router 2 as the default gateway address. In this application of VRRP, routing redundancy is achieved between routers A and B, and the LAN traffic is shared between them to achieve load balancing.

Figure 1-5 IPv6-based VRRP Load Balancing



As shown in Figure 1-5, two virtual routers are configured. For virtual router 1, router A uses the IP address fe80::1/64 of the Ethernet port GigabitEthernet 0/1 as the IP address of the virtual router, so router A becomes the active router and router B becomes the standby router. For virtual router 2, router B takes the IP address fe80::2/64 of Ethernet port GigabitEthernet 0/1 as the IP address of the virtual router, so router B becomes the active router and router A becomes the standby router. On the LAN, hosts A and B use the IP address FE80::1 of virtual router 1 as the default gateway address (manually configured static IPv6 gateway), and hosts C and D use the IP address FE80::2 of virtual router 2 as the default gateway address (manually static configured IPv6 gateway). In this way, routing redundancy is achieved between routers A and B, and the LAN traffic is shared between them to achieve load balancing.

The devices involved are L3 devices or routers. L2 devices are not applicable.

- Before implementation: All traffic goes through the active router, causing a certain burden to the active router. Before the active router fails, the standby router has been always idle, leading to a resource waste.
- After implementation: Each device is not a single active router or standby router, but acts as different roles in different groups. Therefore, the connected devices can be assigned to different groups and forward traffic to different gateways to achieve traffic load balancing. Moreover, no device is kept idle all the time, and resources are fully utilized.

1.1.3 Packet Structure

VRRP has two versions available: VRRPv2 and VRRPv3. VRRPv2 is only applicable to IPv4 networks, and VRRPv3 is applicable to both IPv4 and IPv6 networks. A VRRP packet is used to send the priority and state of

the active device to all standby devices in the same standby group in multicast mode (with the destination address 224.0.0.18). The packet is encapsulated in an IP packet. The source address is the primary IP address (not the virtual IP address) of the packet sending port, the time to live (TTL) is 255, and the protocol ID is 112. Figure 1-6 and Figure 1-7 show the structures of VRRPv2 and VRRPv3 packets respectively.

Figure 1-6 Structure of VRRPv2 Packets

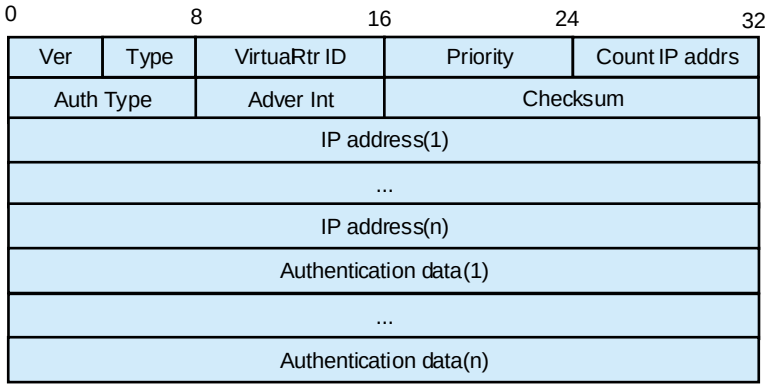


Figure 1-7 Structure of VRRPv3 Packets

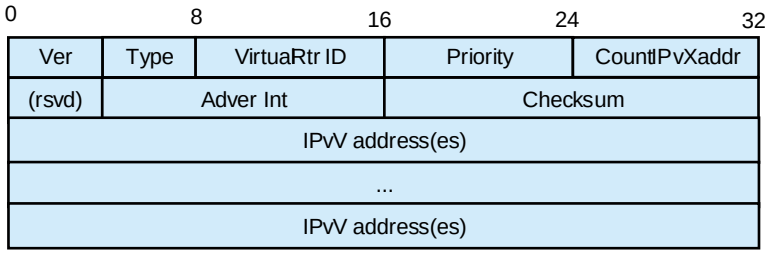


Table 1-1 Fields of VRRP Packets and Description

Packet Field	Description	
	VRRPv2	**VRRPv3**
Ver	Protocol version number. The value is 2.	Protocol version number. The value is 3.
Type	VRRP packet type. The value is 1, indicating that only the advertisement packets of VRRP are supported.	VRRP packet type. The value is 1, indicating that only the advertisement packets of VRRP are supported.
Virtual Rtr ID (VRID)	Virtual router ID (also the standby group ID). The value range is from 1 to 255.	Virtual router ID (also the standby group ID). The value range is from 1 to 255.

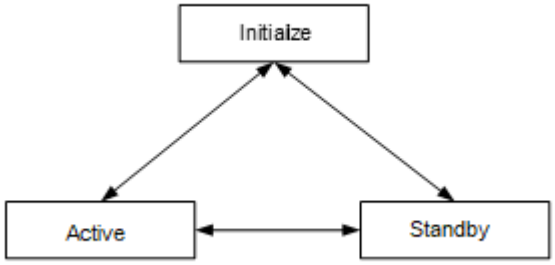
Packet Field	Description	
	VRRPv2	**VRRPv3**
Priority	Priority of the router in the standby group. The value range is from 0 to 255, and the default value is **100**. A larger value indicates a higher priority. When the priority is 0, the current active node no longer joins the VRRP standby group and the standby router is triggered to become the active node, instead of waiting for the current active node to time out. The priority 255 is reserved for the IP address owner, and the configurable range is from 1 to 254 for other routers.	Priority of the router in the standby group. The value range is from 0 to 255, and the default value is **100**. A larger value indicates a higher priority. When the priority is 0, the current active node no longer joins the VRRP standby group and the standby router is triggered to become the active node, instead of waiting for the current active node to time out. The priority 255 is reserved for the IP address owner, and the configurable range is from 1 to 254 for other routers.
Count IP Addrs/Count IPvX Addr	Number of virtual IPv4 addresses in the standby group.	Number of virtual IPv4 or virtual IPv6 addresses in the standby group.
Auth Type	Authentication type. Three authentication types are defined: ● **0**: Indicates no authentication. ● **1**: Indicates simple password (plaintext) authentication. ● **2**: Indicates MD5 authentication.	-
Advertise Interval	Interval of sending advertisement packets, in seconds. The default value is **1** second.	Interval of sending advertisement packets, in centiseconds. The default value is **100** centiseconds, namely, **1** second.
Checksum	16-bit checksum, used to check data corruption in the VRRP packet.	16-bit checksum, used to check data corruption in the VRRP packet.
IP Address/IPvX Address(es)	Virtual IPv4 addresses in the standby group. The number of addresses contained is defined in the **Count IP Addrs** field.	Virtual IPv4 addresses or virtual IPv6 addresses in the VRRP standby group. The number of addresses contained is defined in the **Count IPvX Addrs** field.
Authentication Data	Authentication key of the VRRP packet. At present, the authentication key is only used for simple password authentication. For the other authentication methods, **0** is filled in this field.	-

Packet Field	Description	
	VRRPv2	VRRPv3
rsvd	-	Reserved field for the VRRP packet. The field must be set to **0**.

1.1.4 VRRP State

There are three VRRP states: initialize, standby, and active. Figure 1-8 shows the state transition process.

Figure 1-8 VRRP State Transition Diagram



1. Initialize

Initialize is the initial state of VRRP. The system enters this state after startup. It only responds to startup, that is, it does not process VRRP packets.

- When the system is started, if the priority of the local VRRP standby group is 255 (namely, the IP address owner), then:
  - VRRP packets are sent immediately.
  - If the local VRRP standby group is an IPv4 virtual router, gratuitous ARP packets are sent to all the IPv4 addresses of the virtual router. If the local VRRP standby group is an IPv6 virtual router, gratuitous NA packets are sent to all the IPv6 addresses of the virtual router, the routing flag bit is 1, the request flag bit is 0, the coverage flag bit is 1, the target address is the IPv6 address of the virtual router (not destination address), and the link layer address in the target link layer option is the MAC address of the virtual router.
  - The advertisement information timer is configured to send VRRP advertisements regularly, that is, **Adver_Timer** is set to **Advertisement_Interval**.
  - The local VRRP standby group Device to the active state.
- When the system is started, if the priority of the local VRRP standby group is not 255, then:
  - **Active_Adver_Interval** is set to **Advertisement_Interval**.
  - **Active_Down_Timer** is set to **Active_Down_Interval**.
  - The local VRRP standby group Device to the standby state.

2. Standby

A standby virtual router is used to monitor the validity and state of the active virtual router.

- An IPv4 standby virtual router does not respond to the ARP request for the IP address of the virtual router, and an IPv6 standby virtual router does not respond to the neighbor solicitation (NS) packet for the IPv6 address of the virtual router.
- The standby virtual router discards packets whose destination MAC addresses are the MAC address of the VRRP standby group and packets whose destination IP addresses are the IP address of the VRRP standby group.
- The standby virtual router does not send protocol packets. It must receive the VRRP multicast packets sent by the active router to know the state of the active router.
- When a shutdown event is received, the standby virtual router deletes the host timeout timer and Device to the initialize state.
- If a timeout event of the host timeout timer is received, the standby virtual router sends VRRP advertisement information immediately. If the standby virtual router is an IPv4 virtual router, it broadcasts ARP packets to all the IPv4 addresses of the virtual router. If the standby virtual router is an IPv6 virtual router, it computes and joins the solicited-node multicast address for the IPv6 address associated with the virtual router, and sends non-request (gratuitous) NA packets to all the IPv6 addresses of the virtual router. The routing flag bit is 1, the request flag bit is 0, the coverage flag bit is 1, the target address is the IPv6 address of the virtual router (not destination address), and the link layer address in the target link layer option is the MAC address of the virtual router. **Adver_Timer** is set to **Advertisement_Interval**, and the virtual router Device to the active state.
- If a VRRP packet is received and the priority is 0, the host timeout timer **Active_Down_Timer** is reset to **Skew_Time**; if the priority of the received VRRP packet is greater than 0, and preemption is not allowed or the priority of the packet is greater than or equal to the priority of the local virtual router, **Adver Interval** is set to the **Active_Adver_Interval** field value in the packet and **Active_Down_Interval** is set to **Active_Down_Timer**, otherwise the packet is discarded.

3. Active

As the forwarding role in a VRRP standby group, the virtual router in the active state forwards virtual gateway packets.

- It sends VRRP multicast packets regularly, and forwards relevant packets passing through the virtual IP address and packets whose destination MAC address is the VRRP virtual MAC address.
- If the virtual router is an IPv4 virtual router, the VRRP virtual MAC address must be used to respond to the ARP requests whose destination IP addresses are the virtual IP address of the virtual router. If the virtual router is an IPv6 virtual router, it must respond to the NS packets whose destination IP addresses are the virtual IPv6 address of the virtual router and must be a member of the solicited-node multicast address for the IPv6 address associated with the virtual router.
- If **Accept_Mode** or **IPvX Owner** is configured, that is, the virtual router is the IPv4/IPv6 address owner, the packets whose destination IP addresses are the virtual IP address of the virtual router must be received, otherwise the packets are discarded.
- When the network port where the VRRP virtual router is located receives a shutdown event, the virtual router deletes the regular advertisement timer, removes the value of **Adver_Timer**, immediately sends a VRRP

advertisement packet with priority 0, and Device to the initialize state.

- If the timer for sending VRRP packets times out or a VRRP packet with priority 0 is received, the VRRP advertisement message is sent immediately, the timer for sending VRRP packets is reset, and **Adver_Timer** is assigned to **Advertisement_Interval**.
- If the priority of the received VRRP packet is greater than or equal to that of the corresponding VRRP virtual router at the local end and the source IP address of the received packet is greater than the primary IP address at the local end, the host timeout timer is reset, the regular advertisement timer is configured, and the virtual router Device to the standby state.
- If none of the above requirements is met, the packets are discarded.

1.1.5 VRRP Election

VRRP provides a simple mechanism for electing the active router, which provides the actual routing and forwarding service. The active router is elected in the following order:

(1) VRRP judges whether the VRRP device is an IP address owner. If so, VRRP Device to the active state.

(2) VRRP judges the VRRP priority in the same VRRP standby group. The device with a larger priority is in the active state.

(3) VRRP judges the primary IP address of the network port. The device with a larger primary IP address changes to the active state.

(4) The standby device determines whether to switch the state by monitoring the VRRP packet regularly sent by the active device. It Device the state in the following two cases:

- If the received advertisement state of the active device is normal, the priority of the active device's advertisement is lower than that of the standby device, and the preemption mode has been configured, the standby device Device to the active state.
- If no state advertisement is received from the active device within the timeout period and the preemption mode is configured, the standby device Device to the active state.

1.1.6 VRRP Timer

The main VRRP timers include the VRRP advertisement sending interval timer and the VRRP preemption delay timer.

- VRRP advertisement sending interval timer

You can adjust the time interval for the active router to send VRRP advertisement packets by configuring the VRRP advertisement sending interval timer. If the standby router still fails to receive a VRRP advertisement packet after waiting for at least three intervals, it regards itself as the active router, and sends a VRRP advertisement packet to reelect the active router.

- VRRP preemption delay timer

To avoid frequent active/standby switching between the members of the standby group and allow the standby router to have enough time to collect necessary information (such as the routing information), the standby router does not immediately preempt the position of the active router after receiving the advertisement packet with a priority lower than the local priority, but waits for a period of time (preemption delay) before sending a VRRP advertisement packet to take the place of the original active router.

1.1.7 VRRP Tracking

By correlating with the tracking function, VRRP monitors port state changes and the IP host accessibility and automatically adjusts the priority of the standby group to trigger role competition in the standby group, which causes state switching in the standby group and achieves fault monitoring and automatic switching. In this way, the network traffic is not interrupted. Two detection methods are available:

- Port state: VRRP detects the port state by receiving port state change messages. When a port is no longer in the up state, the priority of the current active router is automatically reduced to enable other standby routers to have the opportunity to become the active router. This avoids the failure of the entire virtual router due to the disconnection of a certain line.
- Host accessibility: Internet Control Message Protocol (ICMP) detection packets are sent regularly, and the time interval for sending detection packets, timeout time for waiting for responses, and continuous timeout count are configured to judge accessibility of the destination addresses.

⚠ Caution

- A monitored port must be a routable L3 logical port (for example, a routed port, an SVI, a loopback port, or a tunnel port). If a VRRP standby group is an IPv6 group, the IPv6 function needs to be enabled on the port first.
- The decrease in the priority value must be greater than the priority difference between the active and standby devices; otherwise the priority of the active device may be the same as that of the standby device after priority adjustment, and the active/standby state cannot be switched.

1.1.8 Protocols and Standards

- RFC 2338: Virtual Router Redundancy Protocol
- RFC 3768: Virtual Router Redundancy Protocol (VRRP)
- RFC 5798: Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6

1.2 Restrictions and Guidelines

- The IPv6 VRRP and IPv4 VRRP standby groups share the VRRP standby group ID. One VRRP standby group ID is applicable to an IPv4 VRRP standby group and an IPv6 VRRP standby group configured on the same port.
- When the IPv4 VRRP and IPv6 VRRP standby groups of the same port use the same VRRP standby group ID, two VRRP standby group IDs are used.

1.3 Configuration Task Summary

VRRP configuration includes the following tasks:

(1) Configuring IPv4 VRRP

(2) Configuring IPv6 VRRP

(3) Configuring IPv4 VRRP Tracking

(4) Configuring IPv6 VRRP Tracking

(5) (Optional) Configuring VRRP Attributes

All the following configuration tasks are optional and may be selected as needed.

- Configuring VRRP Basic Attributes
- Configuring a Method of Sending IPv4 VRRP Packets on a Super VLAN Port
- Configuring the Dual-Active Mode for an IPV4 VRRP Group
- Configuring the IPv4 VRRP Packet Version

1.4 Configuring IPv4 VRRP

1.4.1 Overview

IPv4 VRRP creates a standby group on a specified LAN segment after the standby group ID and virtual IPv4 address are configured, and then the VRRP single standby function is enabled on the corresponding port. You can configure multiple VRRP standby groups on the same Ethernet port to achieve load balancing and offer more stable and reliable network services through mutual standby.

1.4.2 Restrictions and Guidelines

- To achieve VRRP, the routers in a VRRP standby group must be configured with the same virtual IPv4 address.
- To achieve mutual standby between multiple IPv4 VRRP standby groups, configure multiple same IPv4 VRRP standby groups with different priorities on ports so that they act as the active and standby groups mutually.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN router interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable IPv4 VRRP.

vrrp-config *group-id* **ip** *ipv4-address* [**secondary**]

IPv4 VRRP is disabled on a port by default.

(5) Configure a priority for the IPv4 VRRP standby group.

vrrp-config *group-id* **priority** *priority*

The default priority of an IPv4 VRRP standby group is **100**.

(6) (Optional) Configure a VRRP packet standard for the IPv4 VRRP standby group.

vrrp-config *group-id* **version** { **2** | **3** }

IPv4 VRRP adopts the VRRPv2 standard by default.

(7) (Optional) Configure an authentication string for the IPv4 VRRP standby group.

vrrp-config *group-id* **authentication** *authentication-string*

No authentication string of IPv4 VRRP is configured by default.

(8) (Optional) Configure a name for the IPv4 VRRP standby group.

vrrp-config *group-id* **description** *group-name*

No VRRP group name is configured by default.

(9) (Optional) Configure a VRRP advertisement interval for the IPv4 VRRP active router.

vrrp-config *group-id* **timers advertise** { *advertise-interval* | **csec** *centisecond-interval* }

The default VRRP advertisement interval of the VRRP active router is **1** second.

(10) (Optional) Enable the IPv4 VRRP timer learning function.

vrrp-config *group-id* **timers learn**

The advertisement interval timer learning function is disabled for an IPv4 VRRP standby group by default.

(11) (Optional) Configure the preemption mode for the IPv4 VRRP standby group.

vrrp-config *group-id* **preempt** [**delay** *delay-seconds*]

IPv4 VRRP runs in the preemption mode without preemption delay by default.

(12) (Optional) Configure a startup delay for the VRRP standby group on the port.

vrrp-config delay { **protocol-up** *min-seconds* | **reload** *reload-seconds* }

No VRRP group startup delay is configured by default.

1.5 Configuring IPv6 VRRP

1.5.1 Overview

IPv6 VRRP creates a standby group on a specified LAN segment after the standby group ID and virtual IPv6 address are configured, and then the VRRP single standby function is enabled on the corresponding port. You can configure multiple VRRP standby groups on the same Ethernet port to achieve load balancing and offer more stable and reliable network services through mutual standby.

1.5.2 Restrictions and Guidelines

- The first configured virtual address for IPv6 VRRP must be a link-local address, which can be deleted only

after other virtual addresses are deleted.

- The IPv6 function must be enabled on a port before IPv6 VRRP is configured.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN router interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable the IPv6 function on the port.

ipv6 enable

(5) Enable IPv6 VRRP.

vrrp-config *group-id* **ipv6** *ipv6-address*

IPv6 VRRP is disabled on a port by default.

(6) Enable the function of receiving packets whose destination addresses are the IPv6 address of the virtual router.

vrrp-config ipv6 accept_mode

By default, an IPv6 VRRP group in the active state is not permitted to receive packets whose destination addresses are the IPv6 address of a virtual router, except the IPv6 VRRP group in owner state or the received NA/NS packets.

(7) Configure a priority for the IPv6 VRRP standby group.

vrrp-config ipv6 *group-id* **priority** *priority*

The default priority of an IPv6 VRRP standby group is **100**.

(8) (Optional) Configure a name for the IPv6 VRRP standby group.

vrrp-config ipv6 *group-id* **description** *group-name*

No VRRP group name is configured by default.

(9) (Optional) Configure a VRRP advertisement interval for the IPv6 VRRP active router.

vrrp-config ipv6 *group-id* **timers advertise** { *advertise-interval* | **csec** *centisecond-interval* } [**multiplier** *interval*]

The default VRRP advertisement interval of the VRRP active router is **1** second, and the timeout time is three times of the interval.

(10) (Optional) Enable the IPv6 VRRP timer learning function.

vrrp-config ipv6 *group-id* **timers learn**

The advertisement interval timer learning function is disabled for an IPv6 VRRP group by default.

(11) (Optional) Configure the preemption mode for the IPv6 VRRP standby group.

vrrp-config ipv6 *group-id* **preempt** [**delay** *delay-seconds*]

IPv6 VRRP runs in the preemption mode without preemption delay by default.

1.6 Configuring IPv4 VRRP Tracking

1.6.1 Overview

IPv4 VRRP tracking monitors the changes in port state and the IPv4 host accessibility. You can configure the optional packet detection parameter and bidirectional forwarding detection (BFD) correlation. The priority of a standby group can be adjusted dynamically to implement automatic switching and recovery of the standby group in case of a failure monitored.

1.6.2 Restrictions and Guidelines

- A monitored port must be a routable L3 logical port (for example, a routed port, an SVI, a loopback port, or a tunnel port).
- The support for BFD depends on the product version.
- If a VRRP group uses the actual IP address of an Ethernet port (in the owner state), the group priority is 255, and the monitored IP address or port can be configured, but the priority of the VRRP group is not changed.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN router interface configuration mode.

  interface overlayrouter *interface-number*

(4) Configure a port to be monitored by an IPv4 VRRP standby group.

vrrp-config *group-id* **track** { *interface-type interface-number* | **bfd** *interface-type interface-number ipv4-address* } [*priority*]

No port to be monitored by an IPv4 VRRP standby group is configured by default. If a port to be monitored by an IPv4 VRRP standby group is configured and the priority parameter is not specified, the priority change value is 10.

(5) Configure an IP address to be monitored by the IPv4 VRRP standby group.

vrrp-config *group-id* **track** *ipv4-address* [**interval** *interval-value*] [**timeout** *timeout-value*] [**retry** *retry-value*] [*priority*]

No IP address to be monitored by an IPv4 VRRP standby group is configured by default. If an IP address to be monitored by an IPv6 VRRP group is configured, the time interval for sending detection packets is 3 seconds, and the timeout time for waiting for a response to a sent detection packet is 1 second, and the consecutive timeout count for judging that a tracked IP address is unreachable is 3.

(6) Configure BFD correlation with the IPv4 VRRP group.

vrrp-config *group-id* **bfd** *ipv4-address*

BFD correlation with an IPv4 VRRP group is not configured on a port by default.

(7) Return to the global configuration mode.

exit

(8) Configure global IPv4 VRRP BFD.

vrrp-config bfd *interface-type interface-number ipv4-address*

By default, no global BFD mode is configured for IPv4 VRRP to detect whether the active device is active.

1.7 Configuring IPv6 VRRP Tracking

1.7.1 Overview

IPv6 VRRP tracking monitors the changes in port state and the IPv6 host accessibility. You can configure the optional packet detection parameter to dynamically adjust the priority of a standby group and implement automatic switching and recovery of the standby group in case of a failure monitored.

1.7.2 Restrictions and Guidelines

- A monitored port must be a routable L3 logical port (for example, a routed port, an SVI, a loopback port, or a tunnel port).
- The IPv6 function needs to be first enabled on a port.
- If a tracked host IP address is a link-local address, specify a network port.
- If a VRRP group (in the owner state) uses the actual IP address of an Ethernet port, the group priority is 255,

and the monitored IP address or port can be configured, but the priority of the VRRP group is not changed.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the SVI configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN router interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable the IPv6 function on the port.

ipv6 enable

(5) Configure a port to be monitored by an IPv6 VRRP group.

vrrp-config ipv6 *group-id* **track** { *interface-type interface-number* | **bfd** *interface-type interface-number peer-ipv6-address* } [*priority*]

No port to be monitored by an IPv6 VRRP group is configured by default. If a port to be monitored by an IPv6 VRRP group is configured and the priority parameter is not specified, the priority change value is 10.

(6) Configure an IPv6 address to be monitored by the IPv6 VRRP standby group.

vrrp-config ipv6 *group-id* **track** { *ipv6-global-address* | *ipv6-linklocal-address interface-type interface-number* } [**interval** *interval-value*] [**timeout** *timeout-value*] [**retry** *retry-value*] [*Priority*]

No IP address to be monitored by an IPv6 VRRP group is configured by default. If an IP address to be monitored by an IPv6 VRRP group is configured, the time interval for sending detection packets is 3 seconds, the timeout time for waiting for a response to a sent detection packet is 1 second, and the consecutive timeout count for judging that a tracked IP address is unreachable is 3.

(7) Configure correlation between an IPv6 VRRP group and BFD.

vrrp-config ipv6 *group-id* **bfd** *ipv6-address*

Correlation between an IPv6 VRRP group and BFD is not configured on an interface by default.

(8) Return to the global configuration mode.

exit

(9) Configure global IPv6 VRRP BFD.

vrrp-config ipv6 bfd *interface-type interface-number ipv6-address*

No global BFD mode is configured for IPv6 VRRP to detect whether the active device is active by default.

1.8 Configuring VRRP Attributes

1.8.1 Overview

If a VRRP standby group has been set up, you can configure enhancement to increase the usability of the VRRP standby group.

1.8.2 Configuration Tasks

The VRRP attribute configuration includes the following tasks:

All the following configuration tasks are optional and may be selected as needed.

- Configuring VRRP Basic Attributes
- Configuring a Method of Sending IPv4 VRRP Packets on a Super VLAN Port
- Configuring the Dual-Active Mode for an IPV4 VRRP Group
- Configuring the IPv4 VRRP Packet Version

1.8.3 Notes

This section describes notes for configuring VRRP attribute functions.

Table 1-2 describes the VRRP attribute functions supported by different products.

Table 1-2 VRRP Attribute Functions Supported by Different Products

Function	Product Supporting the Function
Configuring a Method of Sending IPv4 VRRP Packets on a Super VLAN Port	S6920-4C, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ

1.8.4 Configuring VRRP Basic Attributes

1. Overview

If a VRRP standby group has been set up, you can configure basic attributes according to the actual situation.

2. Prerequisites

A VRRP standby group has been set up.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure VRRP basic attributes according to the actual situation.

- Configure a priority for the IPv4 VRRP standby group.

  vrrp-config *group-id* **priority** *priority*

  The default priority of an IPv4 VRRP standby group is **100**.

- Configure an authentication string of IPv4 VRRP.

  vrrp-config *group-id* **authentication** *authentication-string*

  No authentication string of IPv4 VRRP is configured by default.

- Configure a VRRP advertisement interval for the IPv4 VRRP active router.

  vrrp-config *group-id* **timers advertise** { *advertise-interval* | **csec** *centisecond-interval* } [**multiplier** *interval*]

  The default VRRP advertisement interval of the VRRP active router is **1** second, and the timeout time is three times of the interval.

- Configure the preemption mode for the IPv4 VRRP standby group.

  vrrp-config *group-id* **preempt** [**delay** *delay-seconds*]

  IPv4 VRRP runs in the preemption mode without preemption delay by default.

- Enable the IPv4 VRRP timer learning function.

  vrrp-config *group-id* **timers learn**

  The advertisement interval timer learning function is disabled for an IPv4 VRRP standby group by default.

- Configure a name for the IPv4 VRRP group.

  vrrp-config *group-id* **description** *group-name*

  No VRRP group name is configured by default.

- Configure a startup delay for the VRRP standby group on a port.

  vrrp-config delay { **protocol-up** *min-seconds* | **reload** *reload-seconds* }

  No VRRP group startup delay is configured by default.

1.8.5 Configuring a Method of Sending IPv4 VRRP Packets on a Super VLAN Port

Note

Only S6920-4C, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ support configuring a method for sending IPv4 VRRP packets on a super VLAN interface.

1. Overview

The IPv4 VRRP protocol packets can be sent by the following three methods in a super VLAN. You can configure the method of sending advertisement packets.

- Sending IPv4 VRRP packets only to the first up sub VLAN in a super VLAN.
- Sending IPv4 VRRP packets to a specified sub VLAN in a super VLAN.
- Sending IPv4 VRRP packets to all the sub VLANs in a super VLAN.

2. Restrictions and Guidelines

- The configuration is effective only on super VLAN interfaces.
- If VRRP and VRRP plus are both enabled on a super VLAN interface, VRRP packets are sent to all the up sub-VLAN interfaces of the super VLAN interface.

3. Prerequisites

VRRP has been enabled on a super VLAN port.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI configuration mode.

interface vlan *vlan-id*

(4) Configure a sub-VLAN range.

subvlan *vlan-range*

No sub-VLAN range is configured by default.

(5) Configure a method of sending IPv4 VRRP packets on a super VLAN port.

vrrp-config detection-vlan { **first-subvlan** | *subvlan-id* }

The IPv4 VRRP packets on a super VLAN port are sent to the first up sub VLAN only by default.

1.8.6 Configuring the Dual-Active Mode for an IPV4 VRRP Group

1. Overview

A VRRP group works in active/standby mode by default. To set both ends of a VRRP group in the active state and avoid sending keepalive packets, configure the dual-active mode for the VRRP group.

2. Restrictions and Guidelines

- The dual-active mode can be configured only on an SVI.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the SVI configuration mode.

interface vlan *vlan-id*

(4) Configure a VRRP group to work in dual-active state and configure the VRRP group not to send keepalive packets.

vrrp-config mode dual-active

A VRRP group works in active/standby mode by default.

1.8.7 Configuring the IPv4 VRRP Packet Version

1. Overview

The devices belonging to the same VRRP group should send packets according to the unified packet protocol: VRRPv2 or VRRPv3. By default, VRRPv2 is used for sending packets. VRRPv3 routers can recognize VRRPv2 packets, but VRRPv2 routers cannot recognize VRRPv3 packets. Configuring this function enables VRRPv3 routers to send VRRPv2 packets so as to communicate with VRRPv2 routers.

2. Restrictions and Guidelines

- The configuration is applicable to IPv4 VRRP only.
- If simple authentication has been configured for a VRRP group and the VRRP group is configured to send VRRPv2 packets, the authentication information is carried. If a VRRP group is configured to send VRRPv3 packets, the authentication information is not carried.
- The configuration only applies to the sending mode of VRRP packets and does not affect the receiving mode. For example, a VRRPv3 router can identify both VRRPv2 and VRRPv3 packets.

3. Prerequisites

IPv4 VRRP has been configured.

4. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the VRRP packet standard for an IPv4 VRRP standby group.

vrrp-config *group-id* **version** { **2** | **3** }

IPv4 VRRP adopts the VRRPv2 standard by default.

1.9 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

You can run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-3 VRRP Monitoring

Command	Purpose
display [**ipv6**] **vrrp-config** [**brief** | *group-id*]	Displays the brief or detailed information of IPv4/IPv6 VRRP.
display [**ipv6**] **vrrp-config interface** *interface-type interface-number* [**brief**]	Displays the information of an IPv4/IPv6 VRRP group on a specified port.
display vrrp-config [**brief** | *group-id*]	Displays the brief or detailed information of IPv4 VRRP.
display vrrp-config interface *interface-type interface-number* [**brief**]	Displays the information of an IPv4 VRRP group on a specified port.
display vrrp-config packet statistics [*interface-type interface-number* | **total**]	Displays the statistics of VRRP packets.
debugging [**ipv6**] **vrrp-config**	Debugs VRRP errors, events, packets, and state.
debugging [**ipv6**] **vrrp-config errors**	Debugs VRRP errors.
debugging [**ipv6**] **vrrp-config events**	Debugs VRRP events.
debugging vrrp-config packets [**acl** *acl-id* **|** [**icmp** | **protocol**] **interface** *interface-type interface-number* [*group-id*]] **debugging ipv6 vrrp-config packets** [**acl** *acl-name* | [**icmp** | **protocol**] **interface** *interface-type interface-number* [*group-id*]]	Debugs VRRP packets.
debugging [**ipv6**] **vrrp-config state**	Debugs VRRP state.
debugging vrrp-config	Debugs VRRP errors, events, packets, and state.
debugging vrrp-config errors	Debugs VRRP errors.
debugging vrrp-config events	Debugs VRRP events.
debugging vrrp-config packets [**acl** *acl-id* **|** [**icmp** | **protocol**] **interface** *interface-type interface-number* [*group-id*]]	Debugs VRRP packets.
debugging vrrp-config state	Debugs VRRP state.

1.10 Configuration Examples

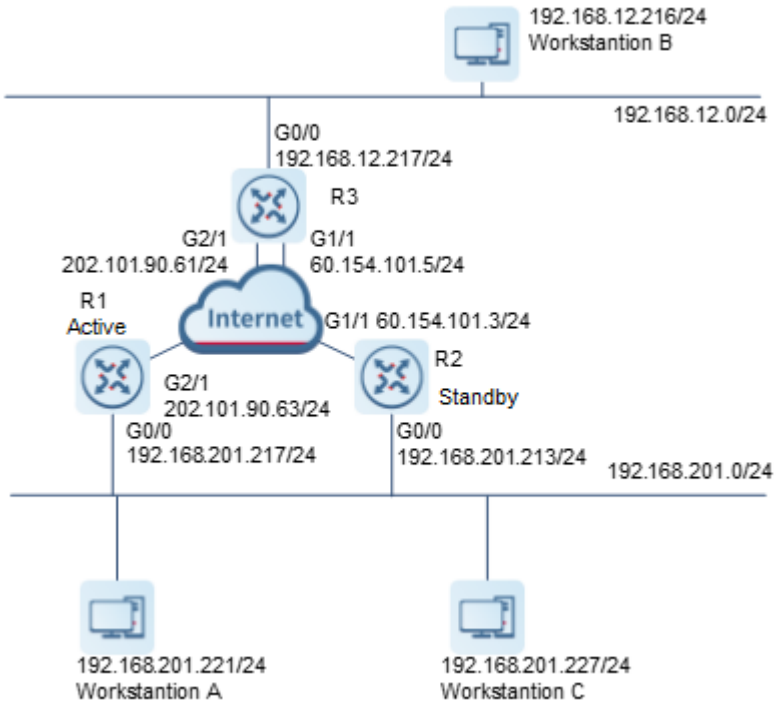
1.10.1 Configuring IPv4 VRRP

1. Requirements

An IPv4 VRRP group and a to-be-monitored port need to be configured.

2. Topology

Figure 1-9 Configuring IPv4 VRRP



3. Notes

- Configure the work station cluster (192.168.201.0/24) to use the standby group composed of routers R1 and R2 and set the gateway address to the virtual router IP address configured for the standby group (192.168.201.1) so that the work station cluster accesses the remote work station cluster (192.168.12.0 /24) through the virtual router 192.168.201.1.
- On router R1, configure GigabitEthernet 2/1 as a VRRP monitored port.
- No VRRP but an ordinary routing function is configured on router 3.

4. Procedure

Perform the following configuration on router 3:

```
R3> privilege
R3# config-system terminal
R3(cfg)# interface GigabitEthernet 0/0
R3(cfg-if-GigabitEthernet 0/0)# unset switch-port
```

```
R3(cfg-if-GigabitEthernet 0/0)# ip address 192.168.12.217 255.255.255.0
R3(cfg-if-GigabitEthernet 0/0)# exit
R3(cfg)# interface GigabitEthernet 1/1
R3(cfg-if-GigabitEthernet 1/1)# unset switch-port
R3(cfg-if-GigabitEthernet 1/1)# ip address 60.154.101.5 255.255.255.0
R3(cfg-if-GigabitEthernet 1/1)# exit
R3(cfg)# interface GigabitEthernet 2/1
R3(cfg-if-GigabitEthernet 2/1)# unset switch-port
R3(cfg-if-GigabitEthernet 2/1)# ip address 202.101.90.61 255.255.255.0
R3(cfg-if-GigabitEthernet 2/1)# exit
R3(cfg)# router-protocol ospf
R3(cfg-router)# network-advertise 202.101.90.0 0.0.0.255 area 10
R3(cfg-router)# network-advertise 192.168.12.0 0.0.0.255 area 10
R3(cfg-router)# network-advertise 60.154.101.0 0.0.0.255 area 10
```

Perform the following configuration on router 1:

```
R1> privilege
R1# config-system terminal
R1(cfg)# interface GigabitEthernet 0/0
R1(cfg-if-GigabitEthernet 0/0)# ip address 192.168.201.217 255.255.255.0
R1(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 priority 120
R1(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 timers advertise 3
R1(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 ip 192.168.201.1
R1(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 track GigabitEthernet 2/1 30
R1(cfg-if-GigabitEthernet 0/0)# exit
R1(cfg)# interface GigabitEthernet 2/1
R1(cfg-if-GigabitEthernet 2/1)# ip address 202.101.90.63 255.255.255.0
R1(cfg-if-GigabitEthernet 2/1)# exit
R1(cfg)# router-protocol ospf
R1(cfg-router)# network-advertise 202.101.90.0 0.0.0.255 area 10
R1(cfg-router)# network-advertise 192.168.201.0 0.0.0.255 area 10
```

Perform the following configuration on router 2:

```
R2> privilege
R2# config-system terminal
R2(cfg)# interface GigabitEthernet 0/0
R2(cfg-if-GigabitEthernet 0/0)# ip address 192.168.201.213 255.255.255.0
R2(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 ip 192.168.201.1
R2(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 timers advertise 3
R2(cfg-if-GigabitEthernet 0/0)# exit
R2(cfg)# interface GigabitEthernet 1/1
R2(cfg-if-GigabitEthernet 1/1)# unset switch-port
R2(cfg-if-GigabitEthernet 1/1)# ip address 60.154.101.3 255.255.255.0
R2(cfg-if-GigabitEthernet 1/1)# exit
R2(cfg)# router-protocol ospf
R2(cfg-router)# network-advertise 60.154.101.0 0.0.0.255 area 10
R2(cfg-router)# network-advertise 192.168.201.0 0.0.0.255 area 10
```

5. Verification

Run the **display vrrp-config** command to verify the configuration.

Check whether router 1, which acts as the active router, reduces its VRRP standby group priority from 120 to 90 when GigabitEthernet2/1 connected to the wide area network (WAN) is unavailable. If yes, router 2 becomes the active router.

Check whether router 1 resumes its VRRP standby group priority from 30 to 120 when GigabitEthernet 2/1 connected to the WAN recovers. If yes, router 1 is reelected as the active router.

```
R1# display vrrp-config
GigabitEthernet 0/0 - Group 1
   State is Active
   Virtual IP address is 192.168.201.1 configured
   Virtual MAC address is 0000.5e00.0101
   Advertisement min_tx_interval is 3 sec
   Preemption is enabled
      min delay is 0 sec
   Priority is 120
   Active Router is 192.168.201.217 (local), priority is 120
   Active Down interval is 10.59 sec
   Tracking state of 1 interface, 1 up:
up     GigabitEthernet 2/1 priority reduce=30
```

Check the configuration on router 2.

```
R2# display vrrp-config
GigabitEthernet 0/0 - Group 1
   State is Standby
   Virtual IP address is 192.168.201.1 configured
   Virtual MAC address is 0000.5e00.0101
   Advertisement min_tx interval is 3 sec
   Preemption is enabled
      min delay is 0 sec
   Priority is 100
   Active Router is 192.168.201.217 , priority is 120
   Active Advertisement min_tx interval is 3 sec
   Active Down min_tx interval is 10.82 sec
```

6. Configuration Files

- Device R1 configuration file

```
hostname R1
!
interface GigabitEthernet 0/0
 unset switch-port
 ip address 192.168.201.217 255.255.255.0
 vrrp-config 1 priority 120
 vrrp-config 1 timers advertise 3
 vrrp-config 1 ip 192.168.201.1
```

```
 vrrp-config 1 track GigabitEthernet 2/1 30
!
interface GigabitEthernet 2/1
 unset switch-port
 ip address 202.101.90.63 255.255.255.0
!
router-protocol ospf
 network 202.101.90.0 0.0.0.255 area 10
 network 192.168.201.0 0.0.0.255 area 10
!
end
```

- Device R2 configuration file

```
hostname R2
!
interface GigabitEthernet 0/0
 unset switch-port
 ip address 192.168.201.213 255.255.255.0
 vrrp-config 1 timers advertise 3
 vrrp-config 1 ip 192.168.201.1
!
interface GigabitEthernet 1/1
 unset switch-port
 ip address 60.154.101.3 255.255.255.0
!
router-protocol ospf
 network 60.154.101.0 0.0.0.255 area 10
 network 192.168.201.0 0.0.0.255 area 10
!
end
```

- Device R3 configuration file

```
hostname R3
!
interface GigabitEthernet 0/0
 unset switch-port
 ip address 192.168.12.217 255.255.255.0
!
interface GigabitEthernet 1/1
 unset switch-port
 ip address 60.154.101.5 255.255.255.0
!
interface GigabitEthernet 2/1
 unset switch-port
 ip address 202.101.90.61 255.255.255.0
!
router-protocol ospf
```

```
 network 202.101.90.0 0.0.0.255 area 10
 network 192.168.12.0 0.0.0.255 area 10
 network 60.154.101.0 0.0.0.255 area 10
!

end
```

7. Common Errors

- Different virtual IP addresses of VRRP are configured on the routers in the same VRRP group, resulting in multiple active routers in the group.
- Different VRRP advertisement transmission intervals are configured on the routers in the same VRRP group and the timer learning function is not configured. As a result, multiple active routers arise in the group.
- Different VRRP packet versions are configured on the routers in the same VRRP group, resulting in multiple active routers in the group.
- For VRRPv2, the Ethernet ports of the routers in the same VRRP group all use the plaintext password authentication mode but use different authentication strings. As a result, multiple active routers arise in the group.

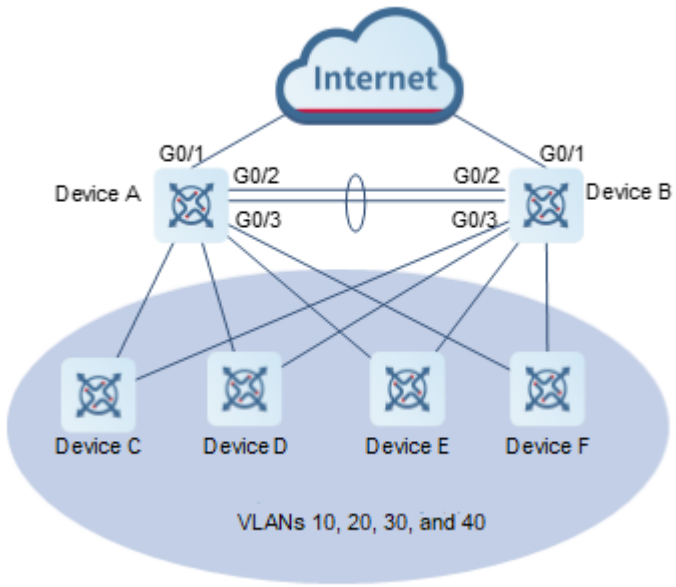
1.10.2 Configuring VRRP + MSTP

1. Requirements

As shown in Figure 1-10, VRRP and Multiple Spanning Tree Protocol (MSTP) can be configured to build a dual link standby network by using access devices and aggregation devices.

2. Topology

Figure 1-10 Configuring VRRP + MSTP



3. Notes

- Enable MSTP on devices (Device A, B, C, D, E, and F in this example). Configure VLAN-instance mappings (map VLANs 10 and 20 to instance 1, VLANs 30 and 40 to instance 2, and the rest VLANs to instance 0), and configure gateway devices (Device A and B in this example) as the root bridges of corresponding instances.
- Add the SVIs of all VLANs to corresponding VRRP standby groups, and set gateways to the active and standby routers of corresponding standby groups.Table 1-4 lists the configuration details in the example.

Table 1-4 Parameters for VRRP + MSTP Topology Configuration

Gateway	VLAN ID	SVI	Standby Group	Virtual IP Address	State
Device A	10	192.168.10.2	VRRP 10	192.168.10.1	Active
Device B		192.168.10.3			Standby
Device A	20	192.168.20.2	VRRP 20	192.168.20.1	Active
Device B		192.168.20.3			Standby
Device A	30	192.168.30.2	VRRP 30	192.168.30.1	Standby
Device B		192.168.30.3			Active
Device A	40	192.168.40.2	VRRP 40	192.168.40.1	Standby
Device B		192.168.40.3			Active

- Configure the uplink ports (port GigabitEthernet 0/1 of Device A and B in the example) of active routers in standby groups as the monitored ports of the active routers.
  - Create VLANs. Create VLANs 10, 20, 30, and 40 on Device A and B.
  - Configure MST regions. Map VLANs 10 and 20 to instance 1, VLANs 20 and 30 to instance 2, and the rest VLANs to instance 0 on Device A and B.
  - Configure Device A as the root bridge of MST 0 and MST 1, and Device B as the root bridge of MST 2.
  - Enable MSTP.
  - Configure SVIs for all the VLANs, add the SVIs to corresponding standby groups, and configure virtual IP addresses for the standby groups. Table 1-4 provides the corresponding configuration parameters.
  - Configure active routers and standby routers for all the groups.
  - Configure the uplink ports of active routers in VRRP standby groups as monitored ports of the VRRP groups.
  - Configure the interconnected ports of the core routers as a Layer 2 link aggregation.

4. Procedure

(1) Perform the following configuration on Device A:

Create VLANs 10, 20, 30, and 40.

```
DeviceA > privilege
DeviceA # config-system terminal
DeviceA (cfg)# vlan range 10,20,30,40
DeviceA (cfg)# exit
```

Map VLANs 10 and 20 to instance 1, VLANs 30 and 40 to instance 2, and the rest VLANs to instance 0.

```
DeviceA (cfg)# spanning-tree mst configuration
DeviceA (cfg-mst)# instance 1 vlan 10,20
DeviceA (cfg-mst)# instance 2 vlan 30,40
DeviceA (cfg-mst)# exit
```

On Device A, set the priority of MST 0 and MST 1 to **4096**, and the priority of MST 2 to **8192**.

```
DeviceA (cfg)# spanning-tree mst 0 priority 4096
DeviceA (cfg)# spanning-tree mst 1 priority 4096
DeviceA (cfg)# spanning-tree mst 2 priority 8192
```

Enable MSTP.

```
DeviceA (cfg)# spanning-tree
```

Configure SVIs for all the VLANs, add the SVIs to corresponding standby groups, and configure virtual IP addresses for the standby groups.

```
DeviceA (cfg)# interface vlan 10
DeviceA (cfg-if-VLAN 10)# ip address 192.168.10.2 255.255.255.0
DeviceA (cfg-if-VLAN 10)# vrrp-config 10 ip 192.168.10.1
DeviceA (cfg-if-VLAN 10)# exit
DeviceA (cfg)# interface vlan 20
DeviceA (cfg-if-VLAN 20)# ip address 192.168.20.2 255.255.255.0
DeviceA (cfg-if-VLAN 20)# vrrp-config 20 ip 192.168.20.1
DeviceA (cfg-if-VLAN 20)# exit
DeviceA (cfg)# interface vlan 30
DeviceA (cfg-if-VLAN 30)# ip address 192.168.30.2 255.255.255.0
DeviceA (cfg-if-VLAN 30)# vrrp-config 30 ip 192.168.30.1
DeviceA (cfg-if-VLAN 30)# exit
DeviceA (cfg)# interface vlan 40
DeviceA (cfg-if-VLAN 40)# ip address 192.168.40.2 255.255.255.0
DeviceA (cfg-if-VLAN 40)# vrrp-config 40 ip 192.168.40.1
DeviceA (cfg-if-VLAN 40)# exit
```

Increase the priority of standby groups 10 and 20 on Device A to **120**.

```
DeviceA (cfg)# interface vlan 10
DeviceA (cfg-if-VLAN 10)# vrrp-config 10 priority 120
DeviceA (cfg-if-VLAN 10)# exit
DeviceA (cfg)# interface vlan 20
DeviceA (cfg-if-VLAN 20)# vrrp-config 20 priority 120
DeviceA (cfg-if-VLAN 20)# exit
```

Configure the port GigabitEthernet 0/1 of Device A as a route port, and set the IP address to10.10.1.1/24.

```
DeviceA (cfg)# interface gigabitEthernet 0/1
DeviceA (cfg-if-GigabitEthernet 0/1)# unset switch-port
```

```
DeviceA (cfg-if-GigabitEthernet 0/1)# ip address 10.10.1.1 255.255.255.0
DeviceA (cfg-if-GigabitEthernet 0/1)# exit
```

Configure the port GigabitEthernet 0/1 of Device A as the monitored port of standby groups 10 and 20, and set *Priority decrement* to 30.

```
DeviceA (cfg)# interface vlan 10
DeviceA (cfg-if-VLAN 10)# vrrp-config 10 track gigabitEthernet 0/1 30
DeviceA (cfg-if-VLAN 10)# exit
DeviceA (cfg)# interface vlan 20
DeviceA (cfg-if-VLAN 20)# vrrp-config 20 track gigabitEthernet 0/1 30
DeviceA (cfg-if-VLAN 20)# exit
```

Configure the ports GigabitEthernet 0/2 and GigabitEthernet 0/3 to belong to a Layer 2 link aggregation, and configure the link aggregation as a trunk port.

```
DeviceA (cfg)# interface range gigabitEthernet 0/2-3
DeviceA (cfg-if-list)# port-group 1
DeviceA (cfg)# interface aggregateport 1
DeviceA (cfg-if-AggregatePort 1)# switch-port mode trunk
```

(2) Perform the following configuration on Device B:

Create VLANs 10, 20, 30, and 40.

```
DeviceB > privilege
DeviceB # config-system terminal
DeviceB (cfg)# vlan range 10,20,30,40
DeviceB (cfg-vlan-range)# exit
```

Map VLANs 10 and 20 to instance 1, VLANs 30 and 40 to instance 2, and the rest VLANs to instance 0.

```
DeviceB (cfg)# spanning-tree mst configuration
DeviceB (cfg-mst)# instance 1 vlan 10,20
DeviceB (cfg-mst)# instance 2 vlan 30,40
DeviceB (cfg-mst)# exit
```

On Device B, set the priority of MST 2 to **4096**, and the priority of MST 0 and MST 1 to **8192**.

```
DeviceB (cfg)# spanning-tree mst 2 priority 4096
DeviceB (cfg)# spanning-tree mst 0 priority 8192
DeviceB (cfg)# spanning-tree mst 1 priority 8192
```

Enable MSTP.

```
DeviceB (cfg)# spanning-tree
```

Configure SVIs for all the VLANs, add the SVIs to corresponding standby groups, and configure virtual IP addresses for the standby groups.

```
DeviceB (cfg)# interface vlan 10
DeviceB (cfg-if-VLAN 10)# ip address 192.168.10.3 255.255.255.0
DeviceB (cfg-if-VLAN 10)# vrrp-config 10 ip 192.168.10.1
DeviceB (cfg-if-VLAN 10)# exit
DeviceB (cfg)# interface vlan 20
DeviceB (cfg-if-VLAN 20)# ip address 192.168.20.3 255.255.255.0
DeviceB (cfg-if-VLAN 20)# vrrp-config 20 ip 192.168.20.1
```

```
DeviceB (cfg-if-VLAN 20)# exit
DeviceB (cfg)# interface vlan 30
DeviceB (cfg-if-VLAN 30)# ip address 192.168.30.3 255.255.255.0
DeviceB (cfg-if-VLAN 30)# vrrp-config 30 ip 192.168.30.1
DeviceB (cfg-if-VLAN 30)# exit
DeviceB (cfg)# interface vlan 40
DeviceB (cfg-if-VLAN 40)# ip address 192.168.40.3 255.255.255.0
DeviceB (cfg-if-VLAN 40)# vrrp-config 40 ip 192.168.40.1
DeviceB (cfg-if-VLAN 40)# exit
```

Increase the priority of VRRP 30 and VRRP 40 on Device B to **120**.

```
DeviceB (cfg)# interface vlan 30
DeviceB (cfg-if-VLAN 30)# vrrp-config 30 priority 120
DeviceB (cfg-if-VLAN 30)# exit
DeviceB (cfg)# interface vlan 40
DeviceB (cfg-if-VLAN 40)# vrrp-config 40 priority 120
DeviceB (cfg-if-VLAN 40)# exit
```

Configure the port GigabitEthernet 0/1 of Device B as a route port, and set the IP address to 10.10.2.1/24.

```
DeviceB (cfg)#interface gigabitEthernet 0/1
DeviceB (cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceB (cfg-if-GigabitEthernet 0/1)# ip address 10.10.2.1 255.255.255.0
DeviceB (cfg-if-GigabitEthernet 0/1)# exit
```

Configure the port GigabitEthernet 0/1 of Device B as the monitored port of standby groups 30 and 40, and set *Priority decrement* to 30.

```
DeviceB (cfg)# interface vlan 30
DeviceB (cfg-if-VLAN 30)# vrrp-config 30 track gigabitEthernet 0/1 30
DeviceB (cfg-if-VLAN 30)# exit
DeviceB (cfg)# interface vlan 40
DeviceB (cfg-if-VLAN 40)# vrrp-config 40 track gigabitEthernet 0/1 30
DeviceB (cfg-if-VLAN 40)# exit
```

Configure the ports GigabitEthernet 0/2 and GigabitEthernet 0/3 to belong to a Layer 2 link aggregation, and configure the link aggregation as a trunk port.

```
DeviceB  (cfg)# interface range gigabitEthernet 0/2-3
DeviceB  (cfg-if-list)# port-group 1
DeviceB  (cfg)# interface aggregateport 1
DeviceB  (cfg-if-AggregatePort 1)# switch-port mode trunk
```

5. Verification

Run the **display current-config** and **display vrrp-config brief** commands to check whether the configuration is correct.

```
DeviceA # display current-config
!
vlan 10
!
vlan 20
```

```
!
vlan 30
!
vlan 40
!
spanning-tree
spanning-tree mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 20
 instance 2 vlan 30, 40
spanning-tree mst 0 priority 4096
spanning-tree mst 1 priority 4096
spanning-tree mst 2 priority 8192
interface GigabitEthernet 0/1
 unset switch-port
 unset ip proxy-arp
 ip address 10.10.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 port-group 1
!
interface GigabitEthernet 0/3
 port-group 1
!
interface AggregatePort 1
 switch-port mode trunk
!
interface VLAN 10
 unset ip proxy-arp
 ip address 192.168.10.2 255.255.255.0
 vrrp-config 10 priority 120
 vrrp-config 10 ip 192.168.10.1
 vrrp-config 10 track GigabitEthernet 0/1 30
!
interface VLAN 20
 unset ip proxy-arp
 ip address 192.168.20.2 255.255.255.0
 vrrp-config 20 priority 120
 vrrp-config 20 ip 192.168.20.1
 vrrp-config 20 track GigabitEthernet 0/1 30
!
interface VLAN 30
 unset ip proxy-arp
 ip address 192.168.30.2 255.255.255.0
 vrrp-config 30 ip 192.168.30.1
!
```

```
interface VLAN 40
 unset ip proxy-arp
 ip address 192.168.40.2 255.255.255.0
 vrrp-config 40 ip 192.168.40.1
```

Check the VRRP state of device A.

```
DeviceA # display vrrp-config brief
Interface   Grp   Pri   timer   Own   Pre   State     Active addr     Group addr
VLAN 10      10    120   3        -     P     Active    192.168.10.2    192.168.10.1
VLAN 20      20    120   3        -     P     Active    192.168.20.2    192.168.20.1
VLAN 30      30    100   3        -     P     Standby   192.168.30.3    192.168.30.1
VLAN 40      40    100   3        -     P     Standby   192.168.40.3    192.168.40.1
```

Disconnect the uplink of Device A and check the VRRP state of the device.

```
DeviceA # display vrrp-config brief
Interface   Grp   Pri   timer   Own   Pre   State     Active addr     Group addr
VLAN 10      10    90    3        -     P     Standby   192.168.10.3    192.168.10.1
VLAN 20      20    90    3        -     P     Standby   192.168.20.3    192.168.20.1
VLAN 30      30    100   3        -     P     Standby   192.168.30.3    192.168.30.1
VLAN 40      40    100   3        -     P     Standby   192.168.40.3    192.168.40.1
```

Check the configuration of Device B.

```
DeviceB # display current-config
!
vlan 10
!
vlan 20
!
vlan 30
!
vlan 40
!
spanning-tree
spanning-tree mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 20
 instance 2 vlan 30, 40
spanning-tree mst 0 priority 8192
spanning-tree mst 1 priority 8192
spanning-tree mst 2 priority 4096
interface GigabitEthernet 0/1
 unset switch-port
 unset ip proxy-arp
 ip address 10.10.2.1 255.255.255.0
!
interface GigabitEthernet 0/2
 port-group 1!
interface GigabitEthernet 0/3
```

```
 port-group 1
!
interface AggregatePort 1
 switch-port mode trunk
!
interface VLAN 10
 unset ip proxy-arp
 ip address 192.168.10.3 255.255.255.0
 vrrp-config 10 ip 192.168.10.1
!
interface VLAN 20
 unset ip proxy-arp
 ip address 192.168.20.3 255.255.255.0
 vrrp-config 20 ip 192.168.20.1
!
interface VLAN 30
 unset ip proxy-arp
 ip address 192.168.30.3 255.255.255.0
 vrrp-config 30 priority 120
 vrrp-config 30 ip 192.168.30.1
 vrrp-config 30 track GigabitEthernet 0/1 30
!
interface VLAN 40
 unset ip proxy-arp
 ip address 192.168.40.3 255.255.255.0
 vrrp-config 40 priority 120
 vrrp-config 40 ip 192.168.40.1
 vrrp-config 40 track GigabitEthernet 0/1 30
```

Check the VRRP state of device B.

```
DeviceB # display vrrp-config brief
Interface   Grp   Pri   timer   Own   Pre   State     Active addr     Group addr
VLAN 10      10    100   3        -     P     Standby   192.168.10.2   192.168.10.1
VLAN 20      20    100   3        -     P     Standby   192.168.20.2   192.168.20.1
VLAN 30      30    120   3        -     P     Active    192.168.30.3   192.168.30.1
VLAN 40      40    120   3        -     P     Active    192.168.40.3   192.168.40.1
```

Disconnect the uplink of Device B and check the VRRP state of the device.

```
DeviceB # display vrrp-config brief
Interface   Grp   Pri   timer   Own   Pre   State     Active addr     Group addr
VLAN 10      10    100   3        -     P     Active    192.168.10.3   192.168.10.1
VLAN 20      20    100   3        -     P     Active    192.168.20.3   192.168.20.1
VLAN 30      30    120   3        -     P     Active    192.168.30.3   192.168.30.1
VLAN 40      40    120   3        -     P     Active    192.168.40.3   192.168.40.1
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
vlan 10
!
vlan 20
!
vlan 30
!
vlan 40
!
spanning-tree
spanning-tree mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 20
 instance 2 vlan 30, 40
spanning-tree mst 0 priority 4096
spanning-tree mst 1 priority 4096
spanning-tree mst 2 priority 8192
interface GigabitEthernet 0/1
 unset switch-port
 unset ip proxy-arp
 ip address 10.10.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 port-group 1
!
interface GigabitEthernet 0/3
 port-group 1
!
interface AggregatePort 1
 switch-port mode trunk
!
interface VLAN 10
 unset ip proxy-arp
 ip address 192.168.10.2 255.255.255.0
 vrrp-config 10 priority 120
 vrrp-config 10 ip 192.168.10.1
 vrrp-config 10 track GigabitEthernet 0/1 30
!
interface VLAN 20
 unset ip proxy-arp
 ip address 192.168.20.2 255.255.255.0
 vrrp-config 20 priority 120
 vrrp-config 20 ip 192.168.20.1
 vrrp-config 20 track GigabitEthernet 0/1 30
!
```

```
interface VLAN 30
 unset ip proxy-arp
 ip address 192.168.30.2 255.255.255.0
 vrrp-config 30 ip 192.168.30.1
!
interface VLAN 40
 unset ip proxy-arp
 ip address 192.168.40.2 255.255.255.0
 vrrp-config 40 ip 192.168.40.1
```

- Device B configuration file

```
hostname DeviceB
!
vlan 10
!
vlan 20
!
vlan 30
!
vlan 40
!
spanning-tree
spanning-tree mst configuration
 instance 0 vlan 1-9, 11-19, 21-29, 31-39, 41-4094
 instance 1 vlan 10, 20
 instance 2 vlan 30, 40
spanning-tree mst 0 priority 8192
spanning-tree mst 1 priority 8192
spanning-tree mst 2 priority 4096
interface GigabitEthernet 0/1
 unset switch-port
 unset ip proxy-arp
 ip address 10.10.2.1 255.255.255.0
!
interface GigabitEthernet 0/2
 port-group 1!
interface GigabitEthernet 0/3
 port-group 1
!
interface AggregatePort 1
 switch-port mode trunk
!
interface VLAN 10
 unset ip proxy-arp
 ip address 192.168.10.3 255.255.255.0
 vrrp-config 10 ip 192.168.10.1
!
```

```
interface VLAN 20
 unset ip proxy-arp
 ip address 192.168.20.3 255.255.255.0
 vrrp-config 20 ip 192.168.20.1
!
interface VLAN 30
 unset ip proxy-arp
 ip address 192.168.30.3 255.255.255.0
 vrrp-config 30 priority 120
 vrrp-config 30 ip 192.168.30.1
 vrrp-config 30 track GigabitEthernet 0/1 30
!
interface VLAN 40
 unset ip proxy-arp
 ip address 192.168.40.3 255.255.255.0
 vrrp-config 40 priority 120
 vrrp-config 40 ip 192.168.40.1
 vrrp-config 40 track GigabitEthernet 0/1 30
```

7. Common Errors

- Different VRRP authentication modes are configured on the Ethernet ports of routers in the same VRRP group, resulting in multiple active routers in the group.
- For VRRPv2, the Ethernet ports of the routers in the same VRRP group all use the plaintext password authentication mode but use different authentication strings. As a result, multiple active routers arise in the group.
- Different VRRP advertisement transmission intervals are configured on the routers in the same VRRP group and the timer learning function is not configured. As a result, multiple active routers arise in the group.
- Different virtual IP addresses of VRRP are configured on the routers in the same VRRP group, resulting in multiple active routers in the group.

Contents

1 Configuring VRRP Plus 1

1.1 Introduction 1

1.1.1 Basic Concepts 1

1.1.2 Packet Structure 4

1.2 Configuration Task Summary 5

1.3 Configuring VRRP Plus 5

1.3.1 Overview 5

1.3.2 Restrictions and Guidelines 5

1.3.3 Procedure 6

1.4 Monitoring 7

1.5 Configuration Examples 8

1.5.1 Configuring IPv4 VRRP Plus 8

1.5.2 Configuring IPv6 VRRP Plus 10

1 Configuring VRRP Plus

1.1 Introduction

Virtual Router Redundancy Protocol Plus (VRRP Plus) is an extension of VRRP. It uses VRRP to implement gateway standby and load balancing in an IEEE 802.3 local area network (LAN). A standby router with VRRP enabled cannot forward packets. To use VRRP to implement load balancing, you need to manually configure multiple VRRP groups and set the gateway addresses of hosts in the LAN to virtual IP addresses of different VRRP groups. This increases the workload of the network administrator. VRRP Plus is designed to address this issue. VRRP Plus automatically implements load balancing. It automatically distributes traffic of different hosts to members of a VRRP Plus group, and you do not need to configure multiple VRRP groups or set the gateway addresses of hosts in the LAN to virtual IP addresses of different VRRP groups. This reduces the workload of the network administrator. You can enable a load balancing policy within a VRRP group without configuring multiple VRRP groups or different default gateways for hosts.

1.1.1 Basic Concepts

1. BVG

The balancing virtual gateway (BVG) allocates virtual MAC addresses to members of a VRRP Plus group. It responds to the gateway Address Resolution Protocol/Neighbor Discovery (ARP/ND) requests in the LAN, and forwards packets of hosts in the LAN.

2. BVF

The balancing virtual forwarder (BVF) forwards packets of hosts in the LAN. If a virtual MAC address is allocated to a BVF, the BVF participates in packet forwarding; otherwise, the BVF does not participate in packet forwarding.

3. Correlation with VRRP

VRRP Plus relies on VRRP, and runs in the following way:

- A master device in VRRP corresponds to a BVG in VRRP Plus.
- A standby device in VRRP corresponds to a BVF in VRRP Plus. Gateway addresses of hosts in a LAN are set to the virtual IPv4/IPv6 address of VRRP.

4. MAC Address Allocation

A BVG allocates virtual MAC addresses to BVFs. For an IPv4 VRRP Plus group, the BVG directly uses the virtual MAC address of VRRP to ensure compatibility between IPv4 VRRP Plus and IPv4 VRRP. That is, the virtual MAC address used by the BVG is 00-00-5E-00-01-{VRID} (VRID is the VRRP group number). The virtual MAC address used by a BVF is 00-1A-A9-16-{MemberID}-{VRID} (MemberID is the member ID of the BVF in the VRRP Plus group). Currently, a VRRP Plus group can have up to four members. The BVG uses the member ID 01, and the other BVFs use the member IDs 02 to 04.

For an IPv6 VRRP Plus group, the BVG directly uses the virtual MAC address of IPv6 VRRP to ensure compatibility between IPv6 VRRP Plus and IPv6 VRRP. That is, the virtual MAC address used by the BVG is

00-00-5E-00-02-{VRID} (VRID is the IPv6 VRRP group number). The virtual MAC address used by a BVF is 14-14-4B-72-{76 + MemberID }-{VRID}. The value range of MemberID is from 01 to 03.

5. Load Balancing Policies

The BVG responds to the gateway ARP/NS requests sent from hosts in a LAN. According to a specific load balancing policy, the BVG responds to hosts by using different virtual MAC addresses. The following load balancing policies are available:

- Host-dependent policy: A specific virtual MAC address is used to respond to the requests sent by a specific host.
- Round-robin policy: Virtual MAC addresses in a standby group are used in a cyclic manner to respond to the gateway ARP/NS requests sent by hosts.
- Weighted policy: The ARP/NA requests are responded based on the forwarding capability of a device.

Policy switching is described below:

- If the load balancing mode is changed, load balancing is always implemented in the new load balancing mode. For example, if the round-robin load balancing mode is previously used, and later the weighted mode is used, load balancing is implemented in weighted mode. Load balancing cannot be implemented for hosts that have learned the VRRP virtual gateway addresses before the configuration of VRRP Plus. Therefore, if VRRP Plus is configured after a VRRP group switches to the master state, real load balancing cannot be implemented before aging of the ARP/ND entries learned by hosts. Load balancing is implemented only after the gateway ARP/ND entries learned by the hosts age and the hosts request new gateway addresses.
- Weight configuration means configuring different weights for different devices. In this way, more traffic is distributed to the device with a greater weight and less traffic is distributed to the device with a smaller weight, thereby fully utilizing the forwarding performance of different devices (if the total weight of virtual routers in a VRRP Plus group is 0, the ARP/NS requests are not responded). When the weight of a BVF in a standby group is smaller than the lower threshold, the BVF automatically exits the forwarding role. When the weight recovers and is greater than the upper threshold, the BVF automatically applies to the forwarding role. The forwarding role can be recovered when one or more remaining virtual MAC addresses or proxy virtual MAC addresses exist.
- The function of periodically sending of gratuitous ARP packets on an interface also affects the load balancing function of VRRP Plus. When VRRP Plus is enabled, the function of sending gratuitous ARP/NA packets of VRRP virtual IP addresses will be disabled. When a virtual IP address overlaps with an actual IP address, gratuitous ARP/NA packets of this address are no longer sent.
- When an address conflict occurs between a host and the local device, the ARP/NA module will broadcast gratuitous ARP/NA packets of this address. If a conflict occurs on the VRRP Plus virtual address, sending gratuitous ARP/NA packets results in re-learning of the host's gateway MAC address, which negatively affects the load balancing function of VRRP Plus. Therefore, the load balancing function of VRRP Plus is currently not supported in this scenario.

6. Proxy Virtual MAC Address

When a device with a virtual MAC address assigned becomes faulty in a standby group, traffic of hosts that use this virtual MAC address as the gateway MAC address will be interrupted. The BVG in the VRRP Plus standby group can quickly detect the fault, and automatically allocate the virtual MAC address of the faulty BVF to another device in the standby group. The new device acts as the proxy of the faulty device to forward packets of

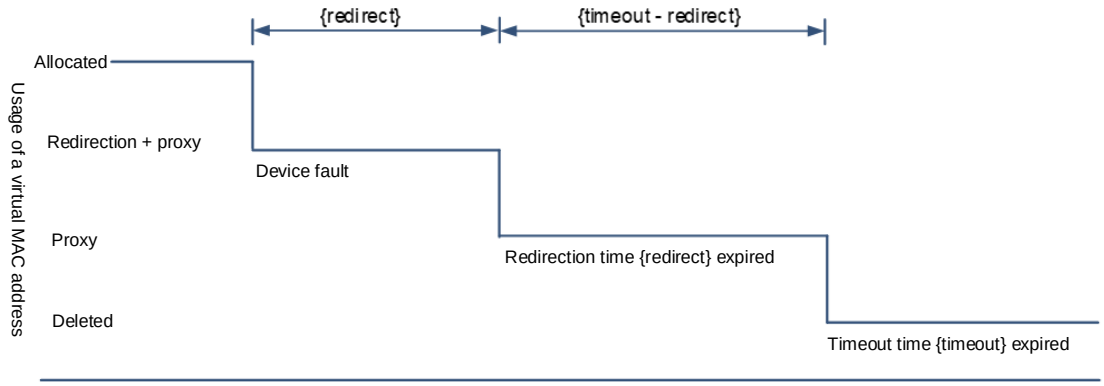
the virtual MAC address. In addition, this proxy device takes over traffic of original hosts to prevent traffic interruption. The virtual MAC address allocated to a device in a standby group can be called a master virtual MAC address, and the virtual MAC address used by this device on behalf of another device is called a proxy virtual MAC address.

7. Redirection Interval and Timeout

VRRP Plus provides the virtual MAC address proxy function so that another device can take the place of a faulty device with a virtual MAC address assigned to forward packets. If a BVF recovers from the fault, its forwarding role should be restored and the BVF continues to forward packets of the virtual MAC address allocated to this BVF. If the faulty BVF does not recover all the time, the standby group stops redirecting traffic to this virtual MAC address. That is, when gateway ARP requests are received again, this virtual MAC address is no longer used to give responses. After a sufficient long period of time, it is believed that hosts that use the virtual MAC address as the gateway MAC address already update the ARP/ND table entry of the gateway address, and the traffic is already taken over by other devices. Now, this virtual MAC address can be deleted, and packets sent to this virtual MAC address should be discarded.

VRRP Plus allows configuring the redirection interval and timeout for a standby group. When a device is faulty, the standby group allocates the virtual MAC address of the faulty device to another device. Within the redirection interval, the standby group continues to use this virtual MAC address to respond to the ARP/NS requests. When the redirection time expires, the standby group no longer uses this virtual MAC address to respond to the requests. When the timeout time elapses, the standby group deletes this virtual MAC address and stops using it for proxy forwarding. Figure 1-1 shows the changes in the role of a virtual MAC address within the redirection interval and timeout time.

Figure 1-1 Diagram of Changes in the Role of an Allocated Virtual MAC Address



8. Correlation with Track

VRRP Plus supports correlation with bidirectional forwarding detection (BFD) and can adjust the weight based on the link status. Each device in a standby group can associate its weight with the link status. When a link is abnormal or unreachable, the device automatically decreases its weight. When the weight is too low, the device automatically exits the forwarding role. If a standby group is currently using the weighted load balancing policy, traffic can be distributed based on the new weight. When the associated link recovers, the device automatically restores its original weight and the forwarding role. If the standby group is currently using the weighted load balancing policy, traffic is distributed based on the restored weight.

9. Forwarding Preemption

VRRP Plus supports the function of preempting the forwarding role. In VRRP Plus, at most four devices can participate in load balancing. That is, a VRRP Plus standby group generates at most four virtual MAC addresses. If more than four devices are added to a VRRP Plus group, only four devices participate in packet forwarding. The remaining devices only listen to the statuses of other devices and do not participate in packet forwarding. Only when a device participating in packet forwarding is faulty, another device that originally does not participate in packet forwarding takes the place of the faulty device to forward packets. Assume that a VRRP Plus standby group already has four devices and all these devices participate in packet forwarding; a fifth device is added to the VRRP Plus group, and the forwarding capability of this device is strong or a device taking the forwarding role encounters a link failure and consequently the degradation of forwarding performance. In this case, if the preemption mode is enabled, the fifth device preempts the forwarding role from a device with a smaller weight (that is, with weaker forwarding capability). A greater weight is configured for a device with stronger forwarding capability. When the weight of a device in listening state is greater than that of a forwarding device, the device in listening state automatically preempts the forwarding role from the forwarding device. That is, the device with stronger forwarding capability forwards packets, whereas the device with lower forwarding capability is in listening state. This minimizes the waste of resources.

The BVG in a standby group is responsible for allocating virtual MAC addresses. Therefore, the BVG role cannot be preempted, and only the forwarding role of a BVF can be seized. If the BVG is faulty, VRRP re-elects a new master device, which assumes the BVG role.

1.1.2 Packet Structure

Figure 1-2 shows the structure of VRRP Plus packets.

Figure 1-2 Structure of VRRP Plus Packets

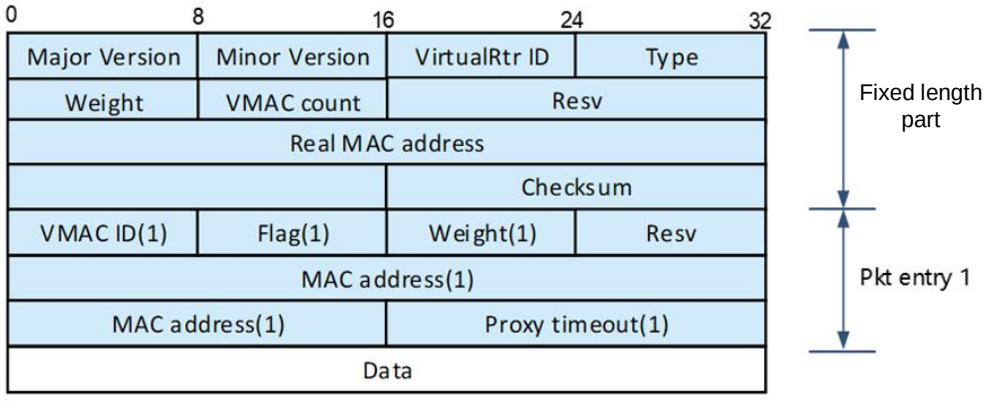


Table 1-1 Field Description of VRRP Plus Packets

Packet Field	Description
Major Version	Major version number. The current version is 1.
Minor Version	Minor version number. The current version is 0.
VirtualRtr ID	Virtual router ID, that is, the ID of a VRRP Plus standby group.

Packet Field	Description
Type	Type of VRRP Plus packets. Values and corresponding types are as follows: ● **0**: Indicates a tricky packet. ● **1**: Indicates a virtual MAC request packet. ● **2**: Indicates a keepalive packet of a BVF. ● **3**: Indicates the virtual MAC address release packet sent by the BVF to the BVG. ● **4**: Indicates the virtual MAC address allocation packet sent by the BVG to a BVF.
Weight	Weight of the device.
VMAC count	Number of virtual MAC addresses filled in the packet.
Resv	Reserved field in the fixed length part.
Real MAC address	Real MAC address of the interface on the local device.
Checksum	Packet checksum (including the fixed length part and filling part).
VMAC ID	Filled virtual MAC ID.
Flag	Flag information of the virtual MAC address.
Weight	Weight of the device to which the virtual MAC address is allocated.
Resv	Reserved field in the filling part.
MAC address	Real MAC address of the interface on the device to which the virtual MAC address is allocated.
Proxy timeout	Remaining time before timeout of the virtual MAC address filling in the case of a proxy virtual MAC address.

1.2 Configuration Task Summary

VRRP Plus configuration includes the following tasks:

(1) Configuring VRRP Plus

1.3 Configuring VRRP Plus

1.3.1 Overview

This section describes how to enable the VRRP Plus function and adjust the VRRP Plus weight based on the track object, to implement automatic balancing among multiple standby groups.

1.3.2 Restrictions and Guidelines

To enable the VRRP Plus function, you must configure a VRRP virtual IP address for the corresponding standby group.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the switch virtual interface (SVI) configuration mode.

  interface vlan *interface-number*

- Enter the VXLAN router interface configuration mode.

  interface overlayrouter *interface-number*

(4) Enable IPv4 VRRP.

vrrp-config *group-id* **ip** *ipv4-address* [**secondary**]

IPv4 VRRP is disabled on a port by default.

(5) Configure a priority for the IPv4 VRRP standby group.

vrrp-config *group-id* **priority** *Priority*

The default priority of an IPv4 VRRP standby group is **100**.

(6) Enable the VRRP Plus function on the interface.

vrrp-config [**ipv6**] *group-id* **balance**

The VRRP Plus function is disabled by default.

(7) Configure a VRRP Plus load balancing policy.

vrrp-config [**ipv6**] *group-id* **load-balancing** { **host-dependent** | **round-robin** | **weighted** }

By default, VRRP Plus uses the host-dependent load balancing policy, that is, host-dependent.

(8) Configure the redirection interval and timeout time for proxy virtual MAC addresses of the VRRP Plus standby group.

vrrp-config [**ipv6**] *group-id* **timers redirect** *redirect-interval redirect-timeout*

The default redirection interval of proxy virtual MAC addresses of a VRRP Plus standby group is **300** seconds and the default redirection timeout time is **14400** seconds.

(9) Configure the weight and upper and lower thresholds for the VRRP Plus standby group.

vrrp-config [**ipv6**] *group-id* **weighting** *weight-limit* [**lower** *min-weight-value*] [**upper** *max-weight-value*]

The default weight, default upper threshold, and default lower threshold of a VRRP Plus standby group are **100**, **1**, and **100** respectively.

(10) Configure the forwarding preemption function for the VRRP Plus standby group.

vrrp-config [**ipv6**] *group-id* **forwarder preempt**

The forwarding preemption function is enabled for a VRRP Plus standby group by default.

(11) Configure a track object for adjusting the weight for the VRRP Plus standby group.

vrrp-config [**ipv6**] *group-id* **weighting track** *object-number* [**reduce** *value*]

VRRP Plus tracking is not configured by default.

1.4 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-2 VRRP Plus Monitoring

Command	Purpose
display [**ipv6**] **vrrp-config balance**	Displays the brief or detailed information about VRRP Plus.
display [**ipv6**] **vrrp-config balance interface** *interface-type interface-number* [**brief**]	Displays the actions of a VRRP Plus group on a specified interface.
debugging [**ipv6**] **vrrp balance**	Debugs the VRRP Plus function.
debugging [**ipv6**] **vrrp balance error**	Debugs errors.
debugging [**ipv6**] **vrrp balance event**	Debugs events of a VRRP Plus group.
debugging [**ipv6**] **vrrp balance messages**	Debugs messages of VRRP Plus.
debugging [**ipv6**] **vrrp balance packets**	Debugs the VRRP Plus packets.
debugging [**ipv6**] **vrrp balance state**	Debugs the VRRP Plus group status.
debugging vrrp balance timer	Debugs the timers of a VRRP Plus group.
debugging [**ipv6**] **vrrp balance**	Debugs the VRRP Plus function.
debugging [**ipv6**] **vrrp balance error**	Debugs errors.

1.5 Configuration Examples

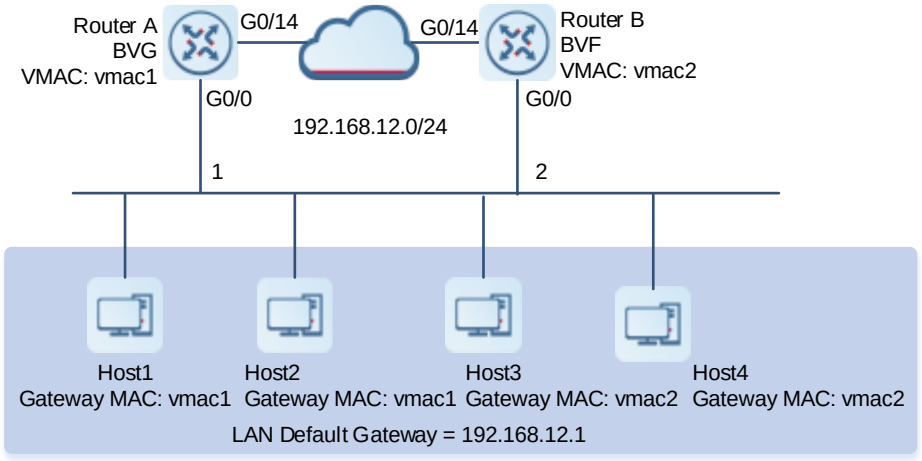
1.5.1 Configuring IPv4 VRRP Plus

1. Requirements

A load balancing policy needs to be enabled in an IPv4 VRRP group.

2. Topology

Figure 1-3 Configuring IPv4 VRRP Plus



3. Notes

- Configure a VRRP group and enable VRRP Plus on both routers A and B. Configure the local IP addresses so that router A becomes the BVG (master) device and router B becomes a BVF (standby) device.
- Configure the weighted load balancing policy for VRRP Plus. Configure a track object for adjusting the weight, and set the weight decrement value to 100.
- Retain default configurations for the weight, upper and lower thresholds, redirection interval, timeout time, and forwarding preemption function of the standby group.
- Set the default gateway addresses of hosts 1 to 4 in the LAN to the virtual IP address of VRRP, that is, 192.168.12.1.

4. Procedure

Perform the following configuration on router A:

```
RouterA> privilege
RouterA# config-system
RouterA(cfg)# interface GigabitEthernet0/0
RouterA(cfg-if-GigabitEthernet 0/0)# unset switch-port
RouterA(cfg-if-GigabitEthernet 0/0)# ip address 192.168.12.3 255.255.255.0
RouterA(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 ip 192.168.12.1
RouterA(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 balance
RouterA(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 load-balancing weighted
```

Perform the following configuration on router B:

```
RouterB> privilege
RouterB# config-system
RouterB(cfg)# interface GigabitEthernet0/0
RouterB(cfg-if-GigabitEthernet 0/0)# unset switch-port
RouterB(cfg-if-GigabitEthernet 0/0)# ip address 192.168.12.2 255.255.255.0
RouterB(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 ip 192.168.12.1
RouterB(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 balance
RouterB(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 load-balancing weighted
```

5. Verification

Run the **display vrrp-config balance** command on Router A to display the configuration of the VRRP Plus standby group. . If the standby group has packet forwarding tasks, **local** is displayed in the **Forwarder** column, and the virtual MAC address allocated to this standby group is also displayed.

```
RouterA# display vrrp-config balance interface GigabitEthernet0/0
State is BVG
  Virtual IP address is 192.168.12.1
  Hello time 1 sec, hold time 3 sec
  Load balancing: weighted
  Redirect time 300 sec, forwarder time-out 14400 sec
  Weighting 100 (configured 100), thresholds: lower 1, upper 100
  There are 2 forwarders
  Forwarder 1 (local)
    MAC address:
      0000.5e00.0101
    Owner ID is 0000.0001.0006
    Preemption disabled (BVG cannot be preempted)
  Forwarder 2
    MAC address:
001a.a916.0201
    Owner ID is 00d0.f822.33a3
Preemption enabled
```

Run the **display vrrp-config balance** command on Router B to display the configuration of the VRRP Plus standby group. If the standby group has packet forwarding tasks, **local** is displayed in the **Forwarder** field, and the virtual MAC address allocated to this standby group is also displayed.

```
RouterB# display vrrp-config balance interface GigabitEthernet0/0
  State is BVF
  Virtual IP address is 192.168.12.1
  Hello time 1 sec, hold time 3 sec
  Load balancing: weighted
  Redirect time 300 sec, forwarder time-out 14400 sec
  Weighting 100 (configured 100), thresholds: lower 1, upper 100
  There are 2 forwarders
  Forwarder 1
    MAC address:
      0000.5e00.0101
```

```
    Owner ID is 0000.0001.0006
    Preemption disabled (BVG cannot be preempted)
  Forwarder 2 (local)
    MAC address:
001a.a916.0201
    Owner ID is 00d0.f822.33a3
    Preemption enabled
```

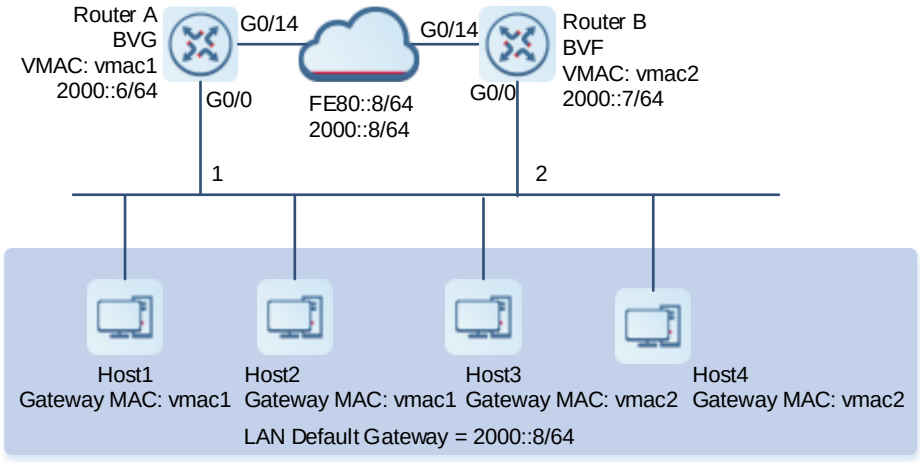
1.5.2 Configuring IPv6 VRRP Plus

1. Requirements

A load balancing policy needs to be enabled in an IPv6 VRRP group.

2. Topology

Figure 1-4 Configuring IPv6 VRRP Plus



3. Notes

- Configure an IPv6 VRRP group and enable IPv6 VRRP Plus on both routers A and B. Configure the priorities so that router A becomes the BVG (master) device and router B becomes a BVF (standby) device.
- Configure the weighted load balancing policy for the IPv6 VRRP Plus standby group.
- Retain default configurations for the weight, upper and lower thresholds, redirection interval, timeout time, and forwarding preemption function of the standby group.
- Hosts 1 to 4 are hosts in the LAN. Set their IPv6 gateway addresses to 2000::8.

4. Procedure

Perform the following configuration on router A:

```
RouterA> privilege
RouterA# config-system
RouterA(cfg)# interface GigabitEthernet0/0
RouterA(cfg-if-GigabitEthernet 0/0)# unset switch-port
RouterA(cfg-if-GigabitEthernet 0/0)# ipv6 address 2000::6/64
```

```
RouterA(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 ipv6fe80::8
RouterA(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 ipv62000::8
RouterA(cfg-if-GigabitEthernet 0/0)# vrrp-config ipv6 1 priority 120
RouterA(cfg-if-GigabitEthernet 0/0)# vrrp-config ipv6 1 load-balancing weighted
RouterA(cfg-if-GigabitEthernet 0/0)# vrrp-config ipv6 1 balance
```

Perform the following configuration on router B:

```
RouterB> privilege
RouterB# config-system
RouterB(cfg)# interface GigabitEthernet0/0
RouterB(cfg-if-GigabitEthernet 0/0)# unset switch-port
RouterB(cfg-if-GigabitEthernet 0/0)# ipv6 address 2000::7/64
RouterB(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 ipv6fe80::8
RouterB(cfg-if-GigabitEthernet 0/0)# vrrp-config 1 ipv62000::8
RouterB(cfg-if-GigabitEthernet 0/0)# vrrp-config ipv6 1 load-balancing weighted
RouterB(cfg-if-GigabitEthernet 0/0)# vrrp-config ipv6 1 balance
```

5. Verification

Run the **display ipv6 vrrp-config balance** command on Router A to display the configuration of the VRRP Plus standby group. If the standby group has packet forwarding tasks, **local** is displayed in the **Forwarder** column, and the virtual MAC address allocated to this standby group is also displayed.

```
RouterA# display ipv6 vrrp-config balance interface GigabitEthernet0/0
GigabitEthernet 0/0 - Group 1
  State is BVG
  Virtual IPv6 address is as follows:
    FE80::8
    2000::8
  Hello time 1 sec, hold time 3 sec
  Load balancing: weighted
  Redirect time 300 sec, forwarder time-out 14400 sec
  Weighting 100 (configured 100), thresholds: lower 1, upper 100
  There are 2 forwarders
  Forwarder 1 (local)
    MAC address:
      0000.5e00.0201
    Owner ID is 00d0.f8fb.96f3
    Preemption disabled (BVG cannot be preempted)
  Forwarder 2
    MAC address:
      1414.4b72.7701
    Owner ID is 00d0.f8fb.6c42
    Preemption enabled
```

Run the **display ipv6 vrrp-config balance** command on Router B to display the configuration of the VRRP Plus standby group. If the standby group has packet forwarding tasks, **local** is displayed in the **Forwarder** field, and the virtual MAC address allocated to this standby group is also displayed.

```
RouterB# display ipv6 vrrp-config balance interface GigabitEthernet0/0
```

```
GigabitEthernet 0/0 - Group 1
  State is BVF
  Virtual IPv6 address is as follows:
    FE80::8
    2000::8
  Hello time 1 sec, hold time 3 sec
  Load balancing: weighted
  Redirect time 300 sec, forwarder time-out 14400 sec
  Weighting 150 (configured 150), thresholds: lower 1, upper 100
  There are 2 forwarders
  Forwarder 1
    MAC address:
      0000.5e00.0201
    Owner ID is 00d0.f8fb.96f3
    Preemption disabled (BVG cannot be preempted)
  Forwarder 2 (local)
    MAC address:
      1414.4b72.7701
    Owner ID is 00d0.f8fb.6c42
Preemption enabled
```

6. Common Errors

- VRRP Plus does not take effect because no VRRP virtual IP address is configured for a group.

Contents

1 Configuring BFD 1

1.1 Introduction 1

1.1.1 Basic Concepts 1

1.1.2 Session Establishment 3

1.1.3 Session Detection 5

1.1.4 BFD Support for Applications 6

1.1.5 BFD Protection 9

1.1.6 BFD Flapping Advertisement Dampening 9

1.1.7 Protocols and Standards 9

1.2 Restrictions and Guidelines 10

1.3 Configuration Task Summary 11

1.4 Notes 11

1.5 Configuring BFD Basic Functions 11

1.5.1 Overview 11

1.5.2 Configuration Tasks 11

1.5.3 Configuring BFD Parameters 11

1.5.4 Configuring the BFD Echo Mode 12

1.5.5 Configuring the BFD Slow Timer 13

1.5.6 Configuring BFD Support for Layer 3 Interfaces 14

1.5.7 Configuring BFD Support for Applications 14

1.6 Configuring BFD Protection 15

1.6.1 Overview 15

1.6.2 Restrictions and Guidelines 15

1.6.3 Prerequisites 16

1.6.4 Procedure 16

1.7 Configuring BFD Flapping Advertisement Dampening 16

1.7.1 Overview 16

1.7.2 Restrictions and Guidelines 16

1.7.3 Procedure 16

1.8 Monitoring 17

1.9 Configuration Examples 17

1.9.1 Configuring BFD Support for OSPF 17

1 Configuring BFD

1.1 Introduction

Bidirectional forwarding detection (BFD) provides a light-load and fast method for detecting the connectivity of the forwarding path between two adjacent routers. It quickly detects faults on the bidirectional forwarding path between two routers for upper layer protocols such as routing protocols and Multi-Protocol Label Switching (MPLS) so that measures can be taken in a timely manner to guarantee services. BFD minimizes the impact of faults on services and improves network availability.

1.1.1 Basic Concepts

1. Packet Format

Detection packets transmitted by BFD are User Datagram Protocol (UDP) packets, which are classified into two types:

- Control packets: The protocol defines a specific format, as shown in Figure 1-1. Currently, there are two versions (v0 and v1) for the format of control packets: v1 is adopted by default for establishing a BFD session; if a device receives packets of v0 from the peer system, the device automatically switches to v0.
- Echo packets: Echo packets are only used by the local system of a BFD session, and the protocol does not define a specific format.

Figure 1-1 BFD Packet Format

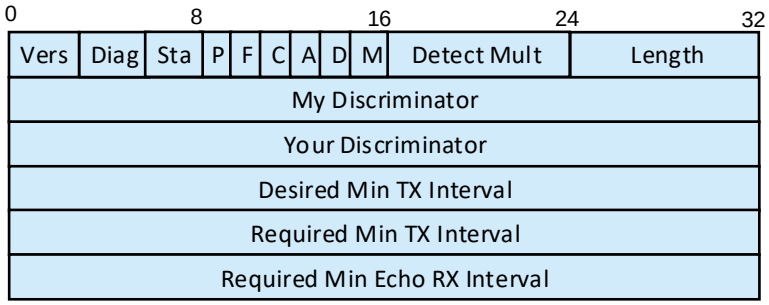


Table 1-1 Description of BFD Packet Fields

Field	Description
Vers	BFD protocol version number, which is 1 currently.

Field	Description
Diag	Causes for the local system to switch from up state to another state last time, including: **0**: Indicates no diagnostic information. **1**: Controls timeout detection. **2**: Indicates that the echo function failed. **3**: Indicates that a neighbor advertises session down. **4**: Indicates forwarding plane reset. **5**: Indicates that the path failed. **6**: Indicates that the connection path failed. **7**: Indicates that the local system enters the AdminDown state.
Sta	BFD local session states, including: **0**: Indicates the AdminDown state. **1**: Indicates the Down state. **2**: Indicates the Init state. **3**: Indicates the Up state.
P	The transmitter sets this flag to **1** in a BFD packet upon parameter changes. The receiver must immediately respond to this packet.
F	This flag must be set to **1** in the response packet for responding to the P flag that is set to **1**.
C	Forwarding/control separation flag. If it is set to **1**, the control plane changes do not affect BFD detection. For example, if the control plane is Open Shortest Path First (OSPF), when OSPF is restarted or experiences graceful restart (GR), BFD can continue to detect the link state.
A	Authentication flag. If it is set to **1**, a session needs to be authenticated.
D	Query request. If it is set to **1**, the transmitter desires the query mode for detecting links.
M	It is used in point-to-multipoint applications in the future. It must be set to **0** currently.
Detect Mult	Detection timeout multiplier. It is used by the detector to calculate the detection timeout time.
Length	Packet length.
My Discriminator	Discriminator of the local end of a BFD session.
Your Discriminator	Discriminator of the remote end of a BFD session
Desired Min Tx Interval	Minimum BFD packet transmission interval supported by the local end.

Field	Description
Required Min RX Interval	Minimum BFD packet receiving interval supported by the local end.
Required Min Echo RX Interval	Minimum echo packet receiving interval supported by the local end. It is set to **0** if the local end does not support the echo function.
Auth Type	Authentication type. It is an optional field and can be set to the following values: Simple Password Keyed MD5 Meticulous Keyed MD5 Keyed SHA1 Meticulous Keyed SHA1
Auth Length	Authentication data length.
Authentication Data	Authentication data area.

2. Session State

BFD migrates the state machine based on the local session state and the BFD packets received from the peer end. A BFD state machine is established and torn down using a three-way handshake mechanism, to ensure that both ends know the state change. A BFD session can be in any of the following four basic states:

- Down: Indicates that a session is in the down state or is established just now.
- Init: Indicates that the local system has communicated with the peer system and desires to bring the session to the up state.
- Up: Indicates that a session has been established successfully.
- AdminDown: Indicates that a session is in the AdminDown state.

3. Transmission Interval and Detection Time

During the establishment of a BFD session, both ends negotiate about BFD parameters to determine the transmission interval and detection time. After a BFD session is established, both ends dynamically negotiate about BFD parameters (for example, the minimum transmission interval and minimum receiving interval). After protocols at both ends transmit relevant negotiation packets, they adopt the new transmission interval and detection time, without affecting the current state of the session.

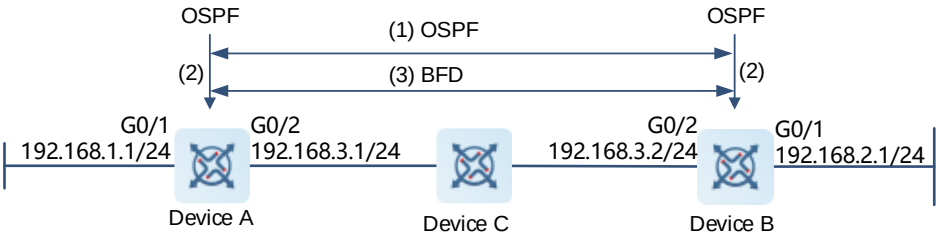
1.1.2 Session Establishment

BFD detection starts from the establishment of a BFD session.

1. Session Establishment Process

BFD itself is unable to discover neighbors. It needs an upper-layer protocol to specify the neighbor, with which BFD needs to establish a session. As shown in Figure 1-2, two routers running OSPF and BFD are connected through an Layer 2 switch.

Figure 1-2 BFD Session Establishment



The BFD session establishment process is as follows:

- OSPF discovers a neighbor and establishes a connection with the neighbor.
- OSPF instructs BFD to establish a session with the neighbor.
- BFD establishes a session with the neighbor.

2. Session Establishment Mode

The BFD protocol specifies that a BFD session can be established in two modes:

- Active mode: Before the establishment of a session, BFD actively transmits a control packet for establishing a BFD session regardless of whether it receives a control packet for establishing a BFD session from the peer end.
- Passive mode: BFD does not actively transmit a control packet for establishing a BFD session before a session is established but waits till it receives a control packet for establishing a BFD session from the peer end.

Specification

The passive mode is not supported currently.

3. Negotiation of Session Parameters

Both ends negotiate about BFD session parameters during the establishment of a BFD session, to determine the transmission interval and detection time. Pay attention to the following points:

- BFD session parameters (including the minimum transmission interval, minimum receiving interval, and detection timeout multiplier) must be set for interfaces at both ends. Otherwise, a BFD session cannot be established. If the BFD session is a multi-hop session, there is no need to configure BFD session parameters on the interface. The default global multi-hop session parameters will be used instead.
- Interfaces at both ends negotiate about BFD session parameters and detect the session based on the parameters during the establishment of a BFD session. If the BFD session is a multi-hop session, the default global multi-hop session parameters will be used for negotiation and session detection.
- After a BFD session is established, both ends dynamically negotiate about BFD parameters (for example, the minimum transmission interval and minimum receiving interval). After protocols at both ends transmit relevant negotiation packets, they adopt the new transmission interval and detection time, without affecting the current state of the session.
- BFD session parameters can be updated only after all the original sessions on a port complete negotiation; otherwise an error is prompted. You can run a related command to query the negotiation status of the original

sessions.

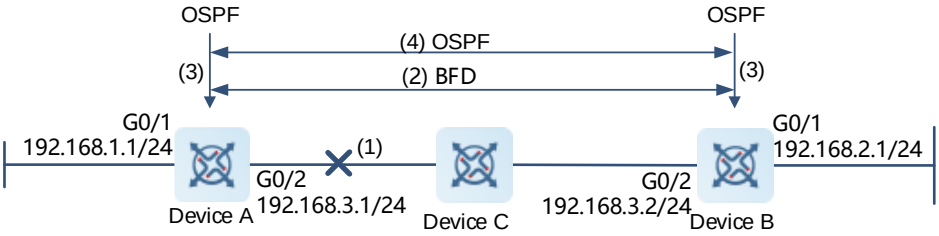
1.1.3 Session Detection

Link detection starts after the establishment of a BFD session. BFD periodically transmits BFD control packets. If it fails to receive BFD packets from the peer end within the detection time, it deems that the session is down and instructs the associated application to accelerate the convergence.

1. Detection Process

As shown in Figure 1-3, two routers running OSPF and BFD are connected through an L2 switch.

Figure 1-3 Topology for BFD Session Detection



The handling process after a BFD session is down is as follows:

- The link between router A and the switch fails.
- The BFD session between routers A and B detects the failure.
- BFD notifies the local OSPF that the forwarding path to the neighbor is faulty.
- OSPF processes the neighbor down situation. If a backup forwarding path is available, it starts protocol convergence to enable the backup forwarding path.

2. Detection Mode

BFD supports the following detection modes:

- Asynchronous mode

  In asynchronous mode, systems transmit BFD control packets periodically to each other. If a system fails to receive BFD control packets from the peer end within the detection time, it advertises that the session is down.

- Query mode

  In query mode, it is assumed that each system has an independent method for confirming its connection with other systems. After a BFD session is established, the system stops transmitting BFD control packets unless it needs to explicitly verify the connectivity. In such a case, the system transmits a shot-sequence BFD control packet. If the system fails to receive a returned packet within the detection time, it advertises that the session is down. If it receives a response from the peer end, the forwarding path is reachable.

- Echo mode

  In echo mode, the local system periodically transmits BFD echo packets and a remote system receives and loops back the packets through the forwarding path. If the local system fails to receive several consecutive echo packets within the detection time, it advertises that the session is down. The echo function can be used together with the preceding two detection modes. The echo packet detection function does not require

the involvement of the control plane of the remote system. Packets are returned by the forwarding plane of the remote system, which reduces the delay and ensures faster fault detection in comparison with transmission of control packets. Enabling the echo function in asynchronous mode greatly reduces transmission of control packets because the detection is accomplished by the echo function. Enabling the echo function in query mode thoroughly cancels transmission of control packets after a session is established. The echo function must be enabled at both ends of a BFD session. Otherwise, the echo function does not take effect.

The echo mode includes the echo mode and one-armed echo mode. The one-armed echo mode is applicable to the single-hop IP link scenario and generally used in two directly connected devices, where one device supports the BFD function and the other does not. The echo mode requires that the echo mode be configured for both detection parties.

Specification

- The query mode is not supported and cannot be configured at present.
- Only BFD session version 1 supports the BFD echo mode.
- The echo mode is not supported for an IPv6 BFD session with the link-local address as the source or destination address.

1.1.4 BFD Support for Applications

By supporting BFD, the associated applications can utilize the fast fault detection of BFD to improve the protocol convergence performance. In general, the fault detection time can be shortened to less than 1 second. After BFD support for a certain application is enabled, BFD rapidly detects faults occurring on an established BFD session. When a link fault occurs, BFD can rapidly identify the fault and notify the associated application to process the fault, thereby improving its convergence. Currently, BFD supports the following applications:

1. BFD Support for RIP

After BFD support for the Routing Information Protocol (RIP) is enabled, RIP can utilize the faster fault detection feature of BFD than the hello mechanism of RIP to improve the protocol convergence. In general, the fault detection time can be shortened to less than 1 second.

Note

For more details about BFD support for RIP, see *Configuring RIP*.

2. BFD Support for OSPF

After BFD support for OSPF is enabled, OSPF can utilize the faster fault detection feature of BFD than the hello mechanism of OSPF to improve the protocol convergence. In general, the fault detection time can be shortened to less than 1 second.

Note

For more details about BFD support for OSPF, see *Configuring OSPFv2*.

3. BFD Support for OSPFv3

After BFD support for OSPFv3 is enabled, OSPFv3 can utilize the faster fault detection feature of BFD than the hello mechanism of OSPFv3 to improve the protocol convergence. In general, the fault detection time can be shortened to less than 1 second.

Note

For more details about BFD support for OSPFv3, see *Configuring OSPFv3*.

4. BFD Support for BGP

After BFD support for Border Gateway Protocol (BGP) is enabled, BGP can utilize the faster fault detection feature of BFD than the hello mechanism of BGP to improve the protocol convergence. In general, the fault detection time can be shortened to less than 1 second.

Note

For more details about BFD support for BGP, see *Configuring BGP*.

5. BFD Support for IS-IS

The Intermediate System to Intermediate System (IS-IS) protocol dynamically discovers neighbors through hello packets. After the BFD function is enabled on IS-IS, a BFD session is set up with a neighbor in up state. The BFD mechanism is used to detect the neighbor state. Once a neighbor failure is detected through BFD, IS-IS performs network convergence immediately. The convergence time can be reduced from 30s to less than 1s. By default, IS-IS hello packets are sent at an interval of **10** seconds in a P2P network, and the time required to detect a neighbor failure is three times the hello packet transmission interval.

Note

For more details about BFD support for IS-IS, see *Configuring IS-IS*.

6. BFD Support for Static Routing

After BFD support for static routing is enabled, BFD prevents routers from selecting an unavailable static route as the forwarding path during routing and enables routers to rapidly switch to an available backup forwarding path. Different from dynamic routing protocols, static routing does not have the ND mechanism. Therefore, after BFD support for static routing is configured, the next-hop reachability of a static route relies on the BFD session state. If a BFD session detects a fault, the next hop of a static route is unreachable and the static route is not added to the routing information base (RIB). If the remote system deletes a BFD session during the establishment of the BFD session, the BFD session becomes down. In this case, the system ensures that the forwarding behavior of static routing is not affected.

Note

For more details about BFD support for static routing, see *Configuring Static Routing*.

7. BFD Support for PBR

After BFD support for policy-based routing (PBR) is configured, BFD prevents routers from selecting an unavailable policy-based route as the forwarding path during routing and enables routers to rapidly switch to an available backup forwarding path. BFD support for PBR is equivalent to BFD support for static routing. BFD tracks and detects the forwarding path to a specified neighbor. When a BFD session fails, BFD notifies the PBR that the next hop is unreachable. Then, the policy-based route to the next hop does not take effect. If the remote system deletes a BFD session during the establishment of the BFD session, the BFD session becomes down. In this case, the system ensures that the PBR forwarding behavior is not affected.

Note

For more details about BFD support for PBR, see *Configuring PBR*.

8. BFD Support for VRRP

The BFD support for the Virtual Router Redundancy Protocol (VRRP) can replace the hello mechanism of VRRP to rapidly detect the running status of the active and standby routers. When a fault occurs, it accelerates the active/standby router switching and improves network performance. In general, the fault detection time can be shortened to less than 1 second. VRRP can also utilize BFD to track a specified neighbor. If a BFD session detects that the forwarding path to the neighbor fails, VRRP automatically lowers the VRRP priority to a certain extent to trigger active/standby router switching. This configuration takes effect only when a dynamic routing protocol or other applications notify BFD to establish a session with a neighbor.

Note

For more details about BFD support for VRRP, see *Configuring VRRP*.

9. BFD Support for VRRP Plus

The BFD support for VRRP Plus can replace the balancing virtual forwarder (BVF) detection conducted by the balancing virtual gateway (BVG) of VRRP Plus to rapidly detect the running status of BVFs. When a fault occurs, it accelerates the forwarding entity switching and improves network performance. In general, the fault detection time can be shortened to less than 1 second. VRRP Plus is based on the VRRP protocol. Therefore, no additional configuration is required for BFD support for VRRP Plus and only VRRP needs to be enabled on devices at both ends and a BFD session is correctly associated.

Note

For more details about BFD support for VRRP Plus, see *Configuring VRRP Plus*.

10. BFD Support for MPLS

The BFD support for Multiprotocol Label Switching (MPLS) enables label switched paths (LSPs) to use BFD to rapidly detect the neighbor status. The following detection modes are supported:

- BFD detects static LSPs.
- BFD detects LSPs generated by the Label Distribution Protocol (LDP).
- BFD detects reverse LSPs by using the IP addresses.

Note

For more details about BFD support for MPLS, see *Configuring MPLS Basics*.

11. BFD Support for Layer 3 Interfaces

BFD supports modifying the status of a Layer 3 interface. In interface configuration mode, run the **bfdbindpeer-ip** command to detect the address of a specified directly connected Layer 3 interface. After this CLI command is executed, a BFD session is created and the status of an Layer 3 interface can be changed based on the detection result of the BFD session, for example, BFD Down or BFD Up. This function is often used in various types of fast reroute (FRR), which use BFD to detect the interface status to implement FRR switching.

Note

Only LDP FRR switching is supported when BFD support for Layer 3 interfaces is configured.

12. BFD Support for AP Member Ports

After BFD support for aggregate port (AP) member ports is enabled, BFD can rapidly detect a fault occurring on a member port link so that traffic on this link is rapidly distributed to other effective member port links. In general, the fault detection time can be shortened to less than 1 second.

Note

For more details about BFD support for AP, see *Configuring Aggregate Interface*.

1.1.5 BFD Protection

If a BFD-enabled device is attacked (for example, attacked by a large number of ping packets), BFD sessions will flap. You can enable the BFD protection to provide protection. If both BFD and BFD protection are enabled on a device, the device discards the BFD packets received from the previous-hop device, which affects the establishment of a BFD session between the previous-hop device and other devices.

1.1.6 BFD Flapping Advertisement Dampening

A BFD session may frequently switch between the down and up states due to link instability. As a result, an associated application (such as static routing) may frequently switch forwarding paths, causing the frequent switching of the BFD session between the down and up states. This function allows users to set the delay for status change advertisement. After a BFD session is up for a certain period of time, BFD informs an associated application of BFD up. Otherwise, BFD informs it of BFD down.

1.1.7 Protocols and Standards

- draft-ietf-bfd-base-09: Bidirectional Forwarding Detection
- draft-ietf-bfd-generic-05: Generic Application of BFD
- draft-ietf-bfd-mib-06: Bidirectional Forwarding Detection Management Information Base
- draft-ietf-bfd-v4v6-1hop-09: BFD for IPv4 and IPv6 (Single Hop)
- draft-ietf-bfd-multihop-07: BFD for IPv4 and IPv6 (Multihop)

- draft-ietf-bfd-mpls-07: BFD For MPLS LSPs

Specification

Currently, draft-ietf-bfd-mib-06 and draft-ietf-bfd-multihop-07 are not supported.

1.2 Restrictions and Guidelines

- You are advised to keep the parameter configuration consistent at both ends of a BFD session. The purpose is to ensure that application protocols associated with BFD take effect simultaneously and prevent occurrence of one-way forwarding path due to different dampening time at both ends.
- Transmission bandwidth differences of different interfaces need to be taken into account during parameter configuration. If the configured minimum Tx interval and minimum Rx interval are very small, BFD may occupy excessive bandwidth and affect data transmission.
- BFD session parameters can be configured only after all the sessions on the current port complete negotiation. Otherwise, a configuration error is prompted. Run the **display bfd adjacencies parm-consult** command to display the negotiation status of all the sessions.
- Ensure that BFD support is enabled on neighbors of a BFD session. Otherwise, a BFD session cannot be established. If a dynamic routing protocol or another application instructs BFD to establish a session with a neighbor, the BFD session can also be established.
- For a single-hop session whose source and destination IP addresses belong to the same network segment, if the interface specified by a BFD session is different from the actual BFD packet outbound interface because of IP routing, or if the interface specified during BFD session creation is different from the actual BFD packet inbound interface, a BFD session cannot be established. For a multi-hop session whose source and destination IP addresses belong to different network segments, associated applications do not need to specify an interface and BFD transmits and receives packets based on IP routing. Such BFD sessions can be established provided that returned BFD packets can be received.
- When BFD session traffic is sent and no BFD session parameters are configured on the device (the **display bfd adjacencies** command shows that the number of BFD sessions is 0), the CPU protect policy (CPP) entries are not delivered. Therefore, the BFD traffic sent to the local device is not matched with the CPP entries and not sent to the local device. In this case, BFD sessions must be configured for sending the traffic to the local device.
- In the process that the forwarding plane of the peer device returns echo packets transmitted by the local end to the local end, the echo packets may be lost due to congestion of the peer device, causing a session detection failure. In this case, you need to configure a quality of service (QoS) policy to ensure that echo packets are processed preferentially or disable the echo function.
- The echo detection function of BFD does not support multi-hop detection. Ensure that the echo function is disabled when configuring multi-hop detection.
- The echo mode takes effect only after it is enabled at both ends of a BFD session.
- Before enabling the echo mode of BFD, run the **unset ip redirects** command on the neighbors of a BFD session to disable the function of sending Internet Control Message Protocol (ICMP) redirection packets, and run the **unset ip deny land** command to disable the distributed denial of service (DDoS) function to prevent the land-based attack.

1.3 Configuration Task Summary

BFD configuration includes the following tasks:

(1) Configuring BFD Basic Functions

- Configuring BFD Parameters
- (Optional) Configuring the BFD Echo Mode
- (Optional) Configuring the BFD Slow Timer
- (Optional) Configuring BFD Support for Layer 3 Interfaces
- Configuring BFD Support for Applications

(2) (Optional) Configuring BFD Protection

(3) (Optional) Configuring BFD Flapping Advertisement Dampening

1.4 Notes

This section describes notes for configuring BFD functions.

Table 1-2 describes BFD functions supported by different products.

Table 1-2 BFD Functions Supported by Different Products

Function	Product Supporting the Function
Configuring the BFD Echo Mode	S6980-64QC, S6920-4C, S6580-48CQ8QC, S6580-24DC8QC, S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ

1.5 Configuring BFD Basic Functions

1.5.1 Overview

You can establish BFD sessions and configure BFD support for applications to perform link fault detection.

1.5.2 Configuration Tasks

The basic function configuration of BFD includes the following tasks:

(1) Configuring BFD Parameters

(2) (Optional) Configuring the BFD Echo Mode

(3) (Optional) Configuring the BFD Slow Timer

(4) (Optional) Configuring BFD Support for Layer 3 Interfaces

(5) Configuring BFD Support for Applications

1.5.3 Configuring BFD Parameters

1. Overview

- Unless otherwise specified, for single-hop sessions, a BFD session should be established between routers

at both ends to be detected by BFD and BFD parameters must be configured at the egresses. Multi-hop sessions use global parameters for session establishment by default. If different negotiation parameters need to be configured for different multi-hop sessions, BFD templates need to be associated with the sessions.

- Transmission bandwidth differences of different interfaces need to be taken into account during parameter configuration. If the configured minimum Tx interval and minimum Rx interval are very small, BFD may occupy excessive bandwidth and affect data transmission.

2. Restrictions and Guidelines

- The fast forwarding function must be enabled before the BFD function is enabled on routers.
- Negotiation parameters can be configured only after all the original sessions on a port complete negotiation.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure parameters for the BFD multi-hop session.

bfd multi-hop min_tx_interval *milliseconds* **min_rx_interval** *milliseconds* **multiplier** *multiplier-value*

(4) By default, the minimum interval to send BFD control packets to BFD peers is 999 ms, the minimum interval to receive BFD control packets from BFD peers is 999 ms, and the multiplexer is 5.Enter the Layer 3 Ethernet interface configuration mode.

interface *interface-type interface-number*

(5) Configure BFD parameters.

bfd min_tx_interval *milliseconds* **min_rx_interval** *milliseconds* **multiplier** *interval-multiplier*

No BFD session parameters are configured by default.

1.5.4 Configuring the BFD Echo Mode

Note

BFD echo mode is supported by only the S6980-64QC, S6920-4C, S6580-48CQ8QC, S6580-24DC8QC, S6520-64CQ. S6510-32CQ, S6510-4C, S6510-48VS8CQ, S6250-48XS8CQ.

1. Overview

- (Optional) Ports run in asynchronous mode by default. If a BFD session needs to run in echo mode, the echo mode needs to be configured. The echo mode includes the echo mode and one-armed echo mode.
- Complete the configuration on ports of switches or routers.
- A BFD session runs in asynchronous mode as long as either of routers at both ends is configured to run in asynchronous mode. If routers at both ends are configured to run in echo mode by default, a BFD session finally runs in echo mode. This function is disabled by default and can be enabled after the **bfd echo** command is executed.
- Run the **bfd echo one-arm** command to enable the one-armed echo mode.

2. Restrictions and Guidelines

By default, when BFD session parameters are configured, the system automatically enables the echo mode. To disable the BFD echo function, run the **unset bfd echo** command in interface configuration mode. The minimum Tx interval and minimum Rx interval of echo packets adopt the **Interval milliseconds** and **min_rx_interval milliseconds** parameters configured for a session.

Before enabling the echo mode of BFD, run the **unset ip redirects** command on the neighbors of a BFD session to disable the function of sending ICMP redirection packets, and run the **unset ip deny land** command to disable the DDoS function to prevent the land-based attack.

The BFD one-armed echo function and **ip redirect-drop** cannot be configured at the same time, otherwise the function fails.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 3 Ethernet interface configuration mode.

interface *interface-type interface-number*

(4) Configure the BFD echo mode.

bfd echo [**one-arm**]

The echo mode is disabled for a BFD session by default.

1.5.5 Configuring the BFD Slow Timer

1. Overview

- The default slow timer is **3000** milliseconds. The value can be changed as required.
- Configure this function in global configuration mode of switches or routers.
- In BFD echo mode or during session establishment, control packets are sent based on the slow timer. A larger value indicates that the required time for negotiating and establishing a BFD session is longer and the time required for transmitting slow BFD packets in echo mode is longer.

2. Restrictions and Guidelines

This command is used to specify the slow timer in echo mode.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the BFD slow timer.

bfd slow-timer [*interval*]

The default slow timer time in the echo mode is **2000** milliseconds.

1.5.6 Configuring BFD Support for Layer 3 Interfaces

1. Overview

- Currently, this function is used only when MPLS LDP is used for FRR switching.
- Complete the configuration on ports of switches or routers.

2. Restrictions and Guidelines

This command is used to enable BFD support for Layer 3 interfaces so as to rapidly detect connectivity of Layer 3 interfaces.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure BFD support for Layer 3 interfaces.

bfd bind peer-ip *ipv4-address* [**source-ip** *ipv4-address*] **process-pst**

BFD support for interfaces is not configured by default.

1.5.7 Configuring BFD Support for Applications

1. Overview

- This function is disabled by default.
- The configuration command varies with the associated applications. For details, see their configuration guides.
- This function must be configured at both ends so that a BFD session can be established.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure BFD support.

bfd all interfaces

In RIP routing configuration mode, enable BFD support for RIP on all interfaces. For details, see *Configuring RIP*.

bfd all interfaces

In OSPF routing configuration mode, enable BFD support for OSPF on all interfaces. For details, see *Configuring OSPFv2*.

bfd all interfaces

In OSPFv3 routing configuration mode, enable BFD support for OSPFv3 on all interfaces. For details, see *Configuring OSPFv3*.

neighbor *address* **fall-over bfd**

In BGP routing configuration mode, enable BFD support for BGP. For details, see *Configuring BGP*.

bfd all interfaces

In IS-IS routing configuration mode, enable BFD support for IS-IS on all interfaces. For details, see *Configuring IS-IS*.

ip route-config static bfd [**vrf** *vrf-name*] *interface-type interface-number gateway* [**source** *ip-address*]

In global configuration mode, enable BFD support for static routing. For details, see *Configuring Static Routing*.

ipv6 route static bfd [**vrf** *vrf-name*] *interface-type interface-number gateway* [**source** *ipv6-address*]

In global configuration mode, enable BFD support for static IPv6 routing. For details, see *Configuring Static Routing*.

set ipnext-hopverify-availability next-hop-address **bfd** [**vrf** *vrf-name*] interface-type interface-number gateway

Enable BFD support for PBR. For details, see *Configuring PBR*.

set ipv6 next-hopverify-availability *next-hop-address* **bfd** [**vrf** *vrf-name*] *interface-type interface-number gateway*

Enable BFD support for IPv6 PBR. For details, see *Configuring PBR*.

vrrp bfd *interface-type interface-number ip-address*

Enable BFD support for VRRP. For details, see *Configuring VRRP*.

VRRP Plus is based on the VRRP protocol. Therefore, no additional configuration is required for BFD support for VRRP Plus and only VRRP needs to be enabled on devices at both ends and a BFD session is correctly associated.

1.6 Configuring BFD Protection

1.6.1 Overview

- If a BFD-enabled device is attacked (for example, attacked by a large number of ping packets), BFD sessions flap. In this case, you can configure BFD protection to provide protection.

1.6.2 Restrictions and Guidelines

- The BFD basic functions must be configured.
- If both BFD and BFD protection are enabled on a device, the device discards the BFD packets received from the previous-hop device, which affects the establishment of a BFD session between the previous-hop device and other devices.

1.6.3 Prerequisites

Basic functions of BFD have been configured.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure BFD protection.

bfd cpp

Configure BFD protection on a switch in the network where attacks exist.

The BFD protection is enabled by default.

1.7 Configuring BFD Flapping Advertisement Dampening

1.7.1 Overview

- A BFD session may frequently switch between the down and up states due to link instability. As a result, a relevant application (such as static routing) may frequently switch forwarding paths and the running services are affected.
- Users can set a delay for status change advertisement, after which BFD informs an associated application of BFD up. After a BFD session is up for a certain period of time, BFD informs an associated application of BFD up. Otherwise, BFD informs it of BFD down. This aims to reduce flapping of associated protocols caused by unstable links.

1.7.2 Restrictions and Guidelines

- The BFD basic functions must be configured.
- If a BFD session does not frequently switch between the down and up states, enabling BFD flapping advertisement dampening will delay notifying an associated application of BFD up.
- Enabling BFD flapping advertisement dampening can relieve frequent processing of associated applications (such as route re-calculation) caused by frequent advertisement of BFD status changes. The larger the configured time is, the longer the required BFD stability time is. BFD notifies an application module of BFD up only after the stability time reaches the configured time.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure flapping dampening parameters.

bfd up-dampening { *milliseconds* }

The BFD flapping advertisement dampening is disabled on ports by default. If a BFD session frequently switches between the down state and up state, you are advised to enable this function.

1.8 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-3 BFD Monitoring

Command	Purpose
display bfd adjacencies [**vrf** *vrf-name*] [**protocol** { **ap** | **backward**-**lsp**-**with**-**ip** | **bgp** | **bgp**-**lsp** | **dhcp** | **isis** | **ldp**-**lsp** | **openflow** | **ospf** | **ospfv3** | **pbr** | **pimdm** | **pimsm** | **pimsmv6** | **pst** | **rip** | **srp** | **srp-v6** | **static**-**lsp** | **static**-**route** | **vrrp** | **vrrp**-**balance** }] [**ipv4** *ipv4-address* | **ipv6** *ipv6-address*] [**details**] [**parm**-**consult** [*interface-type interface-number*]]	Displays BFD session information.
debugging bfd event [**interface** *interface-type interface-number* **| ipv4** *ipv4-address* **| ipv6** *ipv6-address*]	Debugs BFD events.
debugging bfd packet [**interface** *interface-type interface-number* **| ipv4** *ipv4-address* **| ipv6** *ipv6-address*]	Debugs BFD packets.

1.9 Configuration Examples

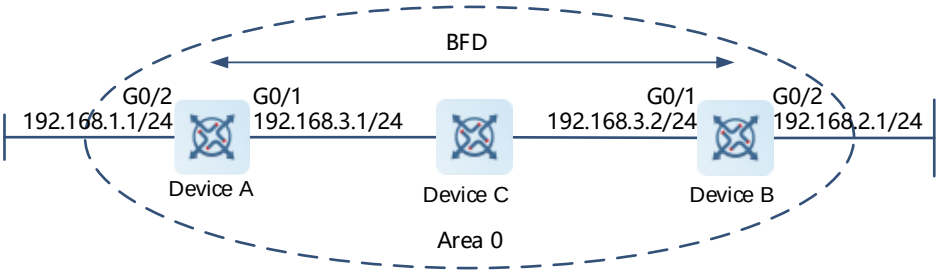
1.9.1 Configuring BFD Support for OSPF

1. Requirements

Three devices are used to build a network, the devices run the OSPF protocol, and BFD support for OSPF is enabled on two devices.

2. Topology

Figure 1-4 Configuring BFD Support for OSPF



3. Notes

- Configure IP addresses for ports connecting devices A and B.
- Run the OSPF protocol on devices A and B.
- Configure BFD parameters for the interconnected ports of devices A and B.
- Enable BFD support for OSPF on devices A and B.

4. Procedure

Perform the following configuration on device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet0/1
DeviceA(cfg-if-GigabitEthernet0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet0/1)# ip address 192.168.3.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet0/1)# bfd min_tx_interval 200 min_rx_interval 200 multiplier 5
DeviceA(cfg-if-GigabitEthernet0/1)# exit
DeviceA(cfg)# interface GigabitEthernet0/2
DeviceA(cfg-if-GigabitEthernet0/2)# unset switch-port
DeviceA(cfg-if-GigabitEthernet0/2)# ip address 192.168.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet0/2)# exit
DeviceA(cfg)# router-protocol ospf 123
DeviceA(cfg-router)# log-adj-changes detail
DeviceA(cfg-router)# network-advertise 192.168.3.0 0.0.0.255 area 0
DeviceA(cfg-router)# network-advertise 192.168.1.0 0.0.0.255 area 0
DeviceA(cfg-router)# bfd all-interfaces
```

Perform the following configuration on device B.

```
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitethernet2/1
DeviceB(cfg-if-GigabitEthernet0/1)# unset switch-port
DeviceB(cfg-if-GigabitEthernet0/1)# ip address 192.168.3.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet0/1)# bfd min_tx_interval 200 min_rx_interval 200 multiplier 5
DeviceB(cfg-if-GigabitEthernet0/1)# exit
DeviceB(cfg)# interface GigabitEthernet0/2
```

```
DeviceB(cfg-if-GigabitEthernet0/2)# unset switch-port
DeviceB(cfg-if-GigabitEthernet0/2)# ip address 192.168.2.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet0/2)# exit
DeviceB(cfg)# router-protocol ospf 123
DeviceB(cfg-router)# log-adj-changes detail
DeviceB(cfg-router)# network-advertise 192.168.3.0 0.0.0.255 area 0
DeviceB(cfg-router)# network-advertise 192.168.2.0 0.0.0.255 area 0
DeviceB(cfg-router)# bfd all-interfaces
```

5. Verification

Run the **display bfd adjacencies detail** command to check whether BFD support configuration is correct.

```
DeviceA# display bfd adjacencies detail
OurAddr          NeighAddr      LD/RD    RH/RS    Holdown(mult)  State     Int
192.168.3.1      192.168.3.2    1/2      Up       532 (3 )       Up        Ge0/1
Local Diag: 0, Demand mode: 0, Poll bit: 0
MinTxInt: 200000, MinRxInt: 200000, Multiplier: 5
Received MinRxInt: 50000, Received Multiplier: 3
Holdown (hits): 600(22), Hello (hits): 200(84453)
Rx Count: 49824, Rx Interval (ms) min/max/avg: 208/440/332
Tx Count: 84488, Tx Interval (ms) min/max/avg: 152/248/196
Registered protocols: OSPF
Uptime: 02:18:49
Last packet: Version: 1 - Diagnostic: 0
I Hear You bit: 1              - Demand bit: 0
Poll bit: 0                    - Final bit: 0
Multiplier: 3              - Length: 24
My Discr.: 2                       - Your Discr.: 1
Min tx interval: 50000         - Min rx interval: 50000
Min Echo interval: 0
```

```
DeviceB# display bfd adjacencies detail
OurAddr          NeighAddr      LD/RD    RH/RS    Holdown(mult)  State     Int
192.168.3.2 192.168.3.1    2/1      Up       532 (5 )       Up        Ge2/1
Local Diag: 0, Demand mode: 0, Poll bit: 0
MinTxInt: 50000, MinRxInt: 50000, Multiplier: 3
Received MinRxInt: 200000, Received Multiplier: 5
Holdown (hits): 600(22), Hello (hits): 200(84453)
Rx Count: 49824, Rx Interval (ms) min/max/avg: 209/440/332 last: 66 ms ago
Tx Count: 84488, Tx Interval (ms) min/max/avg: 153/249/197 last: 190 ms ago
Registered protocols: OSPF
Uptime: 02:18:49
Last packet: Version: 1 - Diagnostic: 0
I Hear You bit: 1              - Demand bit: 0
Poll bit: 0                    - Final bit: 0
Multiplier: 5              - Length: 24
```

```
My Discr.: 1                    - Your Discr.: 2
Min tx interval: 200000         - Min rx interval: 200000
Min Echo interval: 0
```

6. Common Errors

- For a single-hop session, BFD parameters may not be configured for the port on the device at one end of the session.
- If different negotiation parameters need to be configured for different multi-hop sessions, templates need to be associated with the sessions and negotiation parameters need to be configured in the templates.
- For multi-hop BFD sessions, if the sessions use different templates, negotiation parameters in the template used by a session that becomes up first are used for transmitting and receiving packets.
- BFD support for applications is not configured.
- BFD support for applications is enabled only on the device at one end.

Contents

1 Configuring Track 1

1.1 Introduction 1

1.1.1 Basic Concepts 1

1.1.2 ICMP Echo Test 1

1.1.3 UDP Echo Test 1

1.1.4 DHCP Test 2

1.1.5 TCP Test 2

1.1.6 DNS Test 2

1.1.7 Track Support for the RNS 3

1.2 Configuration Task Summary 4

1.3 Configuring Basic Features of RNS 4

1.3.1 Overview 4

1.3.2 Restrictions and Guidelines 4

1.3.3 Procedure 4

1.4 Configuring the ICMP Echo Test 5

1.4.1 Overview 5

1.4.2 Restrictions and Guidelines 5

1.4.3 Procedure 5

1.5 Configuring the DNS Test 6

1.5.1 Overview 6

1.5.2 Restrictions and Guidelines 6

1.5.3 Procedure 6

1.6 Configuring the TCP Connection Test 7

1.6.1 Overview 7
1.6.2 Restrictions and Guidelines 7
1.6.3 Procedure 7
1.7 Configuring the UDP Echo Test 8
1.7.1 Overview 8
1.7.2 Restrictions and Guidelines 8
1.7.3 Procedure 8
1.8 Configuring the DHCP Test 8
1.8.1 Overview 8
1.8.2 Restrictions and Guidelines 8
1.8.3 Procedure 8
1.9 Configuring the Track Support for the RNS 9
1.9.1 Overview 9
1.9.2 Restrictions and Guidelines 9
1.9.3 Procedure 9
1.10 Monitoring 10
1.11 Configuration Examples 11
1.11.1 Configuring the ICMP Echo Test 11
1.11.2 Configuring the DNS Test 12
1.11.3 Configuring the TCP Connect Test 13
1.11.4 Configuring the UDP Echo Test 14
1.11.5 Configuring the DHCP Test 15
1.11.6 Configuring the Device to Track Link Status of an Interface 16
1.11.7 Configuring the Device to Track the Status of Multiple Interfaces 17

1.11.8 Configuring the Device to Track A Single RNS Instance 18

1.11.9 Configuring the Device to Track the Results of Multiple RNS Instances..... 19

1 Configuring Track

1.1 Introduction

The reliable network service (RNS) tests the services provided by a peer device to monitor the service availability, integrity of the end-to-end connection, and service quality. With the RNS test results, you can learn the network performance in time and take measures accordingly to handle related network performance problems. Meanwhile, you can diagnose and locate network faults. The RNS test in this document is also called track test.

1.1.1 Basic Concepts

1. RNS Instance

An RNS instance can be treated as an RNS process. Before performing the RNS, you must create an RNS instance. In the RNS instance, you need to configure the RNS parameters, such as the test type, test destination address, and test frequency. The instance ID is globally unique.

2. RNS Test

RNS test aims to monitor the network connectivity, service availability, integrity of end-to-end connection, and service quality. For example, the RNS tests whether the domain name server (DNS) function of the device is normal. Currently, the RNS supports the following types of tests: RNS Test, DNS, TCP, UDP-Echo, and HTTP.

1.1.2 ICMP Echo Test

The Internet Control Message Protocol (ICMP) echo test is a basic function of the RNS, and is implemented in compliance with the RFC 2925. An ICMP packet is sent to check whether the destination is reachable and to calculate the network response time and packet loss rate.

An ICMP echo request packet is sent to the destination IP address based on the preset test time and frequency. Upon receipt of the ICMP echo request packet, an ICMP echo reply packet is returned from the destination IP address. Through the ICMP echo test, the response time and packet loss rate is calculated based on the information relating to the received ICMP echo reply packet, for example, the receipt time and number of packets. In this way, the current network performance and status are reflected. The ICMP echo test results and historical records are saved, and you can use the command line to display them.

 Caution

The prerequisite for a successful ICMP echo test is that destination devices can correctly respond to ICMP echo request packets.

1.1.3 UDP Echo Test

User Datagram Protocol (UDP) echo is used to test the network accessibility and delay. When UDP packets are used to test the network accessibility and delay, the RNS server must be enabled on the peer end with the UDP port enabled, or the RNS server must be able to respond to ICMP port unreachable packets.

A UDP packet is sent to the destination IP address based on the preset test time and frequency. Upon receipt of the UDP packet, the device of the destination IP address replies by using such packet. Through the UDP echo test, the time required for packet round trip and packet loss rate are calculated based on the information relating to the received UDP packet from the server to reflect the current network performance and network conditions. The UDP echo test results and historical records are saved, and you can use the command line to display them.

1.1.4 DHCP Test

The Dynamic Host Configuration Protocol (DHCP) test is used to test the DHCP service function of the peer device and the network accessibility and delay. When DHCP packets are used to test the network accessibility and delay, the RNS service must be enabled on the peer end.

After the detection time and frequency are set, the DHCP test process is as follows:

1. The client sends a DHCP discovery packet to the test destination IP address, and the destination server replies with an offer packet after receiving the DHCP discovery detection packet.
2. After receiving the offer packet, the client replies with a request packet. The destination server receives the request packet and replies with an ack packet.
3. After receiving the ack packet, the client deems that the detection succeeds, and sends a cancel packet to the server.

The detection client calculates the time required for the client to obtain the IP address and the packet loss rate according to the information of the received DHCP ack packet, reflecting the current network performance and network conditions. The DHCP test results and historical records are saved. You can run the **display ip rns configuration** command to view them.

1.1.5 TCP Test

The Transmission Control Protocol (TCP) test is used to test the availability of a TCP connection. A TCP connection can be established based on the configured destination IP address and port ID. If the TCP connection is established successfully, the test succeeds; otherwise, the test fails.

1.1.6 DNS Test

In a DNS test, a DNS client is simulated to send a domain name resolution request to a specified DNS server. You can determine whether the DNS server is available and the domain name resolution speed by checking the domain name resolution result and the time required for domain name resolution. In the DNS test, the domain name resolution process is simulated, and the mapping between the resolved domain name and the IP address is not saved. The DNS test results and historical records are saved in the test group. You can use the command line to view them.

The procedure for configuring an RNS instance test is as follows:

(1) Create an instance and configure the test based on the test type.

(2) Start the instance.

(3) Use the RNS instance to construct a packet of the specific test type and send the packet to the peer end.

(4) Upon receipt of the test packet, the peer end returns a reply packet of the corresponding type.

(5) The RNS instance calculates the packet loss rate and round trip time based on whether a reply packet is received and the time of reply packet receipt.

(6) Run the **display** or **debugging** command to view the test result.

1.1.7 Track Support for the RNS

Objects that can be tracked include: the test result of an RNS instance, RNS list status, link status on an interface, and track list status. When the track status changes, an action of other modules is triggered. The working process of track is described below by taking different track objects as examples.

- Configure a track object for tracking the test result of an RNS instance.
  - When the test result of the RNS instance changes, the RNS module sends a status change message to the track module.
  - The track module receives the test result. After the preset delay, if the test result remains unchanged, the status of the track object is modified, and the module of the track object is notified of the modification. If the test result recovers within the period, the status of the track object is not modified and the corresponding module is not notified.
- Configure a track object for tracking the link status of an interface.
  - If the link status of the interface is up, the status of the track object is also up.
  - If the link status of the interface is down, the status of the track object is also down.
- Configure a track object for tracking the test result of an RNS instance.
  - If the RNS test succeeds, the status of the track object is also up.
  - If the RNS test fails, the status of the track object is also down.
- Configure a track object for tracking the test result of an RNS list.
  - The result can be the AND or OR operation result of all member status.
  - If the result of this track object is set to the AND operation result of all member status, and the AND result of the status of all the track RNS objects is up, the status of this track object is up. So long as the RNS test result of one track RNS object is down, the status of this track object is also down.
  - If the result of this track object is set to the OR operation result of all member status, and the OR result of the status of all the track RNS objects is down, the status of this track object is down. So long as the RNS test result of one track RNS object is up, the status of this track object is also up.
- Configure a track object for tracking the status of a track list.
  - Set the result of the track object to the AND operation result of all member status.
  - When the AND operation result of all member status is up, the status of this track object is also up.
  - When the AND operation result of all member status is down the status of this track object is also down.
- Configure a track list member.

  Configure a track list member. The status of the member can be the same as or contrary to that of the corresponding track object.
- Adjust the delay for notifying the status change of a track object.
  - A longer delay indicates that it takes more time before the module that is concerned with the track object is informed of the status.
  - A shorter delay indicates that it takes less time before the module that is concerned with the track object is informed of the status.

1.2 Configuration Task Summary

RNS configuration includes the following tasks:

(1) Configuring Basic Features of RNS

(2) Configuring the ICMP Echo Test

(3) Configuring the DNS Test

(4) Configuring the TCP Connection Test

(5) Configuring the UDP Echo Test

(6) Configuring the DHCP Test

(7) Configuring the Track Support for the RNS

1.3 Configuring Basic Features of RNS

1.3.1 Overview

Configure details of an RNS test instance to complete basic configuration of the RNS instance. Brief configuration is also available. Configure and start an RNS instance at a time.

1.3.2 Restrictions and Guidelines

- In detailed configuration mode, if you do not configure the test type after entering the IP RNS mode by running the command, the RNS instance is not created.
- In detailed configuration mode, after configuring an RNS instance, you need to run the **ip rns schedule** command to configure the startup policy; otherwise, the test is not implemented.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Define an RNS operation object.

ip rns *operation-number* [{ **dns** *destination-hostname* **nameserver-config** *ipv4-address* | **icmp-echo** { *destination-ipv4-address* | *destination-ipv6-address* } | **tcp-connect** *destination-ipv4-address port-number* } [**frequency** *interval*] [**timeout** *interval*] [**threshold** *interval*]]

No track operation object is created by default.

(4) Configure the proactive threshold monitoring and triggering mechanism of the RNS test.

ip rns react-operation *operation-number* **react** { **allfail** | **rtt** | **timeout** } [**action track**] [**threshold-mode** { **average** [*number-of-measurements*] | **consecutive** [*occurrences*] | **immediate** | **never** | **xofy** [*x-value y-value*] }] [**threshold-value** *max-threshold min-threshold*]

The proactive threshold monitoring and triggering mechanism is not configured for the test by default.

(5) Enable an RNS instance to trigger another RNS instance.

ip rns trigger-operation *source-operation-number target-operation-number*

The trigger mechanism when the monitoring threshold exceeds the expectation is not configured for the test by default.

(6) Configure scheduling parameters of an RNS instance.

ip rns schedule *operation-number* [**lifetime** { **forever** | *period* }] [**start** { *hh:mm* [*:ss*] [*month day* | *day month*] | **pending** | **now** | **after** *hh*:*mm*:*ss* }] [**cycle**]

No scheduling policy is configured for the test by default.

(7) Configure the **ip rns restart** command to restart an RNS instance.

ip rns restart *operation-number*

To restart an IP RNS instance in pending status of scheduling, perform this configuration (or directly configure scheduling start: **ip rns schedule** *operation-number* **start now**).

The test restart function is not configured by default.

(8) Run the **ip rns reset** command to clear the RNS test configuration.

ip rns reset

Perform this configuration if you need to clear configurations of all the RNS instances, for example, when a lot of instances are configured but configuration errors are found.

The RNS test clearing function is not configured by default.

(9) Configure the **ip rns server** command to enable the server function.

ip rns-server udp-echo *port-number*

This configuration is mandatory when you need to enable the RNS server for UDP echo tests.

The server function of UDP echo test is not configured by default.

1.4 Configuring the ICMP Echo Test

1.4.1 Overview

Create an RNS test instance of ICMP echo type.

1.4.2 Restrictions and Guidelines

You must first configure the RNS basic features.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Create an RNS test instance of ICMP echo type.

icmp-echo oob { *destination-ipv4-address* | destination-ipv6-address | *destination-hostname* [**nameserver-config** { *ipv4-address* | *ipv6-address }*] } [**src-address** { *ipv4-address* | *ipv6-address* }] **via** *interface-name* **next-hop** { *ipv4-address* | *ipv6-address* }

icmp-echo { *destination-ipv4-address* | destination-*ipv6-address* | *destination-hostname* [**nameserver-config** *ipv4-address* | *ipv6-address*] } [**src-address** { *ipv4-address* | *ipv6-address* }] [[**af-direct**] **out-interface** *interface-name* [**next-hop** { *ipv4-address* | *ipv6-address* }]]

No test of ICMP echo type is created by default.

(5) (Optional) Configure the repeat test interval.

frequency *interval*

The default test interval is **60** seconds.

(6) (Optional) Configure the tag parameter.

tag *tag-name*

No test tag is configured by default.

(7) (Optional) Configure the time threshold parameter.

threshold *interval*

The default upper threshold of test is **5** seconds.

(8) (Optional) Configure the timeout period.

timeout *interval*

The default timeout interval of a test is **5** seconds.

(9) (Optional) Configure the type of service (TOS) parameter.

tos *tos-values*

The default TOS value for IPv4 packet of the test is **0**.

(10) (Optional) Configure the virtual routing and forwarding (VRF) of an RNS test.

vrf *vrf-name*

No VRF name is configured for the test by default.

(11) Configure the protocol payload size for the test.

request-data-size *bytes*

The default protocol payload size is **36** bytes for the tests of ICMP echo and UDP echo types.

1.5 Configuring the DNS Test

1.5.1 Overview

Create an RNS test instance of DNS type.

1.5.2 Restrictions and Guidelines

You must first configure the RNS basic features.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an RNS test instance of DNS type.

dns { **oob** *destination-hostname* **nameserver-config** *ipv4-address* [**src-address** *ipv4-address*] **via** *interface-type interface-number* **next-hop** *ipv4-address* } | { *destination-hostname* **nameserver-config** *ipv4-address* [**src-address** *ipv4-address*] [[**af-direct**] **out-interface** *interface-type interface-number* [**next-hop** *ipv4-address*]] }

dns *destination-hostname* **nameserver-config** *ipv4-address* [**src-address** *ipv4-address*] [[**af-direct**] **out-interface** *interface-type interface-number* [**next-hop** *ipv4-address*]]

No test of DNS type is created by default.

⚠ Caution

For the commands (**frequency, tag, threshold, timeout, and tos**) of configuring common optional parameters of a test, see 1.4 Configuring the ICMP Echo Test. They are omitted here.

1.6 Configuring the TCP Connection Test

1.6.1 Overview

Create an RNS test instance of TCP type to implement a TCP connection test.

1.6.2 Restrictions and Guidelines

- You must first configure the RNS basic features.
- The target host must be able to respond to the TCP connection request.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an RNS test instance of TCP type.

tcp-connect { *destination-ipv4-address* | *destination-hostname* [**nameserver-config** *ipv4-address*] } *port-number* [**source-ipaddr** *ipv4-address*] [[**af-direct**] **out-interface** *interface-type interface-number* **next-hop** *ipv4-address*]

No test of TCP type is configured by default.

⚠ Caution

For the commands (**frequency, tag, threshold, timeout, and tos**) of configuring common optional parameters of a test, see 1.4 Configuring the ICMP Echo Test. They are omitted here.

1.7 Configuring the UDP Echo Test

1.7.1 Overview

Create an RNS test instance of UDP echo type and implement the UDP echo test.

1.7.2 Restrictions and Guidelines

- You must first configure the basic features of the RNS and RNS server.
- The RNS server must be enabled to provide server functions.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an RNS test instance of UDP echo type.

udp-echo { **oob** *destination-ipv4-address port-number* [**via** *interface-type interface-number* **next-hop** *next-hop-ip*] | *destination-ipv4-address port-number* [**out-interface** *interface-type interface-number* **next-hop** *next-hop-ip*] }

udp-echo *destination-ipv4-address port-number* [**out-interface** *interface-type interface-number* **next-hop** *next-hop-ip*]

No test object of UDP echo type is configured by default.

⚠ Caution

For the commands (**frequency, tag, threshold, timeout, and tos**) of configuring common optional parameters of a test, see 1.4 Configuring the ICMP Echo Test. They are omitted here.

1.8 Configuring the DHCP Test

1.8.1 Overview

Create an RNS test instance of DHCP type to implement a DHCP test.

1.8.2 Restrictions and Guidelines

You must first configure the basic features of RNS and the IP address of the outbound port.

1.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an RNS test instance of DHCP type.

dhcp *destination-ipv4-address* **out-interface** *out-interface* **next-hop** *next_ip_address*

Caution

For the commands (**frequency, tag, threshold, timeout, and tos**) of configuring common optional parameters of a test, see 1.4 Configuring the ICMP Echo Test. They are omitted here.

1.9 Configuring the Track Support for the RNS

1.9.1 Overview

Configure track support for RNS, including the test results of track RNS, the link status of the interface, the status of track list, and the status of RNS test list.

1.9.2 Restrictions and Guidelines

- To configure the track function to track the test result of an RNS instance, you need to configure the related RNS instance.
- To configure the track function to track the link status of an interface, you need to configure the related interface.
- To configure the track function to track the status of a track list, you need to configure the members for the related track list.
- To configure the track function to track the status of an RNS list, you need to configure the members for the related RNS list.

1.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure the track object. Select at least one of the following to configure.

- Configure to track the link status of an interface.

track *object-number* **interface** *interface-type interface-number* **line-protocol**

- Configure to track the RNS test result.

track *object-number* **rns** *entry-number*

- Configure to track the status of a track list.

track *object-number* **list-object bool-calc** { **and** | **or** }

- Configure to track the status of an RNS list.

track *object-number* **rns-list** *men-list* { **and | or** }

(5) Configure the notification delay of a track object.

state-delay { **up** *interval* [**down** *interval*] **|** [**down** *interval*] **up** *interval* }

Configure the delay for notifying the tracking status change from up to down or from down to up, or configure both at the same time.

No delay is set by default for notifying the track object status change from up to down or from down to up.

(6) Configure a track member.

object *object-number* [**not**]

No member object is configured by default.

1.10 Monitoring

This section describes the **display** commands used for checking the running status of a configured function to verify the configuration effect.

Run the **debugging** command to display debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 RNS Monitoring

Command	Purpose
display ip rns configuration [*operation-number*]	Displays the configuration of one or more RNS instances.
display ip rns statistics-detail [*operation-number*]	Displays the detailed statistics of one or more RNS instances.
display ip rns state [*operation-number*]	Displays the current status of one or more RNS instances.
display ip rns-server udp-echo	Displays the RNS server configuration.
display ip rns react-operation [*operation-number*]	Displays the proactive threshold monitoring information of one or more RNS instances.
display ip rns trigger-operation [*operation-number*]	Displays information about the test triggered by one or more RNS instances.
display ip rns statistics [*operation-number*]	Displays the brief statistics of one or more RNS instances.
display track [*object-number*]	Displays the statistics of one or more track objects.
debugging track { **all	proc-event	rdnd-event	client** }	Debugs the track module.
debugging rns { **all	lib	rdnd-event	restart	rns_id** *rns_id* }	Debugs the RNS module.

Command	Purpose
debugging rns-server {udp-echo} { event | fail | packet }	Debugs the RNS server module.

1.11 Configuration Examples

1.11.1 Configuring the ICMP Echo Test

1. Requirements

Configure an RNS test of ICMP echo type, which must be started immediately and take effect permanently.

2. Topology

Figure 1-1 Configuring the ICMP Echo Test



3. Notes

Configure RNS instance 1 and related parameters on Device A.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip rns 1
DeviceA(cfg-ip-rns)# icmp-echo 10.2.2.2
DeviceA(cfg-ip-rns-icmp-echo)# exit
DeviceA(cfg)# ip rns schedule 1 start now lifetime forever
```

5. Verification

Run the **display ip rns configuration** command to display the instance configuration.

```
DeviceA# display ip rns configuration 1
Entry number: 1
Tag:
Type of operation to perform: icmp-echo
Operation timeout (milliseconds): 5000
Operation frequency (milliseconds): 60000
Threshold (milliseconds): 5000
Recurring (Starting Everyday): FALSE
Life (seconds): foerver
```

```
Next Scheduled Start Time:Start Time already passed
Target address/Source address: 10.2.2.2/0.0.0.0
Request size (ARR data portion): 36
```

1.11.2 Configuring the DNS Test

1. Requirements

Configure an RNS test of DNS type, which must be started immediately and take effect permanently.

2. Topology

Figure 1-2 Configuring the DNS Test



3. Notes

Configure RNS instance 1 and related parameters on Device A.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip rns 1
DeviceA(cfg-ip-rns)# dns www.abc.com nameserver-config 10.2.2.2
DeviceA(cfg-ip-rns-dns)# exit
DeviceA(cfg)# ip rns schedule 1 start now lifetime forever
```

5. Verification

Run the **display ip rns configuration** command to display the instance configuration.

```
DeviceA# display ip rns configuration 1
Entry number: 1
Tag:
Type of operation to perform: dns
Operation timeout (milliseconds): 5000
Operation frequency (milliseconds): 60000
Threshold (milliseconds): 5000
Recurring (Starting Everyday): FALSE
Life (seconds): foerver
Next Scheduled Start Time:Start Time already passed
Target host name: www.abc.com
Name Server: 10.2.2.2
```

6. Common Errors

The DNS IP address is incorrect.

1.11.3 Configuring the TCP Connect Test

1. Requirements

Configure an RNS test of TCP type, which must be started immediately and take effect permanently.

2. Topology

Figure 1-3 Configuring the TCP Connect Test



3. Notes

Configure RNS instance 1 and related parameters on Device A.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip rns 1
DeviceA(cfg-ip-rns)# tcp-connect 10.2.2.2 8000
DeviceA(cfg-ip-rns-tcp)# exit
DeviceA(cfg)# ip rns schedule 1 start now lifetime forever
```

5. Verification

Run the **display ip rns configuration** command to display the instance configuration.

```
DeviceA# display ip rns configuration 1
Entry number: 1
Tag:
Type of operation to perform: tcp-connect
Operation timeout (milliseconds): 5000
Operation frequency (millseconds): 60
Threshold (milliseconds): 5000
Recurring (Starting Everyday): FALSE
Life (seconds): foerver
Next Scheduled Start Time:Start Time already passed
Target Address: 10.2.2.2
Target Port: 8000
```

6. Common Errors

- The target host does not respond to the TCP connection request.
- The configured TCP port for the RNS test is incorrect.

1.11.4 Configuring the UDP Echo Test

1. Requirements

Configure an RNS test of UDP echo type, which must be started immediately and take effect permanently.

2. Topology

Figure 1-4 Configuring the UDP Echo Test



3. Notes

Configure RNS instance 1 and related parameters on Device A.

4. Procedure

Perform the following configuration Device A:

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip rns 1
DeviceA(cfg-ip-rns)# udp-echo 10.2.2.2 1025
DeviceA(cfg-ip-rns-udp-echo)# exit
DeviceA(cfg)# ip rns schedule 1 start now lifetime forever
```

Perform the following configuration on Device B:

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ip rns-server udp-echo 1025
```

5. Verification

Run the **display ip rns configuration** command to display the instance configuration.

```
DeviceA# display ip rns configuration 1
Entry number: 1
Tag:
Type of operation to perform: udp-echo
Operation timeout (milliseconds): 5000
Operation frequency (milliseconds): 60000
Threshold (milliseconds): 5000
Target Address: 10.2.2.2
Target Port: 1025
Request size (ARR data portion): 36
```

Run the **display ip rns-server udp-echo** command to display the RNS server configuration.

```
DeviceA# display ip rns-server udp-echo
```

```
UDP-Echo-Server:10000
Receive packets number: 0
Reflect packets success number: 0
Reflect packets fail number: 0
```

6. Common Errors

The UDP server port configured on the client is not correct.

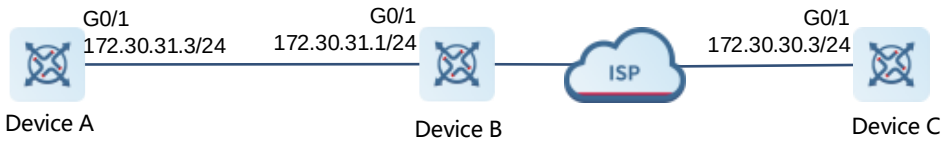
1.11.5 Configuring the DHCP Test

1. Requirements

Configure an RNS test of DHCP type, which must be started immediately and take effect permanently.

2. Topology

Figure 1-5 Configuring the DHCP Test



3. Notes

Configure RNS instance 1 and related parameters on Device A.

4. Procedure

Perform the following configuration on Switch A:

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip rns 1
DeviceA(cfg-ip-rns)# dhcp 172.30.30.3 out-interface GigabitEthernet 1/0/1 next-hop 172.30.31.1
DeviceA(cfg-ip-rns-dhcp)# exit
DeviceA(cfg)# ip rns schedule 1 start now lifetime forever
```

5. Verification

Run the **display ip rns configuration** command to display the instance configuration.

```
DeviceA# display ip rns configuration 1
Entry number: 1
Tag:
Type of operation to perform: dhcp
Operation timeout (milliseconds): 5000
Operation frequency (milliseconds): 60000
Threshold (milliseconds): 5000
Recurring (Starting Everyday): FALSE
Life (seconds): forever
Next Scheduled Start Time: Start Time already passed
```

```
Target address/Source address: 172.30.30.3/172.30.31.252
Out-interface: GigabitEthernet 1/0/1
Next-hop address: 172.30.31.1
```

6. Common Errors

- The DHCP service is not enabled on the server.
- No IP address is configured for the outbound interface of the client.

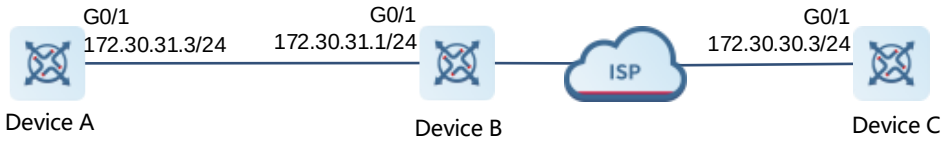
1.11.6 Configuring the Device to Track Link Status of an Interface

1. Requirements

Configure track object 3 to track the link status of the interface GigabitEthernet 0/1.

2. Topology

Figure 1-6 Configuring the Device to Track Link Status of an Interface



3. Notes

- Configure a track object to track the link status of an interface.
- Configure the delay for notifying the status change from up to down.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# track 3 interface gigabitethernet 0/1 line-protocol
DeviceA (cfg-track)# state-delay down 10
```

5. Verification

Disable the interface GigabitEthernet 0/1, namely, change the link status to down. Immediately check the status of the track object, and verify that the status is still up. Check the status of the track object 10s later, and verify that the status changes to down.

```
DeviceA# display track 3
Track 3
  Interface GigabitEthernet 0/1
  The state is Up, delayed Down (5 secs remaining)
    1 change, current state last: 300 secs
  Delay up 0 secs, down 10 secs
```

6. Configuration Files

```
track 3 Interface GigabitEthernet 0/1 line-protocol
 state-delay down 10
!
```

7. Common Errors

- The track object for tracking an RNS test is configured, but the RNS test is not configured.
- The track object for tracking the link status of an interface is configured, but the corresponding interface is not configured.
- The track object for tracking the status of a track list is configured, but no member of the RNS list is configured.
- The track object configured for tracking an RNS list is configured, but the RNS test is not configured.

1.11.7 Configuring the Device to Track the Status of Multiple Interfaces

1. Requirements

Configure track object 3. When track object 1 is up and object 2 is down, track object 3 is up.

2. Topology

Figure 1-7 Configuring the Device to Track the Status of Multiple Interfaces



3. Notes

- Configure track objects 1 and 2.
- Configure track object 3, and its members include track objects 1 and 2.

4. Procedure

Perform the following configuration on Device A:

```
DeviceA# config
DeviceA(cfg)# track 1 interface gigabitEthernet 0/0 line-protocol
DeviceA(cfg-track)# state-delay up 20 down 40
DeviceA(cfg-track)# exit
DeviceA(cfg)# track 2 interface gigabitEthernet 0/1 line-protocol
DeviceA(cfg-track)# state-delay down 30
DeviceA(cfg-track)# exit
DeviceA(cfg)# track 3 list Boolean and
DeviceA(cfg-track)# object 1
DeviceA(cfg-track)# object 2 not
```

5. Verification

```
DeviceA# display track 3
Track 3
   List boolean and
   Object 1
   Object 2 not
   The state is Down
      1 change,current state last:10 secs
   Delay up 0 secs,down 0 secs
```

6. Configuration Files

```
track 1 Interface GigabitEthernet 0/0 line-protocol
 state-delay up 20 down 40
!
track 2 Interface GigabitEthernet 0/1 line-protocol
 state-delay down 30
!
track 3 list boolean and
 object 1
 object 2 not
!
```

1.11.8 Configuring the Device to Track A Single RNS Instance

1. Requirements

Configure track object 5 to track the test result of RNS instance 7.

2. Notes

- Configure an RNS test.
- Configure a track object for tracking the test result of an RNS instance.
- Configure the delay for notifying the test result change from up to down and the delay for notifying the test result change from down to up.

3. Procedure

Perform the following configuration on Device A:

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ip rns 7
DeviceA(cfg-ip-rns)# icmp-echo 2.2.2.2
DeviceA(cfg-ip-rns-icmp-echo)# exit
DeviceA(cfg)# ip rns schedule 7 start now lifetime forever
DeviceA(cfg)# track 5 rns 7
DeviceA(cfg-track)# state-delay up 20 down 30
```

4. Verification

Make the test result of RNS instance 7 change from successful to unsuccessful. When the test result changes to unsuccessful, immediately check the status of track object 7, and verify that the status is still up. Check the status of the track object 30s later, and verify that the status changes to down.

```
DeviceA# display track 5
Track 5
  Reliable Network Service 7
  The state is Down
    2 change, current state last: 10 secs
  Delay up 20 secs, down 30 secs
```

5. Configuration Files

```
ip rns 7
 icmp-echo 2.2.2.2
!
ip rns schedule 7 start now lifetime forever
!
track 5 rns 7
 state-delay up 20 down 30
!
```

1.11.9 Configuring the Device to Track the Results of Multiple RNS Instances

1. Requirements

Configure track object 5 for tracking the test results of an RNS list (consisting of RNS instances 1, 2-5, and 8).

2. Notes

- Configure and start an RNS test.
- Configure a track object for tracking the test result of an RNS list.
- Configure the delay for notifying the test result change from up to down, and the delay for notifying the test result change from down to up.

3. Procedure

Perform the following configuration on Device A:

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# track 5 rns-list 1,2-5,8 and
DeviceA(cfg-track)# state-delay up 20 down 30
```

4. Verification

Make the test result of one of the RNS instances 1, 2-5, and 8 change from successful to unsuccessful. When the test result changes to unsuccessful, immediately check the status of track object 7, and verify that the status is still up. Check the status of the track object 30s later, and verify that the status changes to down.

```
DeviceA# display track 5
Track 5
```

```
   rns-list 1,2-5,8 and
   The state is Down
      2 change, current state last: 10 secs
Delay up 20 secs, down 30 secs
```

5. Configuration Files

```
track 5 rns-list 1,2-5,8 and
  state-delay up 20 down 30
!
```

Contents

1 Configuring IP Event Dampening....1
1.1 Introduction....1
1.1.1 Basic Concepts....1
1.1.2 Protocols and Standards....2
1.2 Restrictions and Guidelines....2
1.3 Configuring IP Event Dampening....2
1.3.1 Overview....2
1.3.2 Restrictions and Guidelines....2
1.3.3 Procedure....2
1.4 Monitoring....3
1.5 Configuration Examples....3
1.5.1 Configuring IP Event Dampening on Layer 3 Interfaces....3

1 Configuring IP Event Dampening

1.1 Introduction

When a Layer 3 network interface frequently goes Up or Down due to manual enabling/disabling or other external causes, the routing table on the device flaps repeatedly. If a routing protocol is configured, the protocol may extend the flap to the entire network, causing repeated updates and recalculation of neighboring routes. As a result, this wastes network bandwidth and destabilizes the network. Repeated route updates and recalculation on devices consume many CPU resources, and thus affect the normal running of customer networks.

The function of IP Event Dampening detects abnormal Up/Down flapping and automatically suppresses frequent interface state changes, which prevents a routing protocol from propagating single-point link failures. When an interface is restored to stable, it is automatically released, thus reducing network flaps and CPU resource consumption while improving network stability.

1.1.1 Basic Concepts

1. Penalty

IP Event Dampening uses the penalty method to perceive and control interface changes. Whenever an interface goes Up or Down, the penalty accumulates, but when the interface is stable, the penalty decreases exponentially. In this way, interface behaviors can be sensed and processed intelligently.

2. Suppress threshold

To identify an interface flap event, the function of IP Event Dampening configures a suppress threshold. When the cumulative penalty of an interface exceeds a suppress threshold, the interface is considered to flap and is suppressed.

3. Half-life period

The half-life period is the period for the penalty to decrease to half of the original value when an interface becomes stable. It defines the speed at which the penalty decays exponentially. The shorter the half-life period, the faster the penalty decays, and the faster the interface is detected to be stable, but the flap detection sensitivity is reduced more greatly.

4. Reuse threshold

When an interface does not flap and its penalty decays to a certain degree (below the suppress threshold), the interface is considered to be stable and is released.

5. Maximum suppress time

When an interface keeps flapping and reaches a very large penalty, the interface is not usable for a long time. To avoid this problem, we define the maximum suppress time to ensure that the interface suppression duration is less than the maximum suppress time no matter how long the interface has flapped.

6. Interface flap suppression

- An interface configured with IP Event Dampening is assigned a penalty. The interface gets a penalty of 1000 whenever it goes Down, and the penalty decreases with time. If the interface goes Down again, the penalty

increases. When the cumulative penalty exceeds the suppress threshold, the interface is suppressed. For the affected upper-layer protocol, the suppressed interface is always Down regardless of the interface state. When the penalty decreases to the reuse threshold, the interface is released and restored to its real state.

- If a Layer 3 interface is not configured with IP Event Dampening or is not suppressed, the routing protocol or other protocols concerned about the interface state still work normally. When the interface is configured with IP Event Dampening and is suppressed, the upper-layer protocol considers the interface to be Down. During this process, any state change of the interface does not affect the routing table or the route calculation and advertisement performed by the upper-layer routing protocol until the interface is released.

1.1.2 Protocols and Standards

- RFC 2439: BGP Route Flap Damping

Note

The suppression algorithm core used by IP Event Dampening is the same as that used by Border Gateway Protocol (BGP) Route Flap Dampening stipulated in RFC 2439.

1.2 Restrictions and Guidelines

- The function of IP Event Dampening is supported only on Layer 3 interfaces.

1.3 Configuring IP Event Dampening

1.3.1 Overview

When an interface configured with IP Event Dampening keeps flapping until the predefined threshold is exceeded, the interface is set to Down.

1.3.2 Restrictions and Guidelines

- When a Layer 3 interface on a device is converted into a Layer 2 interface, the IP Event Dampening command configured on the interface is deleted.
- IP Event Dampening can be configured only on the primary interface of a router but takes effect to all subinterfaces related to the primary interface. IP Event Dampening cannot be configured on **virtual templates**.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 aggregate interface configuration mode.

interface aggregateport *interface-number*

(4) Configure IP Event Dampening.

dampening [*half-life-period* [*reuse-threshold suppress-threshold max-suppress* [**restart** [*restart-penalty*]]]]

IP event dampening is disabled on an interface by default.

1.4 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

Run the **debugging** commands to output debugging information.

⚠ Caution

- Running the **cleanup** command during operation of the device may lose vital information and interrupt services.
- The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 Monitoring

Command	Purpose
cleanup counters	Clears interface statistics.
display dampened-interface interface	Displays statistics about interfaces with IP Event Dampening.
display interfaces [*interface-type interface-number*] **dampened-interface**	Displays IP Event Dampening configurations.
debugging dampening interface	Debugs IP Event Dampening.

1.5 Configuration Examples

1.5.1 Configuring IP Event Dampening on Layer 3 Interfaces

1. Requirements

Configure IP Event Dampening on Layer 3 interfaces.

2. Topology

Figure 1-1 Configuring IP Event Dampening on Layer 3 interfaces



3. Notes

- Enable IP event dampening on interface GigabitEthernet 0/1 on Routers A and B respectively, and set the half-time period to 30s, the reuse threshold to 1500, the suppress threshold to 10,000, and the maximum suppress time to 120s.

4. Procedure

Configure IP event dampening on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# summary gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# dampened-interface 30 1500 10000 100
```

Configure IP event dampening on Device B.

```
DeviceB# config-system terminal
DeviceB(cfg)# summary gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# dampened-interface 30 1500 10000 100
```

5. Verification

Run the **display interfaces dampened-interface** command to display the detailed information about Event Dampening on an interface.

```
DeviceA# display interfaces dampened-interface
GigabitEthernet 0/1
  Flaps Penalty Supp      ReuseTm HalfL    ReuseV   SuppV    MaxSTm   MaxP      Restart
  0     0       FALSE     0       30       1500     1000     100      15119     0
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
summary GigabitEthernet 0/1
 unset switch-port
 dampened-interface 30 1500 10000 100
!
end
```

- Device B configuration file

```
hostname DeviceA
!
summary GigabitEthernet 0/1
```

```
 unset switch-port
 dampened-interface 30 1500 10000 100
!
end
```

7. Common Errors

- IP Event Dampening is configured for an interface on a Layer 3 device before the interface is converted to a Layer 3 interface by running the **unset swithport** command.

Contents

1 Configuring HAM 1

1.1 Introduction 1

1.2 HAM Application 1

1.2.1 Detection Process 1

1.3 Configuring the HA Service of One Frequently Restarting Process 1

1 Configuring HAM

1.1 Introduction

High availability module (HAM) is a self-developed system high availability (HA) technology. It monitors exceptional exit of software processes of a device to automatically start the processes or monitors frequent restart of a software process to automatically restart the device, thus quickly and automatically restore services. Used during patch upgrade, this function can prevent device reset due to patch upgrade exceptions.

1.2 HAM Application

HAM is mainly used for:

- Software process HA service: When a software process exits abnormally, HAM automatically starts the software process to restore services.
- Device restart service: When a software process restarts frequently, HAM attempts to automatically restart the device to restore services.

Typically, the HA service of one frequently restarting process is enabled on all devices. You can run command line interface (CLI) commands on the master device to enable/disable this service and synchronize the service status to other non-master devices. During a patch upgrade, a process may restart due to an abnormal process. If the device automatically restarts, traffic may be interrupted. Therefore, you must disable the HA service of one frequently restarting process. After a patch upgrade, you can enable the HA service of one frequently restarting process when confirming that no process restarts frequently.

⚠ Caution

If you disable the HA service of one frequently restarting process by running a command, services may not automatically be restored to normal because some processes frequently restart due to no automatic restarting of the device. As a result, failure time prolongs.

1.2.1 Detection Process

HAM monitors kernel advertisements to perceive the status of an abnormal process and automatically restores services based on the following rules:

- If one process exits five times within two minutes, HAM attempts to start the process five minutes later.
- If a process restarts twice within 10 minutes, services cannot be restored. In this case, you must restart the device to restore them.

1.3 Configuring the HA Service of One Frequently Restarting Process

1. Overview

When one process exits abnormally, you can automatically restore services by configuring the HA service of one process that restarts frequently. You can further determine the exception to check whether the device needs to be restarted to restore services.

2. Restrictions and Guidelines

If one process frequently restarts, disabling its HA service may lead to a longer failure time. Therefore, you are advised to enable the configuration in scenarios other than patch upgrades.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Disable the HA service of one frequently restarting process.

unset ham single-process-ha auto-restart enable

The HA service of one frequently restarting process is enabled by default.

Contents

1 Configuring DAG 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 2
1.3 Configuring M-LAG DAG 2
1.3.1 Overview 2
1.3.2 Restrictions and Guidelines 2
1.3.3 Procedure 2
1.4 Monitoring 2
1.5 Configuration Examples 3
1.5.1 Configuring the M-LAG DAG 3

1 Configuring DAG

1.1 Introduction

1.1.1 Overview

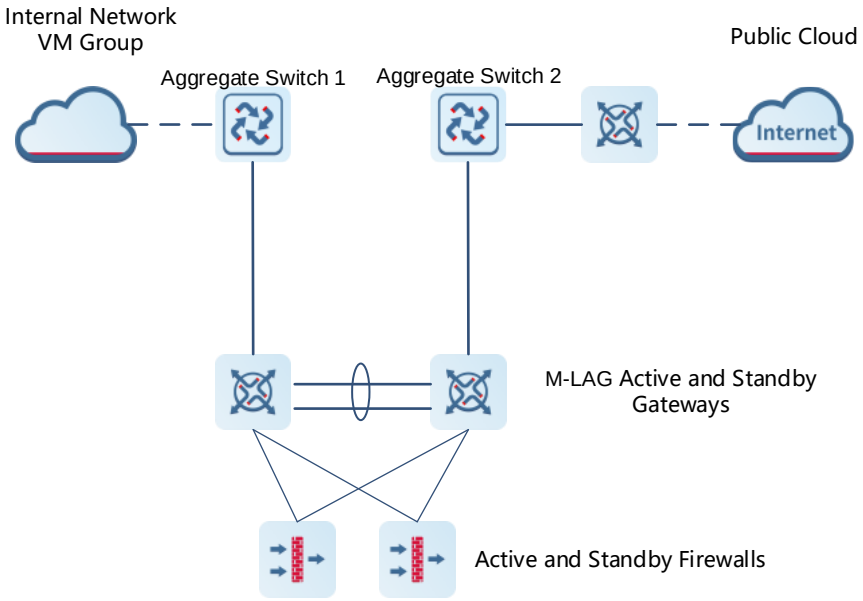
Generally, the same default route is configured for all hosts in a network. The default route points to the egress gateway to allow the hosts to communicate with the external network. When errors occur in the egress gateway, the communication between hosts and the external network is interrupted.

In a dual-active gateway (DAG) mode, multiple gateways (typically two) are combined to form a virtual router. The IP address of the virtual router is used as the default gateway address to communicate with the external network. When one gateway fails, the communication between hosts and the external network is not interrupted. This improves the reliability of the entire network. When all devices run properly, the two gateways can forward traffic at the same time, which helps fully utilize the network bandwidth.

1.1.2 Principles

One common DAG mode is that, in multi-chassis link aggregation group (M-LAG) scenarios, M-LAG active and standby devices are used as a DAG, which is connected to downlink servers through downlink Layer 3 aggregate ports. This DAG mode is called M-LAG DAG.

Figure 1-1 M-LAG DAG in a Public Cloud Scenario



- M-LAG DAG

  In a public cloud scenario, virtual machine (VM) traffic flows through the M-LAG active and standby gateways to the active and standby firewalls. In the reverse direction, the firewalls send the traffic to the public network through the M-LAG active and standby gateways (Figure 1-1).

  M-LAG active and standby gateways connect to the downlink active and standby firewalls through Layer 3 aggregate ports in dual-homing mode. The gateways share the same IP address and MAC address and the

DAG mode is enabled. A static VXLAN tunnel is established between the M-LAG active and standby gateways and reverse path limited (RPL) is disabled. Floating static routes are created in a VRF instance, with one pointing to the firewalls and the other pointing to the peer end of the static VXLAN tunnel. Different administrative distances are configured to ensure that the route pointing to the firewalls has a higher priority. When the DAG runs properly, both gateways forward traffic. When the downlink of one gateway fails, traffic of the faulty gateway is switched to the static VXLAN tunnel (the traffic flows through the peer-link of the M-LAG) to ensure service continuity.

1.2 Configuration Task Summary

Configuring the DAG includes the following tasks:

- Configuring M-LAG DAG

1.3 Configuring M-LAG DAG

- Overview

You can configure the DAG mode on the downlink Layer 3 aggregate ports and use the existing floating static routing function to implement M-LAG DAG.

1.3.1 Restrictions and Guidelines

M-LAG active and standby devices connect to the active and standby firewalls as gateways to ensure the firewall stability. In addition, sub-interfaces of the Layer 3 aggregate ports can be configured to differentiate traffic and expand the capacity of the firewalls to process traffic.

- In normal cases, traffic from the M-LAG gateways is sent to the active firewall. DAG entries are synchronized through the peer-link of the M-LAG.
- When the active firewall fails, the traffic from the M-LAG gateways is switched to the standby firewall.
- When the downlink of one M-LAG gateway fails, the faulty M-LAG gateway sends traffic to the firewall through the peer-link between M-LAG gateways in the static VXLAN tunnel.

1.3.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 3 aggregate port configuration mode.

interface aggregateport *interface-number*

(4) Enable the DAG mode.

mlag dual-active-gateway enable

The DAG mode is disabled by default.

1.4 Monitoring

Run the **debugging** command to output debugging information.

Caution

System resources are occupied when debugging information is output. Therefore, disable the debugging switch immediately after use.

Table 1-1 Monitoring

Command	Purpose
debugging net mlag-dag	Debugs the M-LAG DAG mode.

1.5 Configuration Examples

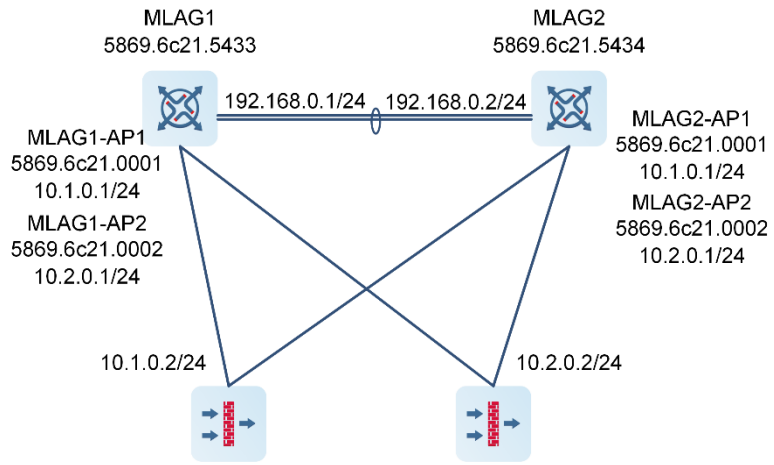
1.5.1 Configuring the M-LAG DAG

1. Requirements

The ARP entries of the DAG can be synchronized, the DAG can respond to the ARP requests of the firewalls as a proxy, and the ARP broadcast requests are not flooded.

2. Topology

Figure 1-2 DAG Configuration Topology



3. Note

- Enable the DAG mode on the downlink Layer 3 aggregate ports.
- Configure the system ID to be used for LACP negotiation between the Layer 3 aggregate ports and downlink firewalls to achieve the dual-homing mode.
- Create a static VXLAN tunnel and two floating static routes.

4. Procedure

(1) On the downlink aggregate ports of M-LAG1, configure the DAG mode, MAC address, IP address, and system ID for LACP negotiation.

```
MLAG1> privilege
MLAG1# config-system terminal
MLAG1(cfg)# interface AggregatePort 1
MLAG1(cfg-if-AggregatePort 1)# mlag dual-active-gateway enable
MLAG1(cfg-if-AggregatePort 1)# mac-address 5869.6c21.0001
MLAG1(cfg-if-AggregatePort 1)# ip address 10.1.0.1/24
MLAG1(cfg-if-AggregatePort 1)# lacp system-id 5869.6c21.54333
MLAG1(cfg-if-AggregatePort 1)# exit
MLAG1(cfg)# interface AggregatePort 2
MLAG1(cfg-if-AggregatePort 2)# mlag dual-active-gateway enable
MLAG1(cfg-if-AggregatePort 2)# mac-address 5869.6c21.0002
MLAG1(cfg-if-AggregatePort 2)# ip address 10.2.0.1/24
MLAG1(cfg-if-AggregatePort 2)# lacp system-id 5869.6c21.54333
MLAG1(cfg-if-AggregatePort 2)# exit
```

(2) On M-LAG1, create a static VXLAN tunnel.

```
MLAG1(cfg)# interface overlaytunnel 1
MLAG1(cfg-if-OverlayTunnel 1)# tunnel source 192.168.0.1
MLAG1(cfg-if-OverlayTunnel 1)# tunnel destination 192.168.0.2
MLAG1(cfg-if-OverlayTunnel 1)# exit
MLAG1(cfg)# interface OverlayRouter 1
MLAG1(cfg-if-OverlayRouter 1)# anycast-gateway
MLAG1(cfg-if-OverlayRouter 1)# exit
MLAG1(cfg)# vxlan 1
MLAG1(cfg-vxlan)# symmetric
MLAG1(cfg-vxlan)# router-interface OverlayRouter 1
MLAG1(cfg-vxlan)# tunnel-interface overlayTunnel 1
MLAG1(cfg-vxlan)# exit
```

(3) On M-LAG1, create a floating static route.

```
MLAG1(cfg)# ip route-config 10.1.0.2 255.255.255.0 aggregatePort 1 10
MLAG1(cfg)# ip route-config 10.2.0.2 255.255.255.0 aggregatePort 2 10
MLAG1(cfg)# ip route-config 10.1.0.2 255.255.255.0 OverlayRouter 1 192.168.0.2 20 router-mac
5869.6c21.5434
MLAG1(cfg)# ip route-config 10.2.0.2 255.255.255.0 OverlayRouter 1 192.168.0.2 20 router-mac
5869.6c21.5434
```

(4) On the downlink aggregate ports of M-LAG2, configure the DAG mode, MAC address, IP address, and system ID for LACP negotiation.

```
MLAG2> privilege
MLA2# config-system terminal
MLAG2(cfg)# interface AggregatePort 1
MLAG2(cfg-if-AggregatePort 1)# mlag dual-active-gateway enable
MLAG2(cfg-if-AggregatePort 1)# mac-address 5869.6c21.0001
MLAG2(cfg-if-AggregatePort 1)# ip address 10.1.0.1/24
MLAG2(cfg-if-AggregatePort 1)# lacp system-id 5869.6c21.54333
MLAG2(cfg-if-AggregatePort 1)# exit
MLAG2(cfg)# interface AggregatePort 2
```

```
MLAG2(cfg-if-AggregatePort 2)# mlag dual-active-gateway enable
MLAG2(cfg-if-AggregatePort 2)# mac-address 5869.6c21.0002
MLAG2(cfg-if-AggregatePort 2)# ip address 10.2.0.1/24
MLAG2(cfg-if-AggregatePort 2)# lacp system-id 5869.6c21.54333
MLAG2(cfg-if-AggregatePort 2)# exit
```

(5) On M-LAG2, create a static VXLAN tunnel.

```
MLAG2(cfg)# interface overlaytunnel 1
MLAG2(cfg-if-OverlayTunnel 1)# tunnel source 192.168.0.2
MLAG2(cfg-if-OverlayTunnel 1)# tunnel destination 192.168.0.1
MLAG2(cfg-if-OverlayTunnel 1)# exit
MLAG2(cfg)# interface OverlayRouter 1
MLAG2(cfg-if-OverlayRouter 1)# anycast-gateway
MLAG2(cfg-if-OverlayRouter 1)# exit
MLAG2(cfg)# vxlan 1
MLAG2(cfg-vxlan)# symmetric
MLAG2(cfg-vxlan)# router-interface OverlayRouter 1
MLAG2(cfg-vxlan)# tunnel-interface overlayTunnel 1
MLAG2(cfg-vxlan)# exit
```

(6) On M-LAG2, create a floating static route.

```
MLAG2(cfg)#ip route-config 10.1.0.2 255.255.255.0 aggregatePort 1 10
MLAG2(cfg)#ip route-config 10.2.0.2 255.255.255.0 aggregatePort 2 10
MLAG2(cfg)# ip route-config 10.1.0.2 255.255.255.0 OverlayRouter 1 192.168.0.2 20 router-mac
5869.6c21.5434
MLAG2(cfg)# ip route-config 10.2.0.2 255.255.255.0 OverlayRouter 1 192.168.0.2 20 router-mac
5869.6c21.5434
```

1. Verification

Run the **display current-config** command to display the configuration.

Network Control and Monitoring Configuration

1. Configuring Network Connectivity Test
2. Configuring Tech-Support
3. Configuring SPAN
4. Configuring Packet Capture
5. Configuring sFlow
6. Configuring IPFIX
7. Configuring the Function of Displaying Device Restart Reasons
8. Configuring NTP
9. Configuring SNTP
10. Configuring FTP Server
11. Configuring FTP Client
12. Configuring TFTP Server
13. Configuring TFTP Client
14. Configuring SNMP
15. Configuring RMON
16. Configuring NETCONF
17. Configuring gRPC
18. Configuring OpenFlow
19. Configuring SEM

20. Configuring IDS
21. Configuring Intelligent Monitoring
22. Configuring HDF
23. Configuring Hotkeys
24. Configuring Network Packet Loss Detection
25. Configuring IFA

Contents

1 Configuring Network Connectivity Test 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.1.3 Protocols and Standards 2

1.2 Restrictions and Guidelines 2

1.3 Configuration Task Summary 2

1.4 Configuring a Ping Connectivity Test 3

1.4.1 Overview 3

1.4.2 Procedure 3

1.5 Configuring a Traceroute Connectivity Test 3

1.5.1 Overview 3

1.5.2 Procedure 3

1.6 Monitoring 4

1.7 Configuration Examples 4

1.7.1 Configuring a Ping Connectivity Test 4

1.7.2 Configuring a Traceroute Connectivity Test 6

1 Configuring Network Connectivity Test

1.1 Introduction

1.1.1 Overview

The network connectivity test tool is used to check whether the network can be connected, and accurately find out the cause of the network failure. Common network communication test tools include packet Internet groper (Ping) and Traceroute. On a network device, the commands corresponding to these two tools are **ping** and **traceroute**.

- The Ping tool is often used to check the network connectivity between devices and the network delay. A larger delay indicates a slower network speed.
- The Traceroute tool helps users learn the topology of the physical and logical connections of the network and efficiency of data transmission.

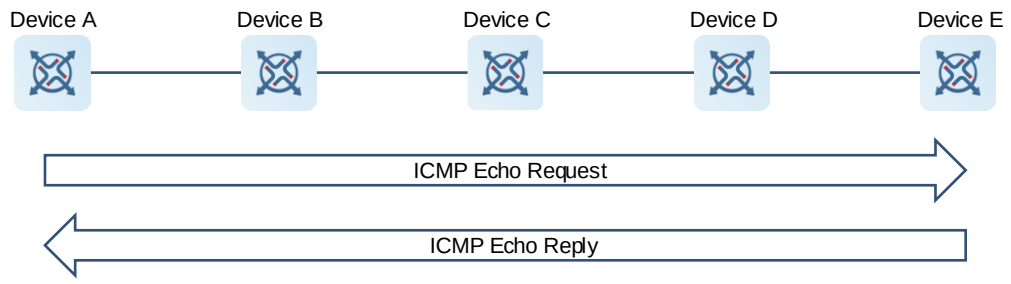
1.1.2 Principles

1. Testing connectivity with Ping

The Ping tool sends an Internet Control Message Protocol (ICMP) Echo Request packet to the destination IP address, requesting the peer to return an ICMP Echo Reply packet to determine whether the two network devices are connected and what is the delay.

The sequence number of the ICMP Echo Request starts from 0. Each time an ICMP Echo Request is sent by the Ping tool, the sequence number increases by 1. The sender sends five ICMP Echo Request packets by default. You can also specify the number of sent ICMP Echo Request packets in the command. If two devices are connected, the receiver returns the same number of ICMP Echo Reply packets.

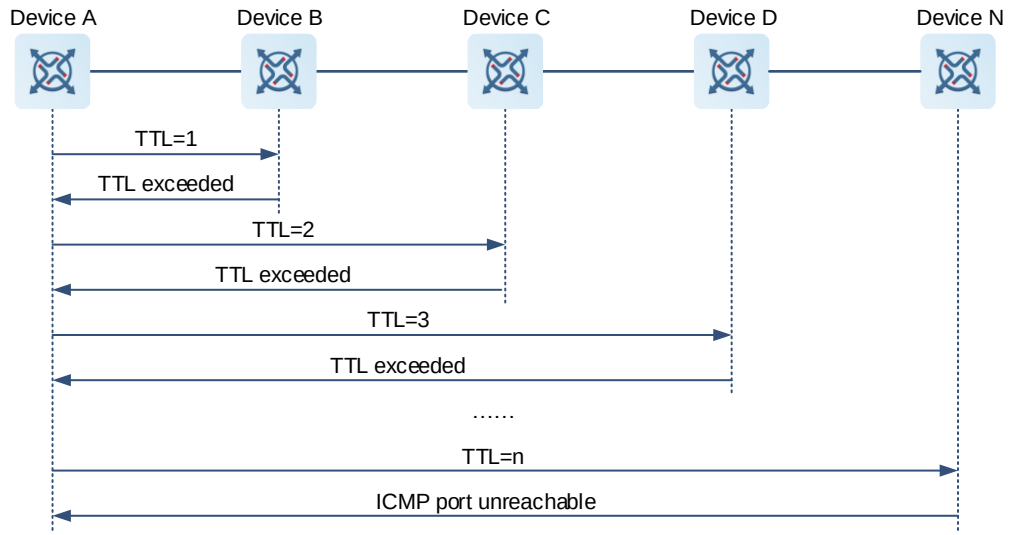
Figure 1-1 Testing Connectivity with Ping



2. Testing connectivity with Traceroute

The Traceroute tool uses the Time To Live (TTL) field in the header of ICMP and IP packets to complete the connectivity check. The process is as follows:

Figure 1-2 Testing Connectivity with Traceroute



(1) Device A sends an ICMP Request packet with the TTL set to 1 through the Traceroute tool. The UDP port number in the packet is an invalid destination port number (greater than 3000).

(2) When the first device, device B, on the path receives this packet, device B decreases the TTL by 1. Then, the TTL changes to 0, and the device discards the packet and returns an ICMP Time Exceeded message.

(3) When Traceroute on device A receives this message, it learns that device B exists on the path and therefore sends a packet with the TTL set to 2 to detect the second device, device C.

(4) Each time when device A sends a packet, Traceroute on device A increases the TTL by 1 to detect another device. This operation repeats until a packet arrives at the destination host. When the packet arrives at the destination host, the host returns an ICMP Port Unreachable message, instead of an ICMP Time Exceeded message. Finally, Traceroute stops the detection and displays the path from the network device to the destination host.

1.1.3 Protocols and Standards

- RFC 792: Internet Control Message Protocol
- RFC 4443: Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification

1.2 Restrictions and Guidelines

An IP address must be configured for the network device where the Ping and Traceroute operations are executed.

1.3 Configuration Task Summary

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring a Ping Connectivity Test
- Configuring a Traceroute Connectivity Test

1.4 Configuring a Ping Connectivity Test

1.4.1 Overview

You can use Ping connectivity test to check whether a network device is connected to the destination host and whether packets can be transmitted between these two devices.

1.4.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Check whether the IP address is reachable.

(IPv4 network)

ping [**ip** | **vrf** *vrf-name*] { *hostname* | *ipv4-address* } [**8021p** *8021p* | **data** *data* | **detail** | **df-bit** | **dscp** *dscp* | **interval** *interval* | **length** *length* | **ntimes** *times* | **out-interface** *interface-type interface-number* [**next-hop** *next-hop*] | [**source** *interface-type interface-number* | **source** *source-ipv4-address*] | **timeout** *time* | **validate**] *

ping oob { *hostname* | *ipv4-address* } [**data** *data* | **detail** | **df-bit** | **interval** *interval* | **length** *length* | **ntimes** *times* | **out-interface** *interface-type interface-number* [**next-hop** *next-hop*] | [**source** *interface-type interface-number* | **source** *source-ipv4-address*] | **timeout** *time* | **validate** | **via** *mgmt-name*] *

(IPv6 network)

ping [**ipv6** | **vrf** *vrf-name*] { *hostname* | ipv6-*address* } [**8021p** *8021p* | **data** *data* | **detail** | **dscp** *dscp* | **interval** *interval* | **length** *length* | **ntimes** *times* | **out-interface** *interface-type interface-number* [**next-hop** *next-hop*] | [**source** *interface-type interface-number* | **source** *source-ipv6-address*] | **timeout** *time*] *

ping oob { *hostname* | *ipv6-address* } [**data** *data* | **detail** | **interval** *interval* | **length** *length* | **ntimes** *times* | **out-interface** *interface-type interface-number* [**next-hop** *next-hop*] | [**source** *interface-type interface-number* | **source** *source-ipv6-address*] | **timeout** *time* | **validate** | **via** *mgmt-name*] *

By default, five 100-byte packets are sent to the specified IP address within two seconds.

To run the **ping oob** command on a device (such as a VSU) with multiple MGMT interfaces, use the **via** field to specify the outbound MGMT interface. This ensures that the Ping function can be used properly.

1.5 Configuring a Traceroute Connectivity Test

1.5.1 Overview

You can use Traceroute connectivity test on a network device to learn the routing topology and the number of gateways between this device and the destination host.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure an IP packet for tracing.

(IPv4 network)

traceroute [**ip** | **vrf** *vrf-name*] { *hostname* | ipv4-*address* } [**out-interface** *interface-type interface-number* [**next-hop** *next-hop*] | **probe** *probe* | [**source** *interface-type interface-number* | **source** *source-ipv4-address*] | **timeout** *time* | **ttl** *minimum maximum*] *

traceroute oob { *hostname* | *ipv4-address* } [**out-interface** *interface-type interface-number* [**next-hop** *next-hop*] | **probe** *probe* | [**source** *interface-type interface-number* | **source** *source-ipv4-address*] | **timeout** *time* | **ttl** *minimum maximum* | **via** *mgmt-name*] *

(IPv6 network)

traceroute [**ipv6** | **vrf** *vrf-name*] { *hostname* | ipv6-*address* } [**out-interface** *interface-type interface-number* [**next-hop** *next-hop*] | **probe** *probe* | [**source** *interface-type interface-number* | **source** *source-ipv6-address*] | **timeout** *time* | **ttl** *minimum maximum*] *

traceroute oob { *hostname* | *ipv6-address* } [**out-interface** *interface-type interface-number* [**next-hop** *next-hop*] | **probe** *probe* | [**source** *interface-type interface-number* | **source** *source-ipv6-address*] | **timeout** *time* | **ttl** *minimum maximum* | **via** *mgmt-name*] *

By default, the timeout time is 3 seconds, the number of probe packets to be sent is 3, the minimum TTL is 1, and the maximum TTL is 255.

To run the **traceroute oob** command on a device (such as a VSU) with multiple MGMT interfaces, use the **via** field to specify the outbound MGMT interface of packets. This ensures that the Traceroute function can be used properly.

1.6 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may loss vital information and thus interrupt services.

Table 1-1 Monitoring

Command	Purpose
cleanup rping table { **all** | **ping-object** *owner-index test-index* | **trace-object** *owner-index test-index* }	Clears the Rping table entries.
display rping detail	Displays detailed information of Rping table entries.

1.7 Configuration Examples

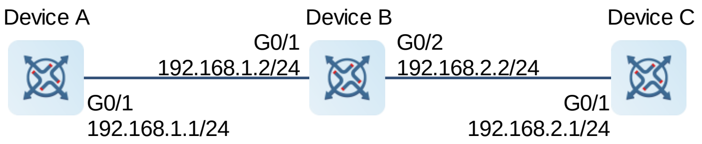
1.7.1 Configuring a Ping Connectivity Test

1. Requirements

Devices A, B, and C are connected to an IP network. It is required to check whether devices A and C can communicate with each other.

2. Topology

Figure 1-3 Configuring a Ping Connectivity Test



3. Notes

- Run the following command in privileged mode: **ping** 192.168.2.1.
- Run the following command in privileged mode: **ping** 192.168.2.1, and specify the length of the sent packet, number of sent packets, and TTL.

4. Procedure

- Use the common Ping function on device A.

  Display the brief Ping result.

```
DeviceA# ping 192.168.2.1
Sending 5, 100-byte ICMP Echoes to 192.168.2.1, timeout is 2 seconds:
< press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/2/10 ms
```

  Display the detailed Ping result.

```
DeviceA# ping 192.168.2.1 detail
Sending 5, 100-byte ICMP Echoes to 192.168.2.1, timeout is 2 seconds:
  < press Ctrl+C to break >
Reply from 192.168.2.1: bytes=100 time=4ms TTL=64
Reply from 192.168.2.1: bytes=100 time=3ms TTL=64
Reply from 192.168.2.1: bytes=100 time=1ms TTL=64
Reply from 192.168.2.1: bytes=100 time=1ms TTL=64
Reply from 192.168.2.1: bytes=100 time=1ms TTL=64

Success rate is 100 percent (5/5), round-trip min/avg/max = 1/2/4 ms.
```

- Use the extended Ping function on device A.

  Display the brief Ping result.

```
DeviceA# ping 192.168.2.1 length 1500 ntimes 100 data ffff source 192.168.1.1 timeout 3
Sending 100, 1500-byte ICMP Echoes to 192.168.2.1, timeout is 3 seconds:
  < press Ctrl+C to break >
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
Success rate is 100 percent (100/100), round-trip min/avg/max = 2/2/3 ms
```

  Display the detailed Ping result.

```
DeviceA# ping 192.168.2.1 length 1500 ntimes 10 data ffff source 192.168.1.1 timeout 3 detail
Sending 20, 1500-byte ICMP Echoes to 192.168.2.1, timeout is 3 seconds:
```

```
  < press Ctrl+C to break >
Reply from 192.168.2.1: bytes=1500 time=1ms TTL=64
Reply from 192.168.2.1: bytes=1500 time=1ms TTL=64
Reply from 192.168.2.1: bytes=1500 time=1ms TTL=64
Reply from 192.168.2.1: bytes=1500 time=1ms TTL=64
Reply from 192.168.2.1: bytes=1500 time=1ms TTL=64
Reply from 192.168.2.1: bytes=1500 time=1ms TTL=64
Reply from 192.168.2.1: bytes=1500 time=1ms TTL=64
Reply from 192.168.2.1: bytes=1500 time=2ms TTL=64
Reply from 192.168.2.1: bytes=1500 time=1ms TTL=64
Reply from 192.168.2.1: bytes=1500 time=1ms TTL=64

Success rate is 100 percent (10/10), round-trip min/avg/max = 1/1/3 ms.
```

5. Verification

Common Ping function: By default, five 100-byte packets are sent to the specified IP address. Within the specified time (2 seconds by default), the detection information is displayed, and statistical information is output.

```
DeviceA# ping 192.168.2.1
Sending 5, 100-byte ICMP Echoes to 192.168.2.1, timeout is 2 seconds:
< press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/2/10 ms
```

Extended Ping function: By default, ten 1500-byte packets are sent to the specified IP address. If there is a response within the specified time (3 seconds), the detection information is displayed, and statistical information is output.

```
DeviceA# ping 192.168.2.1 length 1500 ntimes 10 data ffff source 192.168.1.1 timeout 3
Sending 10, 1500-byte ICMP Echoes to 192.168.2.1, timeout is 3 seconds:
< press Ctrl+C to break >
!!!!!!!!!!
Success rate is 10 percent (10/10), round-trip min/avg/max = 2/2/3 ms
```

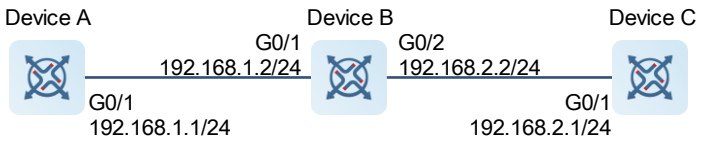
1.7.2 Configuring a Traceroute Connectivity Test

1. Requirements

Devices A, B, and C are connected to an IP network. Device A fails to communicate with device C. It is required to check whether the route between devices A and C is reachable. If the route is unreachable, confirm the faulty network node.

2. Topology

Figure 1-4 Configuring a Traceroute Connectivity Test



3. Notes

Run the following command in privileged mode: **traceroute** 192.168.2.1.

4. Procedure

(1) Configure an IP address on devices A, B, and C, respectively. (omitted)

(2) Configure a static route on device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA (cfg)# ip route-config 0.0.0.0 0.0.0.0 192.168.1.2
DeviceA (cfg)# exit
```

(3) Check whether the route between devices A and C is reachable.

```
DeviceA# ping 192.168.2.1
Sending 5, 100-byte ICMP Echoes to 192.168.2.1, timeout is 2 seconds:
  < press Ctrl+C to break >
.....
Success rate is 0 percent (0/5).
```

(4) Check the faulty network node.

```
DEVICEA# traceroute ip 192.168.2.1
  < press Ctrl+C to break >
Tracing the route to 192.168.2.1
1       192.168.1.2          1 msec    1 msec    1 msec
2       *          *          *
3       *          *          *
4       *          *          *
```

5. Verification

The displayed result indicates that the route between devices A and C is unreachable. The packet sent from device A to device C has arrived at device B, but a fault occurs to the connection between devices B and C.

Contents

1 Configuring Tech-Support 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Enabling the Tech-Support Function 1
1.2.1 Overview 1
1.2.2 Restrictions and Guidelines 2
1.2.3 Procedure 3
1.3 Monitoring 3

1 Configuring Tech-Support

1.1 Introduction

1.1.1 Overview

Tech-Support is an information collection function that provides diversified information collection triggering modes according to requirements of the administrator, and collects information on faulty equipment in different modes.

1.1.2 Principles

1. Basic collection

In basic collection mode, the directory files registered by the log framework are collected, including configuration files, Syslog, and CLI operations.

2. Full collection

In full collection mode, the following fault information is collected:

- All fault log files
- Coredump files
- Service realtime fault information

3. Hotkey collection

When the console fails, you can use the hotkey @@@@f to trigger one-click collection. For details about the collected content, see Full collection.

4. Reset key collection

Press the Reset key on the panel for less than 5s to collect the fault information. Five minutes later, the device is reset. For details about the collected content, see Full collection.

5. Storage of the collected content

The collected fault information is stored in the descending order of USB flash drive, **Flash:/**, and **Tmp:/**.

If the collected fault information is not stored on the USB flash drive, at most two compressed packages are stored.

6. Collection of real-time diagnostic information

When a fault occurs, you can run the **display system diag-info** command and necessary CLI commands in debugging support mode to collect the related fault information and save it in the **diag-info.txt** file on the Flash.

1.2 Enabling the Tech-Support Function

1.2.1 Overview

You can use the Tech-Support function to collect detailed fault information of a device and use it for fault location.

1.2.2 Restrictions and Guidelines

- When the device is connected to the console port, you can also use the hotkey **@@@@f** to collect detailed fault information of the device. The @@@@f hotkey operation is often performed when the console crashes. The fault information is packaged in a way similar to the **tech-support package** command.
- Before enabling the Tech-Support function, you are advised to check whether the device has sufficient space to store the fault information.
- To avoid an information collection failure due to insufficient storage space, you are advised to insert the USB flash drive into the device before enabling Tech-Support to avoid information loss. You can use the TFTP function to transfer the compressed packages of the final fault information to a PC.
- If the collected files are too large, it takes a long time to collect the information and services may be affected. To prevent this problem, if the collected coredump files exceed 200 MB or the number of files is too large, a collection failure notification will be displayed. In this case, you need to use other commands to collect coredump files separately.
- The specific procedure is as follows:

  a Run the **display coredump files** command to display the coredump file names.

```
Device# display coredump files
Directory of /xxx/xxx/xxx/ :
 core-xxx-xxx
```

  b Enter the debugging support mode, run the **execute diagnose-cmd copy** command to copy the coredump files to the **data** directory, and upload the files to the specified TFTP server.

```
Device# debugging support
Device(support)# execute diagnose-cmd copy xxx/xxx/xxx/core-xxx-xxx /data
This operation will modify file? [Y/N]:y
Device(support)# exit
Device# copy flash:/ xxx/xxx/xxx/core-xxx-xxx tftp://xxx.xxx.xxx.xxx/
This operation may overwrite file[tftp:// xxx.xxx.xxx.xxx /core-xxx-xxx]? [Y/N]:y
Press Ctrl+C to quit
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
Copy success.
```

- The compressed packages of fault information are stored as follows:
  - USB flash drive

    If a USB flash drive is inserted into the device, the compressed packages are stored in the root directory of the USB flash drive. You can run the **dir usb0:** command to view these compressed packages.
  - Flash directory

    If no USB flash drive is inserted into the device, the compressed packages are preferentially stored in the Flash directory. You can run the **dir flash:** command to view these compressed packages.
  - **Tmp** directory (memory)

    If no USB flash drive is inserted into the device and the Flash has insufficient space, the compressed packages are stored in the **Tmp** directory. You can run the **dir tmp:** command to view these compressed packages.
- To prevent excessive consumption of space on the storage media after the fault information is collected for

multiple times, a maximum of three compressed packages of fault information can be stored on the same storage medium.

1.2.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(1) Enter the debugging configuration mode.

debugging support

(2) Enable the Tech-Support function. The configuration tasks below are mutually exclusive. Please configure only one task.

- Run the following command to enable the Tech-Support function:

tech-support package [*component-name* | **basic**]

- Use the hotkey to enable the Tech-Support function.

@@@@f

This command takes effect only when the device is connected to the console port, and is invalid for remote connections, for example, Telnet or SSH connection.

(3) Run the diagnose command:

execute diagnose-cmd { [**device** *device-id*] [**slot all**] [**chip all** | **chip** *chip-id*] *shell-command* | **help** }

Note

- The **device** keyword is supported by only the S6520-64CQ, S6510-32CQ, S6510-4C, S6510-48VS8CQ, and S6250-48XS8CQ.

1.3 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Table 1-1 Monitoring

Command	Purpose
display system diag-info	Displays the diagnostic information when a fault occurs.

Contents

1 Configuring SPAN 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Concepts 1
1.1.3 Principles 2
1.2 Restrictions and Guidelines 5
1.3 Configuration Task Summary 6
1.4 Configuring SPAN 7
1.4.1 Overview 7
1.4.2 Restrictions and Guidelines 7
1.4.3 Procedure 7
1.5 Configuring RSPAN 8
1.5.1 Overview 8
1.5.2 Restrictions and Guidelines 8
1.5.3 Procedure 8
1.6 Configuring ERSPAN 9
1.6.1 Configuration Tasks 9
1.6.2 Configuring Basic Functions of ERSPAN 9
1.6.3 Configuring the ERSPAN Sampling Function 10
1.6.4 Configuring ERSPAN Attributes 11
1.7 Monitoring 12
1.8 Configuration Examples 12

1.8.1 Configuring SPAN 12

1.8.2 Configuring RSPAN to Realize One-to-Many SPAN 15

1.8.3 Configuring Basic Functions of ERPSAN 18

1 Configuring SPAN

1.1 Introduction

1.1.1 Overview

With the development of network technologies, the demand for high network availability is increasing. When exceptions occur in a network, the network node or device port needs to be analyzed for data traffic so that the network can return to normal fast. Meanwhile, the device traffic forwarding should not be affected.

The mirroring function copies messages from a specified port to another port connected to a network monitoring device. The device analyzes data traffic of a suspicious network node or device port, without affecting the traffic forwarding of the monitored device. Mirroring is mainly applied to network monitoring and troubleshooting to monitor network information for security and troubleshoot network failures.

1.1.2 Basic Concepts

1. Source Port

A source port is also called a monitored port. The data traffic of a source port is copied to a destination port for network analysis or troubleshooting.

2. Destination Port

A destination port is also called a monitoring port. A destination port is connected to a device to forward packets received from a source port to a monitoring device.

3. Session

A session is a logical concept, which consists of a source port and a destination port.

4. SPAN

Switch port analyzer (SPAN) is also called local port mirroring or local mirroring. In local port mirroring, the source port and the destination port in a mirroring session are in the same device.

5. RSPAN

Remote switch port analyzer (RSPAN) is an extension of SPAN. In RSPAN, the source port and the destination port in a mirroring session are in different devices.

6. ERSPAN

Encapsulated remote switch port analyzer (ERSPAN) is an extension of RSPAN. ERSPAN allows monitoring multiple devices across switched or routing networks.

7. RSPAN VLAN

An RSPAN VLAN is a special VLAN that only transmits mirrored packets to a destination device.

8. Output Port

An output port on a source device sends mirrored packets to a port on an intermediate device or a destination device.

9. Source Device

A source device is a role in RSPAN and has a source port. A source device copies the packets from a source port and outputs them through an output port on the source device to an intermediate device or destination device.

10. Destination Device

A destination device is a role in RSPAN and has a destination port. A destination device forwards the mirrored packets received from an intermediate or source device to a monitoring device through a destination port.

11. Intermediate Device

An intermediate device is a role in RSPAN and resides between a source device and a destination device. An intermediate device forwards mirrored packets to another intermediate device or a destination device. If a source device is directly connected to a destination device, there is no intermediate device.

12. SPAN Data Streams

SPAN data streams refer to mirrored packets. A SPAN session contains data streams in three directions:

- Input data streams

  All packets received by a source port are copied to a destination port. Users can monitor input packets of one or more source ports in a SPAN session.

- Output data streams

  All packets transmitted by a source port are copied to a destination port. Users can monitor output packets of one or more source ports in a SPAN session.

- Bidirectional data streams

  Bidirectional data streams include input data streams and output data streams. In a SPAN session, users can monitor data streams of one or more source ports in the input and output directions.

1.1.3 Principles

1. SPAN Principles

As shown in Figure 1-1, all packets on port 1 are copied to port 10 through configuring SPAN on Device A. The network analyzer connected to port 10 receives all packets that pass through port 1 even if the network analyzer is not directly connected to port 1. Therefore the network analyzer can monitor the data streams that pass through port 1. The specific process is as follows:

(1) The device identifies and marks the incoming and outgoing packets of the source port 1.

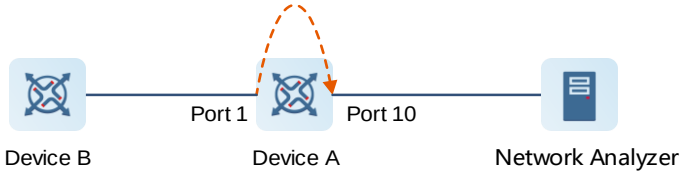
(2) The device searches for the SPAN session to which port 1 belongs.

(3) The device searches for the destination port 10 corresponding to the SPAN session.

(4) The device copies both incoming and outgoing packets of port 1 and sends them to port 10.

(5) Port 10 sends the copied packets to the network analyzer.

Figure 1-1 SPAN Principles



2. RSPAN Working Process

As shown in Figure 1-2, an RSPAN VLAN is created for the source device, intermediate device, and destination device. All ports involved in an RSPAN session must be added to the RSPAN VLAN. Mirrored packets are broadcast in the RSPAN VLAN so that they are forwarded from the source device to the destination device. Then the destination device forwards the mirrored packets to the network analyzer. The following is the specific working process.

(1) Add all ports involved in an RSPAN session of a source device, intermediate device, and destination device to an RSPAN VLAN.

(2) The source device adds the VLAN ID of the RSPAN VLAN to the mirrored packets of the source port. Then the device copies the encapsulated packets to the output port.

Note

As shown in Figure 1-4, for 1:N port mirroring, enable the loopback function on the output port of the source device to make it a loopback port. Add the loopback port (port 1) and the interface of the intermediate device (port 2) to the RSPAN VLAN. The source port adds the VLAN ID of the RSPAN VLAN to the mirrored packets and copies the packets to the loopback port. The loopback port belongs to the RSPAN VLAN and this VLAN is forbidden to learn any MAC address. Therefore the reflected packets are broadcast in the RSPAN VLAN so that all ports in this VLAN can receive the mirrored packets to achieve 1:N port mirroring .

(3) The output port sends the mirrored packets to a port on an intermediate device or a destination device.

(4) The intermediate device broadcasts mirrored packets in the RSPAN VLAN to transparently transmit the packets to another intermediate device or a destination device.

Note

If a source device is directly connected to a destination device, there is no intermediate device.

(5) After receiving the packets, the destination device determines whether they are mirrored packets through the VLAN ID. If the VLAN ID carried by the packets is the same as the RSPAN VLAN ID, the device regards the packet as a mirrored packet and forwards the packet to the network analyzer through the destination port.

Figure 1-2 Principles of RSPAN

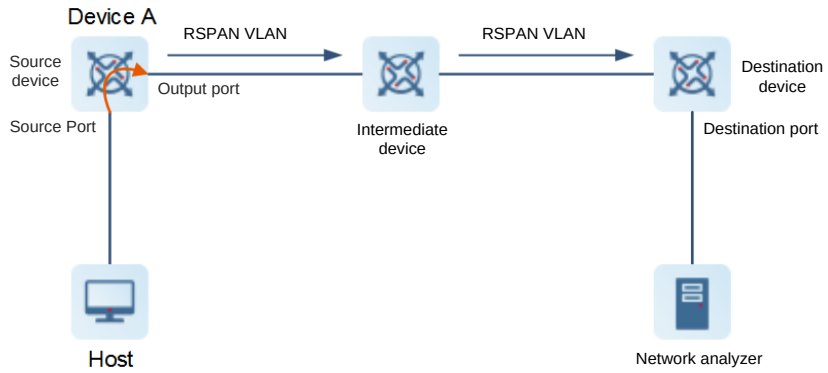
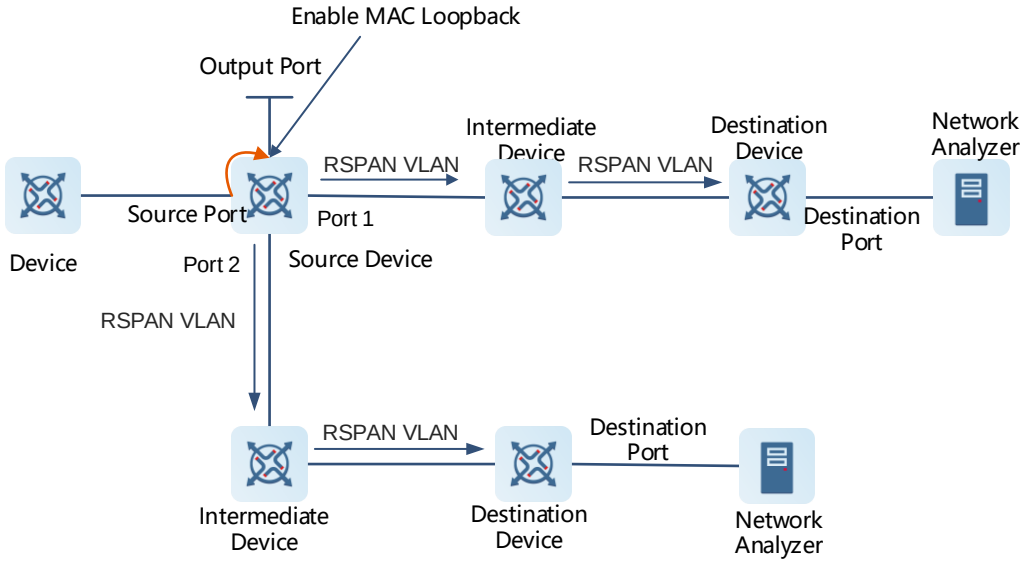


Figure 1-3 Principles of 1:N Port Mirroring Based on RSPAN



3. ERSPAN Working Process

In RSPAN, mirrored packets can be transmitted only at Layer 2 and cannot transmitted across routing networks. However, ERSPAN can encapsulate mirrored packets into IP packets through a generic routing encapsulation (GRE) tunnel and forward the mirrored packets to the destination port of the RSPAN device, thus transmitting mirrored packets across routing networks.

Figure 1-4 shows the specific working process of ERSPAN.

(1) The source device copies and encapsulates the incoming packets of the source port.

- Encapsulate the source and destination IP addresses of the GRE packet.
- Set the TTL and DSCP of the GRE packet.

Note

Figure 1-5 shows the header format of a packet encapsulated in ERSPAN.

(2) The source port forwards the mirrored packets to the destination device through the GRE tunnel.

(3) The destination device decapsulates GRE packets to remove the header encapsulated by ERSPAN, and forwards the packets to the destination port.

(4) The destination port forwards the mirrored packets to the netwokr analyzer.

Caution

The IP packets after GRE encapsulation must be normally routed to the destination device in the network. Therefore, the destination IP address is generally the next-hop IP address of the route. That is, the output port of the source device is the next-hop egress of the Layer 3 path where the destination IP address to be encapsulated is located. The output port must be reachable, and can be an SVI, Layer 3 aggregate interface, or Layer 3 Ethernet interface.

Figure 1-4 ERSPAN Principles

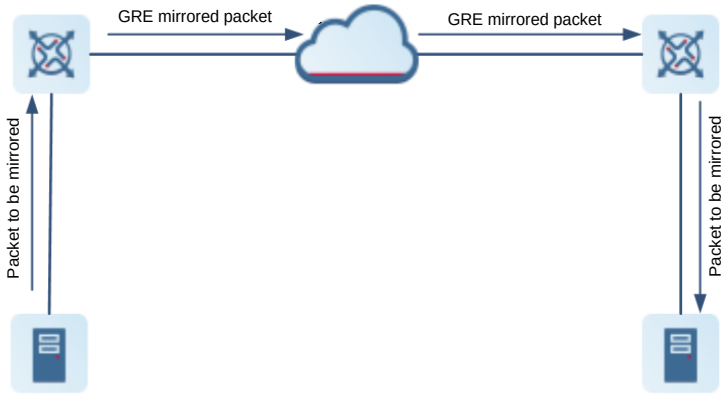


Figure 1-5 Format of an Encapsulated Packet in ERSPAN

MAC_HEADER(14/18B)	IP_HEADER(20B)	GRE_HEADER(48B)	L2_Mirrored_Packet(untagged)

4. Stream Mirroring Principles

Stream mirroring is to associate the source port with an ACL on the basis of SPAN, and filter packets based on the ACL rules, so as to mirror only specified packets. Only the packets matching the permit ACE in the ACL on the source port can be mirrored to the destination port. A source port can be associated with only one ACL. Currently, the standard ACL, extended ACL, MAC ACL, and custom ACL can be associated.

1.2 Restrictions and Guidelines

- A source port has the following features:
  - The source port and the output port cannot be the same one.
  - The source port can be a Layer 2 Ethernet interface, Layer 3 Ethernet interface, Layer 2 link aggregation or Layer 3 link aggregation.
  - Data streams on multiple source ports on the source device can be mirrored to a specified output port.

- When the source port is a Layer 3 link aggregation or Ethernet interface, the port monitors both Layer 2 and Layer 3 packets.
- Upon bidirectional monitoring of multiple ports, a packet is received from one port and output from another port, and monitoring is effective as long as one packet is monitored among these ports.
- When a port is STP enabled and in the block state, the incoming and outgoing packets of the port can also be monitored.
- The source port and the destination port can belong to the same VLAN or different VLANs.

  If a source port is added to a link aggregation, the source port exits the SPAN session. The source port acts as only a member port of the link aggregation.

- A destination port has the following features:
  - A destination port cannot be used as a source port simultaneously.
  - The destination port can be an Ethernet interface, link aggregation or an output group.
  - The SPAN destination port is used for the Spanning Tree Protocol (STP) calculation.
- A source port cannot be added to an output group.
- The destination port is not enabled with switching function and only receives SPAN packets by default. To allow the destination port to forward packets, enable switching function on the port. Otherwise, the packets passing through the port are dropped.
- Due to other reasons (such as port security), incoming packets of a source port may be discarded, but it does not affect the SPAN function, and these packets are still mirrored to the destination port. Due to other reasons, packets sent from other ports to a source port may be discarded, and these packets are not sent to the destination port.
- If the format of a packet output from the source port is changed, for example, the source MAC, destination MAC, VLAN ID, and TTL of the packet change, the format of the packet mirrored to the destination port will change as well.
- Configure the following types of ACL on the source port to mirror specified packets: standard ACL, extended ACL, MAC ACL, and custom ACL.

1.3 Configuration Task Summary

SPAN configuration includes the following tasks:

- Configuring SPAN
- Configuring RSPAN
- Configuring ERSPAN
  - a Configuring Basic Functions of ERSPAN
  - b (Optional) Configuring the ERSPAN Sampling Function
  - c (Optional) Configuring ERSPAN Attributes

1.4 Configuring SPAN

1.4.1 Overview

SPAN copies packets of the designated port or designated VLAN to the port connected to the data monitoring device, so that you can use the data monitoring device to analyze these copied packets for network monitoring and troubleshooting.

1.4.2 Restrictions and Guidelines

- When configuring the SPAN source port, you can specify the following items as the data source simultaneously: some VLANs, an interface, and specified streams on the specified port.
- When configuring the SPAN source port, if you specify a VLAN as the data source, you cannot specify an interface or streams on the specified port as the data source.
- When configuring the SPAN source port, if you specify some VLANs as the data source, you must also specify the source port.
- If you change the VLAN configuration on the source port or destination port, the configuration takes effect immediately.
- If you disable the source or destination port, the SPAN function does not take effect.
- If you configure a VLAN or a VLAN list as the SPAN source, ensure that the destination port has sufficient bandwidth for receiving mirrored data of the entire VLAN (or VLAN list).
- If you add or delete a VLAN source port to or from an effective SPAN session, you need to re-apply the entire SPAN session. Therefore, a few existing mirrored packets may be lost.
- If SPAN source and destination ports are both configured as LACP aggregate interfaces, the LACP flapping occurs repeatedly on the destination port and its peer port.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a SPAN source port. Configure at least one of the following tasks.

- Specify an interface as the data source of SPAN.

  mirror session-id *session-number* **source port** *interface-type interface-number* [**both** | **rx** | **tx**]

- Specify some streams of an interface as the data source of SPAN.

  mirror session-id *session-number* **source port** *interface-type interface-number* **rx acl** { *acl-name* | *acl-number* [**priority** *priority-number*] }

- (Optional) Exclude one or more VLANs as the data source of SPAN.

  mirror session-id *session-number* **filter vlan** *vlan-id-list* **rx**

  This command must be configured together with monitor session source **port** or monitor session source **port acl**

- Specify one or more VLANs as the data source of SPAN.

mirror session-id *session-number* **source vlanid** *vlan-id-list* **rx**

This function is mutually exclusive with the preceding three configurations.

By default, no source port is configured for a SPAN session.

(4) Configure a SPAN destination port.

mirror session-id *session-number* **destination** { **port** *interface-type interface-number* } [**switch**]

By default, no SPAN destination port is configured.

1.5 Configuring RSPAN

1.5.1 Overview

A device enabled with RSPAN can monitor packets acrorss networks for network monitoring and troubleshooting.

1.5.2 Restrictions and Guidelines

- All ports involved in RSPAN must be added to a remote VLAN.
- An RSPAN VLAN must be configured on each device and their VLAN IDs must be consistent. All ports involved in a session must be added to the VLAN. Do not add common ports to an RSPAN VLAN.
- Do not configure a port that is connected to an intermediate device or a destination device as an RSPAN source port. Otherwise, traffic on the network may be in chaos.
- To achieve 1:N port mirroring, configure a port on a source device as a loopback port. The loopback function on the output port will transmit mirrored packets to multiple intermediate devices or destination devices.
- The loopback port cannot serve as a normal port to forward packets. You are advised to configure a port in Down state as the loopback port, and not to add other configurations for this port.
- If RSPAN source and destination ports are both configured as LACP aggregate interfaces, the LACP flapping occurs repeatedly on the destination port and its peer port.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VLAN configuration mode.

vlan *vlan-id*

(4) Configure a remote VLAN.

remote-probe

By default, no remote VLAN is configured.

(5) Exit the VLAN configuration mode.

exit

(6) Configure an RSPAN source device.

mirror session-id *session-number* **remote-source**

By default, no source device is configured for an RSPAN session.

(7) (Optional) Configure the MAC loopback function on the source device.

a Enter the Layer 2 or Layer 3 Ethernet interface configuration mode.

interface *interface-type interface-number*

b Enable the MAC loopback function of the interface.

mac-loopback

By default, MAC loopback is disabled for an interface.

(8) Exit the Layer 2 or Layer 3 Ethernet interface configuration mode.

exit

(9) Configure a source port on the source device.

mirror session-id *session-number* **source port** *interface-type interface-number* [{ **both** | **rx** [**acl** { *acl-name* | *acl-number* [**priority** *priority-number*] | **tx** } }]

By default, no source port on the source device is configured for an RSPAN session.

(10) Configure an RSPAN destination device.

mirror session-id *session-number* **remote-destination**

By default, no destination device is configured for an RSPAN session.

(11) Configure an output port on the source device or a destination port on the destination device.

mirror session-id *session-number* **destination remote vlan** *remote-vlan-id* **port** *interface-type interface-number* [**switch**]

By default, no output port on the RSPAN source device or no destination port on the RSPAN destination device is configured.

1.6 Configuring ERSPAN

1.6.1 Configuration Tasks

ERSPAN configuration includes the following tasks:

(1) Configuring Basic Functions of ERSPAN

(2) (Optional) Configuring the ERSPAN Sampling Function

(3) (Optional) Configuring ERSPAN Attributes

1.6.2 Configuring Basic Functions of ERSPAN

1. Overview

After ERSPAN is configured, a network analyzer can be used to monitor data streams of network devices. Devices can normally exchange data with each other.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ERSPAN session.

mirror session-id *session-number* **erspan**

No ERSPAN session is configured by default.

(4) Specify an interface as the data source of ERSPAN.

source port *interface-type interface-number* [**both** | **rx** [**acl** { *acl-name* | *acl-number* } [**sample | priority** *priority-number*]] | **tx**]

By default, no source port is configured. If a source port is configured, the default mirroring direction is bidirectional.

When a sub-interface is configured as the source port, mirroring is supported only in the inbound direction.

(5) Configure all ports that support mirroring as the mirroring data source.

source port all rx acl { *acl-name* | *acl-number* } [**sample | priority** *priority-number*]

No mirroring source port is configured by default.

(6) Configure the source IP address for encapsulation.

original { **ip** | **ipv6** } **address** *address*

By default, no source IP address is configured for GRE encapsulation.

(7) Configure the destination IP address for encapsulation.

destination { **ip** | **ipv6** } **address** *address*

By default, no encapsulation type or destination IP address is configured.

1.6.3 Configuring the ERSPAN Sampling Function

1. Restrictions and Guidelines

The ERSPAN sampling function takes effect only when the sampling function is configured in an ERSPAN session or source port.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ERSPAN session.

mirror session-id *session-umber* **erspan**

(4) Configure a source port.

source port { **all rx** { **acl** *acl-name* | **acl** *acl-number* } | *interface-type interface-numbere* [{ **both** | **rx** [**acl** *acl-name* | **acl** *acl-number* [**sample** | **priority** *priority-number*]] | **tx** }] }

By default, no source port is configured. If a source port is configured, the mirroring direction is **both** by default.

(5) Configure the source IP address for encapsulation.

original { **ip** | **ipv6** } **address** *address*

By default, no source IP address is configured for GRE encapsulation.

(6) Configure the destination IP address for encapsulation.

destination { **ip** | **ipv6** } **address** *address*

By default, no encapsulation type or destination IP address is configured.

(7) (Optional) Configure the sampling rate.

sampling-rate *rate*

The default sampling rate is **1:1**, that is, every packet is sampled.

1.6.4 Configuring ERSPAN Attributes

1. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an ERSPAN session.

mirror session-id *session-number* **erspan**

(4) Configure attributes of an ERSPAN session. The following configurations are optional. Select at least one of the configurations as actually needed.

- (Optional) Configure the TTL of the encapsulated IP packet.

  ip ttl *ttl-value*

  The default TTL of the encapsulated IP packet is **64**.

- (Optional) Configure the DSCP of the encapsulated IP packet.

  ip dscp *dscp-value*

  The default DSCP of the encapsulated IP packet is **0**.

(Optional) Configure the hop-limit value of encapsulated IPv6 packets.

ipv6 hop-limit *value*

The default hop-limit value of encapsulated IPv6 packets is 64.

Note

This command is supported by only the S6520-64CQ.

- (Optional) Configure the hop-limit value of encapsulated IPv6 packets.

  ipv6 max-hops *value*

  The default hop-limit value of encapsulated IPv6 packets is **64**.

- (Optional) Configure the traffic-class value of encapsulated IPv6 packets.

  ipv6 traffic-class *value*

  The default traffic-class value of encapsulated IPv6 packets is **0**.

Note

This command is supported by only the S6520-64CQ.

- (Optional) Associate ERSPAN with a VRF.

  vrf *vrf-name*

  By default, no VRF is associated with ERSPAN.

  When this function is configured, the VRF must be already configured.

- (Optional) Configure the header type value of ERSPAN packets.

  header type3

Note

This command is supported by only the S6580-48CQ8QC, and S6580-24DC8QC.

(5) (Optional) Shut down the ERSPAN session.

shutdown

The session is enabled by default.

1.7 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Table 1-1 Monitoring

Command	Purpose
display mirror [**session-id** *session-number*]	Displays the SPAN session information.
debugging span	Debugs the SPAN.

1.8 Configuration Examples

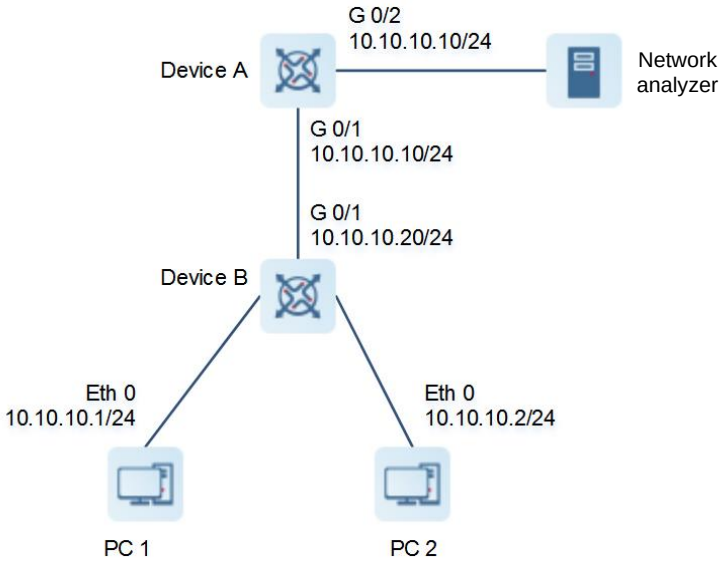
1.8.1 Configuring SPAN

1. Requirements

As shown in Figure 1-6, after appropriate configuration, the network analyzer can monitor all data streams forwarded by device A to device B and specified data streams of device B (for example, data streams fromHost A and Host B).

2. Topology

Figure 1-6 SPAN Topology



3. Notes

- Add GigabitEthernet 0/1 and GigabitEthernet 0/2 of device A to VLAN 1. Create SVI 1 and set the address of SVI 1 to 10.10.10.10/24.
- Add GigabitEthernet 0/1 of device B to VLAN 1. Create SVI 1 and set the address of SVI 1 to 10.10.10.20/24.
- Set the IP addresses of PC 1 and PC 2 to 10.10.10.1/24 and 10.10.10.2/24, respectively.
- Configure SPAN for Device A and configure ports GigabitEthernet 0/1 and GigabitEthernet 0/2 as the source port and destination port of SPAN respectively. Monitor all data streams forwarded by device A to device B and specified data streams of device B.

4. Procedure

Add GigabitEthernet 0/1 and GigabitEthernet 0/2 of device A to VLAN 1.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan 1
DeviceA(cfg-vlan)# exit
```

On device A, create SVI 1 and set the address of SVI 1 to 10.10.10.10/24.

```
DeviceA(cfg)# interface vlan 1
DeviceA(cfg-if-VLAN 1)# ip address 10.10.10.10 255.255.255.0
DeviceA(cfg-if-VLAN 1)# exit
```

Add GigabitEthernet 0/1 of device B to VLAN 1.

```
DeviceB# config-system terminal
DeviceB(cfg)# vlan 1
DeviceB(cfg-vlan)# exit
```

On device B, create SVI 1 and set the address of SVI 1 to 10.10.10.20/24.

```
DeviceB (cfg)# interface vlan 1
DeviceB(cfg-if-VLAN 1)# ip address 10.10.10.20 255.255.255.0
DeviceB(cfg-if-VLAN 1)# exit
```

On device A, configure an ACL to match packets with the source address 10.10.10.20.

```
DeviceA(cfg)# access-list 100 permit ip host 10.10.10.20 any
```

On device A, configure GigabitEthernet 0/1 as the SPAN source port, and monitor all data streams forwarded by device A to device B and specified data streams of device B.

```
DeviceA(cfg)# mirror session-id 1 source interface gigabitethernet 0/1 tx
DeviceA(cfg)# mirror session-id 1 source interface gigabitethernet 0/1 rx acl 100
```

On device A, configure GigabitEthernet 0/2 on device A as the SPAN destination port.

```
DeviceA(cfg)# mirror session-id 1 destination interface gigabitethernet 0/2
```

5. Verification

Run the **display mirror** command to check whether the SPAN configuration is correct.

```
DeviceA# display mirror
sess-num: 1
span-type: LOCAL_SPAN
src-intf:
   GigabitEthernet 0/1          frame-type TX Only
src-intf:
   GigabitEthernet 0/1          frame-type RX Only
rx acl id 100
dest-intf:
   GigabitEthernet 0/2
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip access-control-list extended 100
  10 permit ip host 10.10.10.20 any
!
interface VLAN 1
  ip address 10.10.10.10 255.255.255.0
!
mirror session-id 1 destination interface GigabitEthernet 0/2
mirror session-id 1 source interface GigabitEthernet 0/1 tx
mirror session-id 1 source interface GigabitEthernet 0/1 rx acl 100
!
end
```

- Device B configuration file

```
hostname DeviceB
```

```
!
interface VLAN 1
 ip address 10.10.10.20 255.255.255.0
!
end
```

1.8.2 Configuring RSPAN to Realize One-to-Many SPAN

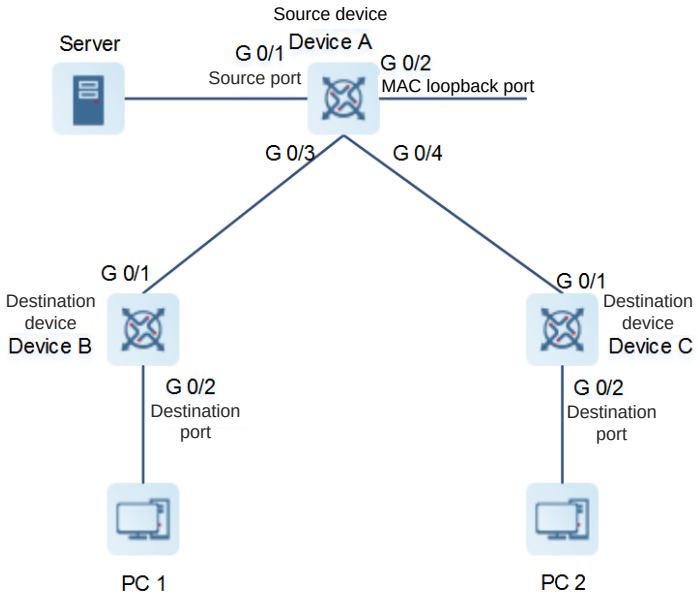
> **Caution**
>
> When deploying One-to-Many SPAN, you must run the **switch-port trunk allowed vlan remove** *vlan-list* command to remove the broadcast VLANs from a trunk interface on each SPAN device. Otherwise, the trunk interface may be fully occupied by the broadcast traffic, causing service interruption.

1. Requirements

As shown in Figure 1-7, the RSPAN function enables the network analyzer to monitor bidirectional data streams of the source device A from the destination devices B and C. The devices can normally exchange data with each other.

2. Topology

Figure 1-7 Configuring RSPAN to Realize One-to-Many SPAN



3. Notes

- Configure device A as the source device. Configure an RSPAN VLAN. Configure GigabitEthernet 0/1 as the source port, GigabitEthernet 0/3 and GigabitEthernet 0/4 connected to the destination device as the output ports, and GigabitEthernet 0/2 as the MAC loopback port to realize the reflection function of 1:N SPAN. Configure the switching function of the output port.
- Configure devices B and C as the destination devices. Configure an RSPAN VLAN. Configure

GigabitEthernet 0/1 connected to the source device as the source port. Configure a trunk port. Configure GigabitEthernet 0/2 connected to the network analyzer as the SPAN destination port, and configure the switching function for the SPAN destination port.

4. Procedure

(1) Configure device A as the source device.

a. Configure a remote VLAN.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# vlan 7
DeviceA(cfg-vlan)# remote-probe
DeviceA(cfg-vlan)# exit
```

b. Configure a source port.

```
DeviceA(cfg)# mirror session-id 1 remote-source
DeviceA(cfg)# mirror session-id 1 source interface gigabitethernet 0/1 both
```

c. Configure an output port.

```
DeviceA(cfg)# mirror session-id 1 destination remote vlan 7 interface gigabitethernet 0/2 switch
```

d. Configure a loopback port.

```
DeviceA(cfg)# port gigabitethernet 0/2
DeviceA(cfg-if)# mac-loopback
DeviceA(cfg-if)# switch-port  access vlan 7
DeviceA(cfg-if)# exit
DeviceA(cfg)# interface range gigabitethernet 0/3-4
DeviceA(cfg-if-list)# switch-port  mode trunk
```

(2) Configure device B as the destination device.

a. Configure a remote VLAN.

```
DeviceB> privilege
DeviceB# config-system
DeviceB(cfg)# vlan 7
DeviceB(cfg-vlan)# remote-probe
DeviceB(cfg-vlan)# exit
```

b. Configure a destination port.

```
DeviceB(cfg)# mirror session-id 1 remote-destination
DeviceB(cfg)# mirror session-id 1 destination remote vlan 7 port gigabitethernet 0/2 switch
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if)# switch-port  mode trunk
```

(3) Configure device C as the destination device.

a. Configure a remote VLAN.

```
DeviceC> privilege
DeviceC# config-system
DeviceC(cfg)# vlan 7
DeviceC(cfg-vlan)# remote-probe
DeviceC(cfg-vlan)# exit
```

b. Configure a destination port.

```
DeviceC(cfg)# mirror session-id 1 remote-destination
DeviceC(cfg)# mirror session-id 1 destination remote vlan 7 port gigabitethernet 0/2 switch
DeviceC(cfg)# interface gigabitethernet 0/1
DeviceC(cfg-if)# switch-port    mode trunk
```

5. Verification

Run the **display mirror** command on device A to view the SPAN source device configurations.

```
DeviceA# display mirror
sess-num: 1
span-type: SOURCE_SPAN
src-intf:
GigabitEthernet 0/1          frame-type Both
dest-intf:
GigabitEthernet    0/2
Remote vlan 7
mtp_switch on
```

Run the **display mirror** command on devices B and C to view the SPAN destination device configurations. Device B is taken as an example below:

```
DeviceB# display mirror
sess-num: 1
span-type: DEST_SPAN
dest-intf:
GigabitEthernet    0/2
Remote vlan 7
mtp_switch on
```

6. Configuration Files

- Device A configuration file of

```
hostname DeviceA
!
vlan 7
!
 remote-probe
!
interface GigabitEthernet 0/2
 mac-loopback
 switch-port    access vlan 7
!
interface GigabitEthernet 0/3
switch-port    mode trunk
interface GigabitEthernet 0/4
switch-port    mode trunk
!
mirror session-id 1 remote-source
```

```
mirror session-id 1 source interface GigabitEthernet 0/1 both
mirror session-id 1 destination remote vlan 7 interface GigabitEthernet0/2 switch
!
end
```

- Device B configuration file

```
hostname DeviceB
!
vlan 7
 remote-probe
!
interface GigabitEthernet 0/1
 switch-port    mode trunk
!
mirror session-id 1 remote-destination
mirror session-id 1 destination remote vlan 7 interface GigabitEthernet0/2 switch
!
end
```

- Device C configuration file

```
hostname Device
!
vlan 7
 remote-probe
!
interface GigabitEthernet 0/1
 switch-port    mode trunk
 !
mirror session-id 1 remote-destination
mirror session-id 1 destination remote vlan 7 interface GigabitEthernet 0/2 switch
!
end
```

7. Common Errors

- A remote VLAN must be configured on the source device, intermediate device, and destination device, and their VLAN IDs must be consistent.
- Packet loss may occur if packets of a port with large bandwidth are mirrored to a port with small bandwidth.

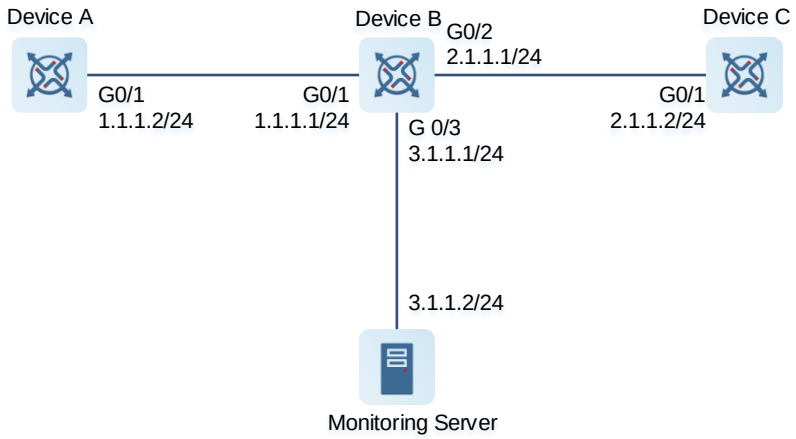
1.8.3 Configuring Basic Functions of ERPSAN

1. Requirements

As shown in Figure 1-8, the network administrator wants to use the network analyzer to monitor traffic on devices A and C. The network administrator monitors bidirectional packets of device A. For device B, the network administrator monitors only some incoming packets. To make reasonable use of resources, the monitored packets are sampled at a sampling rate of 1000.

2. Topology

Figure 1-8 Configuring Basic Functions of ERSPAN



3. Notes

- Configure reachable Layer 3 routes between devices.
- Configure device B as the source device. Configure GigabitEthernet 0/1 and GigabitEthernet 0/2 as the source ports to mirror packets in both directions on GigabitEthernet 0/1, mirror packets with the destination IP address 2.1.1.1 on GigabitEthernet 0/2, and configure GigabitEthernet 0/3 as the output port.
- Set the sampling rate of mirrored packets to 1000.

4. Procedure

Configure the IP address of Device B.

```
DeviceB> privilege
DeviceB # config-system terminal
DeviceB(cfg)# interface gigabitethernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 1.1.1.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
DeviceB(cfg)# interface gigabitethernet 0/2
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 2.1.1.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
DeviceB(cfg)# interface gigabitethernet 0/3
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 3.1.1.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Configure the IP address of device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 1.1.1.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Configure the IP address of device C.

```
DeviceC> privilege
```

```
DeviceC# config-system terminal
DeviceC(cfg)# interface gigabitethernet 0/1
DeviceC(cfg-if-GigabitEthernet 0/1)# ip address 2.1.1.2 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/1)# exit
```

Configure an ACL on Device B.

```
DeviceB(cfg)#access-list 1 permit host 1.1.1.1
```

Create ERSPAN Session 1, and configure Device as the source device. Configure GigabitEthernet 0/1 as the source port, and mirror bidirectional packets. Configure GigabitEthernet 0/2 as the source port, and mirror packets with the destination IP address 1.1.1.1.

```
DeviceB(cfg)# mirror session-id 1 erspan-source
DeviceB(cfg-monitor-erspan)# source interface gigabitethernet 0/1 both
DeviceB(cfg-monitor-erspan)# source interface gigabitethernet 0/2 rx acl 1
DeviceB(cfg-monitor-erspan)# origin ip address 10.1.1.2
DeviceB(cfg-monitor-erspan)# destination ip address 3.1.1.2
```

Enable the sampling function and set the sampling rate to 1000.

```
DeviceB(cfg-monitor-erspan)# sample enable
DeviceB(cfg-monitor-erspan)# sample rate 1000
```

5. Verification

Check that the Layer 3 route between Device B and Device A is reachable.

```
DeviceB# ping 1.1.1.2
Sending 5, 100-byte ICMP Echoes to 1.1.1.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/3/8 ms.
```

Check that the Layer 3 route between Device B and Device C is reachable.

```
DeviceB# ping 2.1.1.2
Sending 5, 100-byte ICMP Echoes to 2.1.1.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/3/8 ms.
```

Check that the Layer 3 route between Device B and the network analyzer is reachable.

```
DeviceB# ping 3.1.1.2
Sending 5, 100-byte ICMP Echoes to 3.1.1.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/3/8 ms.
```

Run the **display access-lists** command to check that the ACL configuration takes effect.

```
DeviceB# display access-lists

ip access-control-list standard 1
  10 permit host 1.1.1.1
```

Run the **display mirror** command to check that the ERSPAN session is in Active state in the configuration result.

```
DeviceB# display mirror
sess-num: 1
span-type: ERSPAN_SOURCE
src-intf:
   GigabitEthernet 0/1          frame-type: Both          TX status: Active     RX status: Active
src-intf:
   GigabitEthernet 0/2          frame-type: RX Only       Rx acl id: 1
status: Active
original ip address: 2.1.1.2
destination ip address: 3.1.1.2
ip ttl: 64
ip dscp: 0
sample rate: 1000
vrf: reset
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 0/1
 ip address 1.1.1.2 255.255.255.0
!
end
```

- Device C configuration file

```
hostname DeviceC
!
interface GigabitEthernet 0/1
 ip address 2.1.1.2 255.255.255.0
!
end
```

- Device B configuration file

```
hostname DeviceB
!
ip access-control-list standard 1
 10 permit host 1.1.1.1
!
interface GigabitEthernet 0/1
 ip address 1.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/2
 ip address 2.1.1.1 255.255.255.0
!
interface GigabitEthernet 0/3
```

```
 ip address 3.1.1.1 255.255.255.0
!
mirror session-id 1 erspan-source
 source interface gigabitEthernet 0/1 both
 source interface gigabitEthernet 0/2 rx acl acl1
 origin ip address 10.1.1.2
 destination ip address 3.1.1.2
 sample enable
 sample rate 1000
!
end
```

7. Common Errors

- The session ID used to configure ERSPAN is already configured with RSPAN or LOCAL SPAN.
- The Layer 3 route between the source device and destination device is unreachable.

Contents

1 Configuring Packet Capture 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 1
1.3 Capturing Packets Destined for the CPU 1
1.3.1 Overview 1
1.3.2 Restrictions and Guidelines 2
1.3.3 Procedure 2
1.4 Capturing Forwarded Packets 3
1.4.1 Overview 3
1.4.2 Restrictions and Guidelines 3
1.4.3 Procedure 4
1.5 Monitoring 5
1.6 Configuration Examples 5
1.6.1 Capturing Packets Destined for the CPU 5
1.6.2 Capturing Forwarded Packets 7

1 Configuring Packet Capture

1.1 Introduction

1.1.1 Overview

Packet capture (PCAP) captures packets sent or received by a network device to analyze network device faults. With network technology development, network deployment becomes more and more complex. The PCAP function configured on a device helps locate faults remotely, facilitating network maintenance.

After packets that meet a capture rule are captured, a network administrator can view the captured packets in the command line interface (CLI) or save them to the storage medium of the device using a *.pcap file and then download the file to a local directory and use software to view the packets. The PCAP function simplifies deployment of the packet analyzer and network monitoring devices and can analyze the causes of network faults, improving network maintenance efficiency and reducing network maintenance costs.

1.1.2 Principles

The device supports the following packet capture methods:

1. Capturing packets destined for the CPU

Configure the packet capture points, specify the packet capture location, direction, matching rule, captured packet quantity, and name of the file for saving captured packets, and analyze different types of captured packets destined for the central processing unit (CPU) to locate the causes of device faults and handle the faults in time to ensure proper network running.

2. Capturing forwarded packets

Associate a packet capture rule with an access control list (ACL) to filter packets and analyze specified packets captured to locate the cause for abnormal data forwarding in the network environment and handle the fault in time to ensure proper network running.

1.2 Configuration Task Summary

PCAP configuration includes the following tasks:

- Capturing Packets Destined for the CPU
- Capturing Forwarded Packets

1.3 Capturing Packets Destined for the CPU

1.3.1 Overview

When a type of packets fails to be sent or received during device running, configure rules to capture and analyze packets destined for the CPU to help locate the fault.

1.3.2 Restrictions and Guidelines

- To capture different types of packets simultaneously, you can define up to eight packet capture points with different names on the device.
- You can define packet capture rules to capture packets that meet requirements. Packet capture rules include:
  - Specifying the location for capturing packets. You can specify a physical port or the system CPU as the packet capture location. Each packet capture point can define only one packet capture location. However, different packet capture points can have the same packet capture location.
  - Specifying the direction for capturing packets. You can specify the outbound direction, inbound direction, or both directions to capture packets.
  - Matching packet feature information to capture specific packets. Packet feature information includes the source media access control (MAC) address, destination MAC address, layer-2 protocol type, source IP address, destination IP address, layer-3 protocol type, and Transmission Control Protocol (TCP)/User Datagram Protocol (UDP) port.
- Specify the name of the file used to save captured packets. Save captured packets to the storage medium of the device using a *.pcap file.
- Specify a condition for automatic stopping of packet capturing, such as the quantity of capture packets, size of the file used to save captured packets, packet capture duration, and less than 1/4 idle space of the PCAP file storage path. After the PCAP function is enabled and any of the conditions for automatic stopping of packet capturing is met, the device automatically stops packet capturing. After the **packet capture stop** command is run, the device stops packet capturing immediately regardless of whether the condition for automatic stopping of packet capturing is met.
- Multiple packet capture rules with different names can be defined on the device simultaneously. A packet capture rule takes effect only when it is referenced by a packet capture point.
- To delete a packet capture rule, you need to delete all the packet capture points referencing the rule.
- The PCAP function can be enabled only when the CPU usage of the device is less than or equal to 70%.
- The PCAP configuration cannot be saved.
- The PCAP configuration cannot be synchronized in VSU mode. Packet capture rules configured on the master device will be lost after a master/slave switchover.
- Rules for capturing packets destined for the CPU and forwarded packets cannot be configured simultaneously. For example, if a rule for capturing forwarded packets is configured, a rule for capturing packets destined for the CPU can be configured only after the previous rule is cleared.
- A switch virtual interface (SVI) cannot be configured as the location for capturing packets destined for the CPU.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure a packet capture rule.

packet capture rule *rule-name* **filter** [**dst-mac** *mac-address*] [**etype** *type*] [**icmpv4_code** *icmpv4-code*] [**icmpv4_type** *icmpv4-type*] [**icmpv6_code** *icmpv6-code*] [**icmpv6_type** *icmpv6-type*] [**ipv4_dip** { *ipv4-address wildcard*| **host** *ipv4-address* }] [**ipv4_dport eq** *port-number*] [**ipv4_sip** { *ipv4-address wildcard* | **host** *ipv4-address* }] [**ipv4_sport eq** *port-number*] [**ipv6_dip** *ipv6-address*] [**ipv6_dport eq** *port-number*] [**ipv6_sip** *ipv6-address*] [**ipv6_sport eq** *port-number*] [**src-mac** *mac-address*] [**v4_protocol** *protocl*] [**v6_protocol** *protocol*] [{ **acl** *acl-name* | **acl** *acl-number* } **interface** *interface-type interface-number* **in**]

No packet capture rule is configured by default.

(3) Configure a packet capture point.

packet capture point *name* **rule** *rule-name* **location** { **control-plane** | **interface** *interface-type interface-number* } { **in** | **out** | **both** }

No packet capture point is configured by default.

(4) Configure the name of the file used to save captured packets.

packet capture file { **flash:** | **tmp: | usb0:** } *filename* [**buffer-size** *size*] [**packet-num** *number*] [**timeout** *timeout*]

No file used to save captured packets is configured by default.

(5) Enable the PCAP function.

packet capture start

The PCAP function is disabled by default.

(6) (Optional) Manually disable the PCAP function.

packet capture stop

Packet capturing stops automatically when any of the conditions for stopping packet capturing is met.

(7) (Optional) Clear configured packet capture rules.

cleanup packet capture point *capture-point-name*

cleanup packet capture rule *rule-name*

1.4 Capturing Forwarded Packets

1.4.1 Overview

Associate a packet capture rule with an ACL to filter packets and analyze specified packets captured to locate the cause for abnormal data forwarding in the network environment and handle the fault in time to ensure proper network running.

1.4.2 Restrictions and Guidelines

- No packet capture point needs to be specified, and 7-tuple matching information is not supported.
- Specify the name of the file used to save captured packets. Save captured packets to the storage medium of the device using a *.pcap file.
- Specify a condition for automatic stopping of packet capturing, such as the quantity of capture packets, size of the file used to save captured packets, packet capture duration, and less than 1/4 idle space of the PCAP file storage path. After the PCAP function is enabled and any of the conditions for automatic stopping of

packet capturing is met, the device automatically stops packet capturing. After the **packet capture stop** command is run, the device stops packet capturing immediately regardless of whether the condition for automatic stopping of packet capturing is met.

- Multiple packet capture rules with different names can be defined on the device simultaneously.
- The PCAP function can be enabled only when the CPU usage of the device is less than or equal to 70%.
- To configure a rule for capturing forwarded packets, configure an ACL rule first. In addition, the permit rule needs to be configured for the ACL rule. Otherwise, packets cannot be captured.
- The PCAP configuration cannot be saved.
- The PCAP configuration cannot be synchronized in VSU mode. Packet capture rules configured on the master device will be lost after a master/slave switchover.
- Rules for capturing packets destined for the CPU and forwarded packets cannot be configured simultaneously. For example, if a rule for capturing packets destined for the CPU is configured, a rule for capturing forwarded packets can be configured only after the previous rule is cleared.
- Configuration requirements on the interface for capturing forwarded packets:
  - A physical interface can be configured as the location for capturing packets.
  - An MGMT interface or virtual switch link (VSL) interface cannot be configured as the location for capturing packets.
  - In a VSU environment, non-master device interfaces cannot be configured as the location for capturing packets. An aggregation port of the master device can be configured as the location for capturing packets even if a member interface of the aggregation port is not a master device interface.

1.4.3 Procedure

(1) Create an ACL and configure rules.

a Enter the privileged EXEC mode.

privilege

b Enter the global configuration mode.

config-system terminal

c Configure an ACL rule.

access-control-list *acl-number* { **permit** | **deny** } { *source-ipv4-address source-ipv4-wildcard* | **host** *source-ipv4-address* | **any** } [**time-range** *time-range-name*]

No ACL rule is configured by default.

d Return to the privileged EXEC mode.

end

(2) Configure a packet capture rule.

packet capture rule *rule-name* **filter** { **acl** { *acl-name* | *acl-number* } [**interface** *interface-type interface-number*] { **in** | **out** | **both** } }

No packet capture rule is configured by default.

(3) Configure the name of the file used to save captured packets.

packet capture file { **flash:** | **tmp: | usb0:** } *filename* [**buffer-size** *size*] [**packet-num** *number*] [**timeout** *timeout*]

No file used to save captured packets is configured by default.

(4) Enable the PCAP function.

packet capture start

The PCAP function is disabled by default.

(5) (Optional) Manually disable the PCAP function.

packet capture stop

Packet capturing stops automatically when any of the conditions for stopping packet capturing is met.

(6) (Optional) Clear configured packet capture rules.

cleanup packet capture rule *rule-name*

1.5 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Table 1-1 Monitoring

Purpose	Command
Displays packet capture information.	**display packet capture status**

1.6 Configuration Examples

1.6.1 Capturing Packets Destined for the CPU

1. Requirements

A TCP connection to GigabitEthernet 0/1 of the device fails. Capture TCP packets over the interface for 1 minute to analyze the cause.

2. Procedure

Configure a packet capture rule named **tcp** to match packets whose layer-2 protocol type is IPv4 and layer-3 protocol type is TCP.

```
Device> privilege
Device# packet capture rule tcp filter etype ip v4_protocol tcp
```

Configure a packet capture point named **tcppoint**, associate it with packet capture rule **tcp**, and capture packets in the inbound and outbound directions of GigabitEthernet 0/1.

```
Device# packet capture point tcppoint rule tcp location interface gigabitethernet 0/1 both
```

Save captured packets to the **tcp.pcap** file in the flash file system. Condition for automatic stopping of packet capturing: The size of the **tcp.pcap** file reaches 2 MB, the number of captured packets reaches 1,500, or the packet capture duration reaches 1 minute.

```
Device# packet capture file flash:tcp.pcap buffer-size 2 packet-num 1500 timeout 1
```

Enable the PCAP function.

```
Device# packet capture start
```

3. Verification

Run the **display packet capture status** command to display the PCAP configuration.

```
Device# display packet capture status

Capture rules:
  Capture rules tcp:
    Etype: 0x0800
    Protocol: 0x6

Capture points:
  Capture point tcppoint:
    Capture rules: tcp
    Location: Gi0/1
    Direction: all
    Packets captured (all): 0

Capture file:
  Filename: flash:tcp.pcap
  Buffer size: 2 (MB)

Capture Statistic:
  Status: running
  Start time: 2021-4-12 10:00:03 AM
  Timeout: 1 (minutes)
  Packets limit: 1500
  Write file packet count: 0
```

Run the **display packet capture status** command to display the packet capture status after packet capturing ends.

```
Device# display packet capture status

Capture rules:
  Capture rules tcp:
    Etype: 0x0800
    Protocol: 0x6

Capture points:
  Capture point tcppoint:
    Capture rules: tcp
    Location: Gi0/1
    Direction: all
    Packets captured (all): 60
```

```
Capture file:
  Filename: flash:tcp.pcap
  Buffer size: 2 (MB)

Capture Statistic:
  Status: stopped
  Stopped reason: Timeout
  Start time: 2021-4-12 10:00:03 AM
  End time: 2021-4-12 10:01:03
  Timeout: 1 (minutes)
  Packets limit: 1500
  Write file packet count: 60
```

1.6.2 Capturing Forwarded Packets

1. Requirements

Traffic over GigabitEthernet 0/1 of the device is abnormal. Capture forwarded packets over the interface for 10 minutes to confirm whether data streams are forwarded to the device.

2. Procedure

Run the packet capture command to save traffic over GigabitEthernet 0/1 that matches the ACL rule to the **tcp.pcap** file and upload the file to a PC to display the captured data.

Configure an ACL rule.

```
Device> privilege
Device(cfg)# config-system terminal
Device(cfg)# access-control-list 100 permit tcp host 1.1.1.1 any
Device(cfg)# exit
```

Configure a packet capture rule named **rule1**, associate it with the ACL rule, and capture packets in the inbound direction of GigabitEthernet 0/1.

```
Device# packet capture rule rule1 filter acl 100 interface gigabitethernet 0/1 in
```

Save captured packets to the **tcp.pcap** file in the flash file system. Condition for automatic stopping of packet capturing: The packet capture duration reaches 10 minutes.

```
Device# packet capture file flash:tcp.pcap timeout 10
```

Enable the PCAP function.

```
Device# packet capture start
```

3. Verification

Run the **display packet capture status** command to display the packet capture status.

```
Device# display packet capture status

Capture rules:
  Capture rules rule1:
```

```
acl_id: 100
acl_name:
acl_ifx: GigabitEthernet 0/1
Direction: in

Capture file:
   Filename: flash:tcp.pcap
   Buffer size: 2 (MB)

Capture Statistic:
   Status: running
   Start time: 2020-10-15 7:48:41 PM
   Timeout: 10 (minutes)
   Packets limit: 1024
   Write file packet count: 10
```

Contents

1 Configuring sFlow....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.1.3 Protocols and Standards....3
1.2 Configuration Task Summary....4
1.3 Configuring sFlow....4
1.3.1 Configuring Basic Functions of sFlow....4
1.3.2 Configuring Optional Parameters of sFlow....6
1.4 Monitoring....7
1.5 Configuration Examples....8
1.5.1 Configuring Basic Functions of sFlow....8

1 Configuring sFlow

1.1 Introduction

1.1.1 Overview

Sampled flow (sFlow) is a network monitoring technology based on message sampling. It is mainly used for traffic statistics analysis when the network traffic is huge.

sFlow has the following advantages:

- Accurate: sFlow supports accurate monitoring of traffic on a Gigabit network or a network with higher bandwidth.
- Scalable: One sFlow Collector can monitor thousands of sFlow Agents, and it has high scalability.
- Low cost: sFlow Agent is embedded in a network device, and its cost is low.

1.1.2 Principles

The sFlow system consists of an sFlow Agent embedded in a device and a remote sFlow Collector. The sFlow Agent obtains interface statistics and data from the network device by using specific sampling technologies, encapsulates the information into sFlow packets, and sends them to the sFlow Collector. The sFlow Collector analyzes sFlow packets and displays the analysis results.

- **sFlow Agent**

  An sFlow Agent is embedded in a network device. Generally, one network device serves as an sFlow Agent.

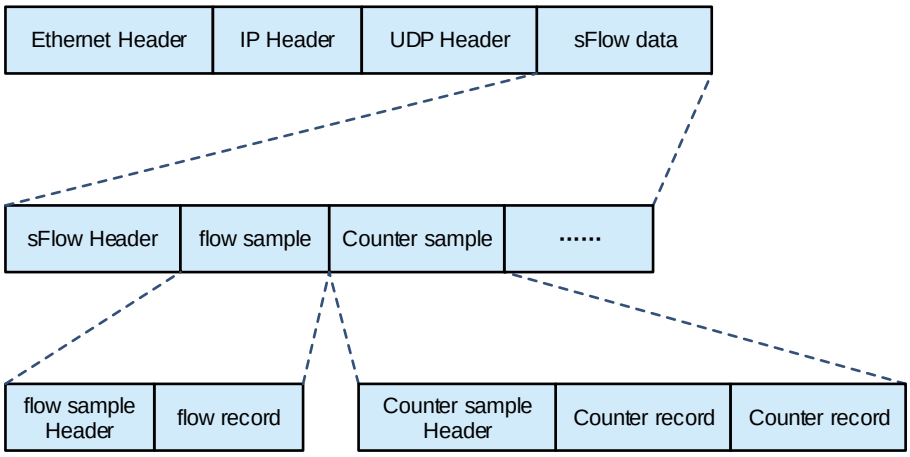
- **sFlow Collector**

  An sFlow Collector can be a PC or server that receives and analyzes sFlow packets from the sFlow Agent. A PC or server installed with the software used for sFlow packet analysis can be regarded as an sFlow Collector.

  The sFlow Agent performs flow sampling and counter sampling, encapsulates sampled data into sFlow packets, and sends them to the sFlow Collector.

  sFlow packets are encapsulated in UDP. Figure 1-1 shows the sFlow packet format.

Figure 1-1 sFlow Packet Format



Note

One sFlow packet may contain one or more flow samples and counter samples.

Figure 1-2 sFlow Header

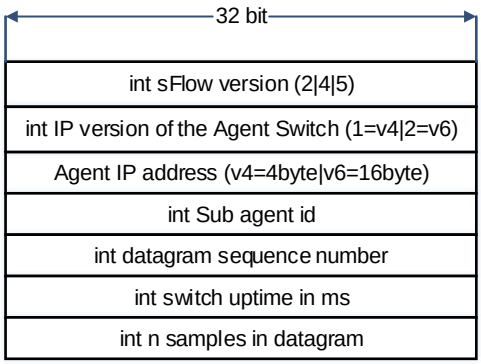


Table 1-1 sFlow Header Description

Field	Description
sFlow version	sFlow version. V2, V4, and V5 are available. Currently, the product supports V5 only.
IP version of the agent/switch	IP address version of the sFlow Agent
Agent IP address	IP address of the sFlow Agent
Sub agent id	Sub agent ID
Datagram sequence number	Serial number of an sFlow packet
Switch uptime	Duration from device startup to the current time

Field	Description
n samples in datagram	Number of samples in an sFlow packet. One sFlow packet contains one or more flow samples and counter samples.

1. Flow sampling

Based on the sampling rate configured for the specified interface, the sFlow Agent device performs flow sampling on packets through this interface, including copying the packet header, extracting the Ethernet header and IP header of each packet, and obtaining the route information of each packet. Then, the sFlow Agent encapsulates the flow sampling result into sFlow packets and sends them to the sFlow Collector for analysis.

Table 1-2 Description of Fields in the Flow Record Part of the sFlow Packet

Field	Description
Raw packet	All or part of the header of the original packet (the extracted length depends on the configuration) is extracted and encapsulated into an sFlow packet and then sent to the Collector.
Extended Switch Data	For an Ethernet packet, the Ethernet header is parsed and encapsulated into an sFlow packet and then sent to the Collector.
Extended VNI Data	VNI ID of a VXLAN packet

2. Counter sampling

In counter sampling, the sFlow Agent periodically obtains the statistics and CPU usage from a specified interface. The sFlow Agent periodically performs interface polling, obtains the statistics from an interface where the counter sampling timer expires, encapsulates the statistics into an sFlow packet, and sends it to the sFlow Collector for analysis.

Table 1-3 Description of Fields in the Counter Sampling Result

Field	Description
Generic Interface Counters	General interface statistics, including basic interface information and general interface traffic statistics
Ethernet Interface Counters	Ethernet-related traffic statistics of an Ethernet interface

1.1.3 Protocols and Standards

- sFlow Version 5: sFlow V5
- RFC 1014: XDR: External Data Representation standard

1.2 Configuration Task Summary

sFlow configuration includes the following tasks:

(1) Configuring Basic Functions of sFlow

(2) (Optional) Configuring Optional Parameters of sFlow

1.3 Configuring sFlow

1.3.1 Configuring Basic Functions of sFlow

1. Overview

- sFlow Agent and sFlow Collector can communicate with each other.
- Traffic flowing through the interface are sampled based on the default sampling rate and sent to the sFlow Collector for analysis.
- Statistics of the interface are periodically sent to the sFlow Collector based on the default sampling interval for analysis.

2. Restrictions and Guidelines

- The forwarding performance of an interface may be affected after flow sampling is enabled.
- To enable the sFlow Collector to analyze the flow sampling results, you must configure the IP address of the sFlow Collector on the sFlow Agent device.
- The sFlow Agent address must be a valid address. That is, the sFlow Agent address cannot be a multicast or broadcast address. It is recommended that the IP address of the sFlow Agent device be used.
- When the **vrf** parameter is configured, the corresponding VRF instance must exist. If you configure a VRF instance for an sFlow Collector address and later remove this VRF instance, the sFlow Collector address will be removed as well.
- When the **oob** parameter is configured, a packet is sent to the sFlow Collector through the management port.
- You can enable flow sampling on an Ethernet interface or a link aggregation.
- **sflow enable**
  - If the direction is not specified, flow sampling is enabled in both the inbound and outbound directions.
  - Counter sampling is enabled when flow sampling is enabled in any direction of an interface.
  - The ACL is supported only in the inbound direction. Configure the ACL before configuring this command.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the sFlow Agent address.

sflow agent { **address** { *ipv4-address* | **ipv6** *ipv6-address* } | **interface** [**ipv6**] *interface-type interface-number* }

No sFlow Agent address is configured by default.

(4) Configure the sFlow Collector address.

sflow collector *collector-id* **destination** { *ipv4-address* | **ipv6** *ipv6-address* } *udp-port-number* [**vrf** *vrf-name* | **oob** [**via mgmt** *mgmt-name*]] [**description** *collector-description*] [**source** { **address** *ip-address* | **interface** *interface-type interface-number* }]

No sFlow Collector address is configured by default.

(5) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 2 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 2 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

- Enter the Layer 3 Ethernet subinterface configuration mode.

  interface *ethernet-type interface-number.subnumber*

- Enter the Layer 3 aggregated subinterface configuration mode.

  interface aggregateport *interface-number.subnumber*

(6) Configure the ID of the sFlow Collector for flow sampling. Configure at least one of the following task.

- Configure the ID of the sFlow Collector for flow sampling.
  - **sflow flow collector** *collector-id*

  No ID of the sFlow Collector is configured for flow sampling by default.

- Configure the ID of the sFlow Collector for counter sampling.
  - **sflow counter collector** *collector-id*

  No ID of the sFlow Collector is configured for counter sampling by default.

(7) Enable the sampling functions.

sflow enable [{ **ingress** } [{ **ingress** | **egress** } [**acl** { acl-name | acl-number }]]

The sFlow sampling function is disabled by default.

1.3.2 Configuring Optional Parameters of sFlow

1. Overview

You can adjust the data sampling accuracy by modifying relevant parameter attributes of sFlow.

2. Restrictions and Guidelines

- You can run the **sflow collector** command to configure the payload size of an sFlow packet, excluding the Ethernet header, IP header, and UDP header. One or more flow samples and counter samples can be encapsulated to an sFlow packet. When the maximum sFlow packet size is configured, different numbers of sFlow packets may be output after the same number of flow samples or counter samples are processed. If the maximum size is greater than MTU, the output sFlow packet is segmented.
- The **sflow sampling-rate** command is used to configure the global flow sampling rate. This configuration applies to all interfaces. The flow sampling rate may affect the sFlow sampling accuracy. A lower sampling rate indicates a higher accuracy and a larger CPU consumption, which may affect the forwarding performance of the interface.
- You can run the **sflow flow max-header** command to modify the configuration of packets sent to the sFlow Collector. For example, if you are concerned about the IP header of a packet, set the length to 56 bytes. During flow sampling, the first 56 bytes of a sample packet are encapsulated to the sFlow packet.
- This command is used to configure the maximum number of bytes that can be copied from the header of the original packet. The copied content is recorded in the generated sample. The protocol requires byte alignment during packet encapsulation, that is, the actual length of a sent packet is a multiple of 4. Therefore, the length of a collected packet may exceed the configured length. For example, when the maximum length is set to any of 21, 22, 23, and 24, the actual output packet length is 24.
- The **sflow counter interval** command is used to configure the global sFlow counter sampling interval. This configuration applies to all interfaces.
- The **sflow source** command is used to configure the source IP address of output packets. The default source address of output sFlow packets is the local device IP address which is used to ping the destination IP address. If the source interface is specified, the primary address (or the first global IPv6 address if any) of the interface is the source IP address of output packets. If the source interface is not specified, the default source address is used.
- Run the **sflow enhance-mode enable** command to enable enhanced sFlow sampling. In the enhanced mode, sFlow packets are sent by the chip to the Collector server, which caters to requirements for high performance. The enhanced mode applies to the only ingress direction. In the enhanced mode, the sampling packets are not sent from the management port. In the enhanced mode, the length of the chip packets and of the copy packet header is fixed. The globally configured maximum length of the chip packets and of the copy packet header applies to only the software sFlow. In the enhanced mode, the sysuptime encrypted in the sFlow packet header uses the IEEE 1588 Time Of Day (TOD) clock, which is not compliant with the standard protocol but does not affect hardware sFlow packet parsing. In the enhanced mode, the ingress sampling on the sub-interface is not supported. When the enhanced mode is enabled on the line card of a modular device, ingress sampling on the sub-interface of the aggregate interface does not conflict with the enhanced mode based on the line card. If the member port of the aggregate interface is on the line card enabled with the enhanced mode, the member port uses the software sFlow actually. In the enhanced mode, sFlow supports only sample_type 1, namely, the sFlow packet does not contain the VLAN ID of the original packet. When the **sample pool** field in the sFlow packet header is 0, the server software (sflowtrend)

fails to parse packets.

- Run the **sflow tx-use-loop disable** command to configure the device to disable loopback mode for sFlow egress sampling. With loopback mode disabled, egress sampling occupies all mirroring resources, leading to conflicts with other mirroring resource-consuming services such as SPAN and IPFIX. With loopback mode disabled, ingress ACL sampling, ingress snoop sampling, or sFlow enhanced mode is not supported.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the maximum size of the output sFlow packet.

sflow collector *collector-id* **max-datagram-size** *datagram-size*

By default, the maximum size of the output sFlow packet is **1400**.

(4) (Optional) Configure the sFlow flow sampling rate.

sflow sampling-rate *sampling-rate*

The default sFlow flow sampling rate is **8192**.

(5) (Optional) Configure the maximum length of the packet header copied during flow sampling.

sflow flow max-header *sampling-length*

By default, the maximum length of the packet header copied during sFlow flow sampling is **64**.

(6) (Optional) Configure the sampling interval.

sflow counter interval *sampling-interval-time*

The interface where counter sampling is enabled sends the interface statistics to the sFlow Collector at the sampling interval.

(7) (Optional) Configure the source address of output sFlow packets.

sflow source { **address** { *ipv4-address* | **ipv6** *ipv6-address* } | **interface** [**ipv6**] *interface-type interface-number* }

By default, the source address of output sFlow packets is the local device IP address which is used to ping the destination IP address.

(8) (Optional) Configure enhanced sFlow sampling.

sflow enhance-mode enable [**device** *devid*] [**slot** *slotid*]

1.4 Enhanced sFlow sampling is disabled by reset. Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Table 1-4 sFlow monitoring

Command	Purpose
display sflow	Displays the sFlow configuration.

Command	Purpose
display sflow stats	Displays sFlow packet statistics.

1.5 Configuration Examples

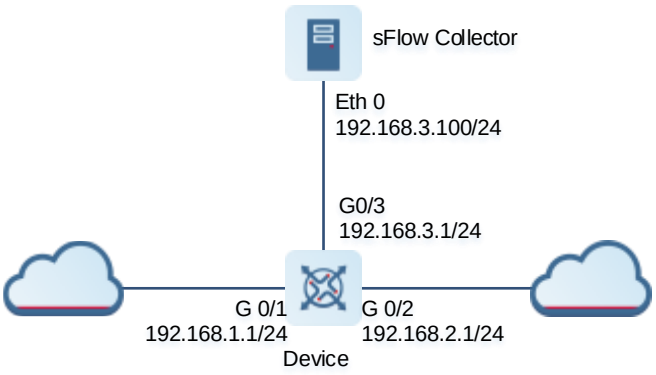
1.5.1 Configuring Basic Functions of sFlow

1. Requirements

Start the device that serves as the sFlow Agent, enable flow sampling and counter sampling on port GigabitEthernet 0/1, monitor the traffic in the 192.168.1.0 network segment, encapsulate the sampling result into sFlow packets at regular intervals or when the buffer is full, and send these packets to the sFlow Collector for traffic analysis.

2. Topology

Figure 1-3 Configuring Basic Functions of sFlow



3. Notes

- Configure the Layer 3 network reachable between the sFlow Agent and the sFlow Collector.
- Configure the sFlow function.

4. Procedure

Configure the Layer 3 network reachable between the sFlow Agent and the sFlow Collector.

```
Device> privilege
Device#config-system terminal
Device(cfg)# interface GigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.1 255.255.255.0
Device(cfg)# interface GigabitEthernet 0/2
Device(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.1 255.255.255.0
Device(cfg)# interface GigabitEthernet 0/3
Device(cfg-if-GigabitEthernet 0/1)# ip address 192.168.3.1 255.255.255.0
```

Configure 192.168.1.1 as the sFlow Agent address.

```
Device(cfg)# sflow agent address 192.168.1.1
```

Configure 192.168.3.100 as the address of sFlow Collector 1, and 6343 as the port number.

```
Device(cfg)# sflow collector 1 destination 192.168.3.100 6343
```

Configure the interface GigabitEthernet 0/1 to output flow samples and counter samples to sFlow Collector 1, and enable sFlow sampling on this interface.

```
Device(cfg)# interface GigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# sflow flow collector 1
Device(cfg-if-GigabitEthernet 0/1)# sflow counter collector 1
Device(cfg-if-GigabitEthernet 0/1)# sflow enable
```

5. Verification

Run the **display sflow** command to check whether the command output is consistent with the configuration.

```
Device# display sflow
sFlow datagram version 5
Global information:
Agent IP: 192.168.1.1
Source IP[1]: 192.168.1.1(used: 192.168.1.1)
Source IP[2]: 192.168.1.2(used: 0.0.0.0)

sflow counter interval:30
sflow flow max-header:64
sflow sampling-rate:8192
Collector information:
ID      IP                                      Port Size VRF                                    Description
1       192.168.3.100                           6343 1400
2       NULL                                    0    1400
Port information
Interface                 CID   FID   Enable    By-snoop    AclId[1]    AclName[1] AclId[2]    AclName[2]
Te0/1                     1     1     B         N           0                      0
Enhance mode:
  disable
Sflow egress sample use loop mode: Enable
```

Figure 1-4 shows the Top N tab of sFlowTrend, which displays the flow sampling result and the source IP addresses with the Top 5 traffic. The inbound traffic 450 kpps, and the outbound traffic is also 450 kpps, which is consistent with the actual traffic.

Figure 1-5 is the Counters tab of sFlowTrend, which displays the counter sampling result. The inbound traffic is 450 kpps, and the outbound traffic is also 450 kpps. All packets are unicast packets.

Figure 1-4 Flow Sampling Results

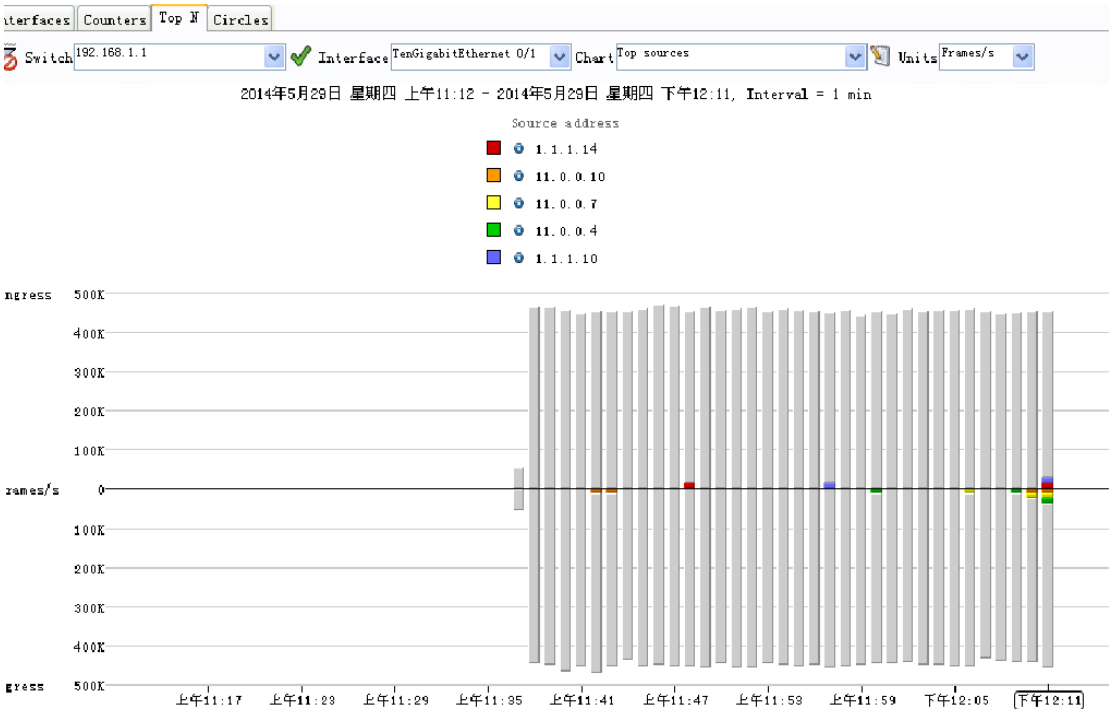
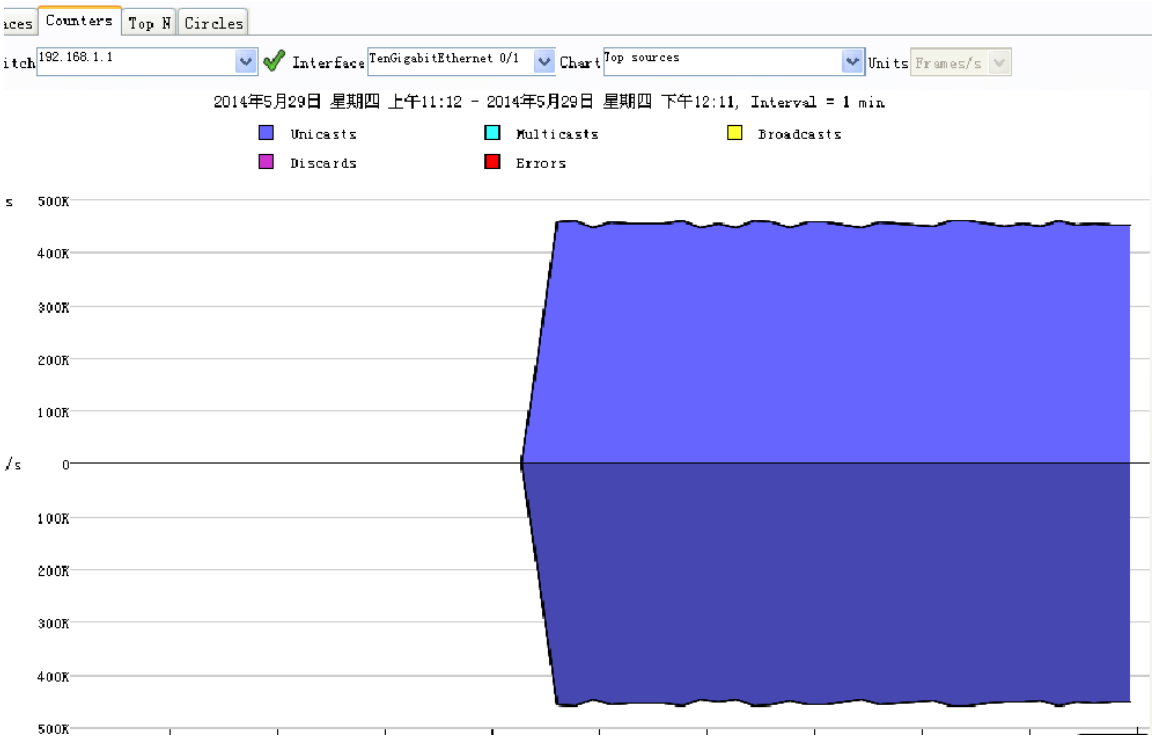


Figure 1-5 Counter Sampling Results



6. Configuration Files

Device configuration file

```
Hostname Device
!
sflow agent address 192.168.1.1
sflow collector 1 destination 192.168.3.100 6343
!
interface GigabitEthernet 0/1
 ip address 192.168.1.1 255.255.255.0
 sflow counter collector 1
 sflow flow collector 1
 sflow enable
!
interface GigabitEthernet 0/2
 ip address 192.168.2.1 255.255.255.0
!
interface GigabitEthernet 0/3
 ip address 192.168.3.1 255.255.255.0
!
end
```

Contents

1 Configuring IPFIX 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Basic Concept 1
1.1.3 IPFIX Sampling 2
1.1.4 IPFIX Flow 2
1.1.5 IPFIX Flow Table 3
1.1.6 Principle 3
1.1.7 IPFIX Flow Statistics 4
1.1.8 IPFIX Flow Aging 5
1.1.9 IPFIX Flow Export 6
1.1.10 Flexible IPFIX Flow Sampling 6
1.1.11 Protocols and Standards 6
1.2 Configuration Task Summary 7
1.3 Configuring IPFIX Flow Statistics 7
1.3.1 Overview 7
1.3.2 Restrictions and Guidelines 7
1.3.3 Procedure 7
1.4 Configuring IPFIX Flow Export 9
1.4.1 Overview 9
1.4.2 Restrictions and Guidelines 9
1.4.3 Procedure 9

1.5 Configuring Parameters of the IPFIX Flow Table 10
1.5.1 Overview 10
1.5.2 Restrictions and Guidelines 11
1.5.3 Procedure 11
1.6 Configuring Flexible Sampling 12
1.6.1 Overview 12
1.6.2 Restrictions and Guidelines 12
1.6.3 Procedure 12
1.7 Monitoring 13
1.8 Configuration Examples 13
1.8.1 Configuring the IPv4 IPFIX Function 13
1.8.2 Configuring the IPv6 IPFIX Function 17

1 Configuring IPFIX

Note

This function is not supported by the S6980-64QC, S6520-64CQ, S6580-48CQ8QC, or S6580-24DC8QC.

1.1 Introduction

1.1.1 Overview

With the rapid development of network technology, stream-based technologies are highly valubale in QoS policy configuration, application deployment, and capacity planning. But there is no standard format for flow export.

IP Flow Information Export (IPFIX) is a standard protocol published by IETF for statistics and analysis of network traffic. It standardizes traffic monitoring, uses a single and unified model, and simplifies export structure so that network equipment and servers from different manufacturers can communicate with each other. Therefore, IPFIX is widely applied to network monitoring and analysis, security analysis and attack check, and traffic-based billing.

1.1.2 Basic Concept

1. IPFIX System

The IPFIX system consists of an Export device, a collector, and an analyzer, as shown in Figure 1-1.

- Export: A network traffic exporter is responsible for analyzing and processing network traffic, extracting eligible traffic statistics, and exporting the statistics to a collector.
- Collector: A network traffic collector is a software specializing in processing traffic records. A collector is responsible for parsing packets received from the exporter and collecting statistics into a database that can be parsed by an analyzer. A collector can receive data from multiple exporters.
- Analyzer: A network traffic analyzer processes the statistical data extracted from a collector and provides a basis for various services. Usually, the analyzer has a graphical user interface to display the analysis results, which is convenient for users to obtain and analyze collected data. An analyzer can receive data from multiple collectors.

In practice, an exporter is a network device, such as a switch, a router and so on. A collector and an analyzer are usually integrated on an IPFIX server, as shown in Figure 1-2.

Figure 1-1 IPFIX System

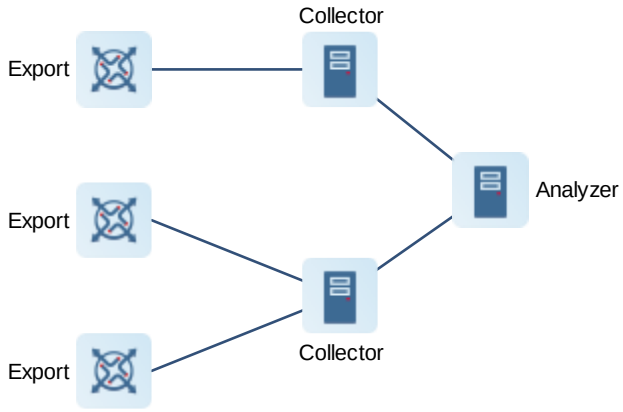
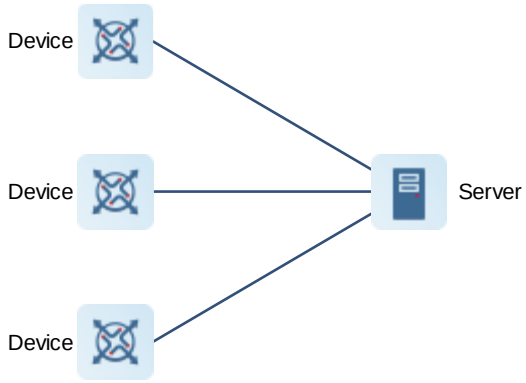


Figure 1-2 IPFIX Networking



2. IPFIX Sampling

When analyzing the traffic of an interface, IPFIX sets the sampling rate on a specified interface to sample the interface traffic and analyze the sample packets.

IPFIX sampling reduces the statistical analysis of packets and the impact on device performance.

You can enable IPFIX to sample the traffic in the inbound or outbound direction of the interface or the traffic in both directions.

3. IPFIX Flow

IPFIX is a flow-based protocol. After sampling traffic, IPFIX processes sampled packets and produces a flow based on the keywords in the packets. The flow is called as an IPFIX flow.

Note

The IPFIX flow here are also called the IPFIX basic flow or primary flow.

A packet consists of a series of attributes, such as the IP address and TCP port ID. Continuous packets that have the same attributes and flow through a detection point within a period of time are called a packet flow. Attributes used to determine whether packets belong to the same flow are called key attributes of the flow. If the key attributes of packets are the same, these packets belong to the same flow. A flow is determined by the common combination of the following key fields:

- Source IP address
- Destination IP address
- Source port ID
- Destination port ID
- L3 protocol type
- ToS
- Packet sampling interface index

These seven key fields determine a unique flow. Packets with the same seven fields belong to the same flow. If a packet has one key field that is different from that of another packet, they belong to different flows. A flow also contains other statistical fields, such as the number of packets, next-hop IP address, and total number of bytes in the flow. Flow information also includes:

- Flow start time
- Flow end time
- Total number of packets
- Total number of bytes
- Flow ingress interface index (for sampling in the ingress direction)
- Flow egress interface index (for sampling in the egress direction)

4. IPFIX Flow Table

A flow table stores measured IP flow information. For an IP packet, the device first looks up the corresponding flow record in the main cache based on key fields of its flow. If the flow record is not configured, the device creates a new flow record, and records the flow information. If the flow record is already configured, the device updates the flow statistics in the flow record, such as the number of packets and the number of bytes in the flow.

Note

The IPFIX flow table here is also called the IPFIX basic flow table or primary flow table, and used to store basic flows.

1.1.3 Principles

The basic functions of an IPFIX device include:

- Flow statistics: Obtains attributes from transmitted packets, and looks up the flow table based on the seven key fields. If no matched record is found in the flow table, the packet belongs to a new flow, and a flow record is created to record the flow statistics. If the packet belongs to an existing flow, the flow statistics is updated.
- Exporting flow information to the server: Sends the flow statistics to the server for analysis at the appropriate time.

Note

The IPFIX specification defines only how an IPFIX device exports flow records, but not how to save and analyze flow records.

1. IPFIX Flow Statistics

IPFIX flow statistics is used to build flow tables, extract information in packets, and count traffic in the network. Figure 1-3 shows the IPFIX flow statistics process.

(1) A network device forwards packets normally. A packet is input from one interface and output from another interface after being forwarded.

(2) The IPFIX module of the device extracts the packet-related attributes, such as the IP address and packet length, from the packet header.

(3) It also adds a time stamp to the packet to record the start time, end time, and duration of the flow.

(4) The IPFIX module samples packets according to the sampling rate.

(5) IPFIX filters the packets as needed. IPFIX filters the packets by ACLs to process the specified packets only.

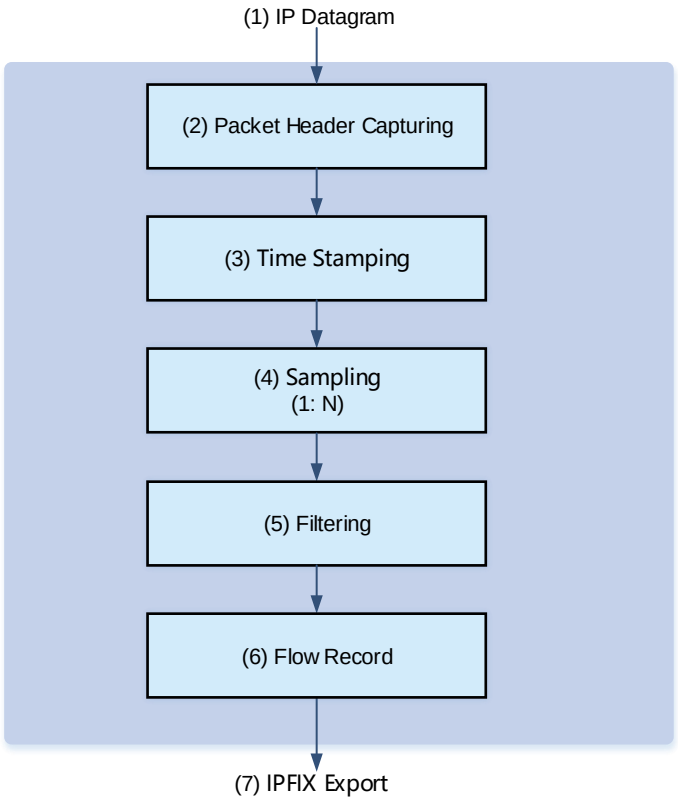
(6) The device save flow information. If the flow record is not configured, the device adds a flow record. If the flow record already is configured, the device updates the flow information, such as the number of packets and the number of bytes.

(7) When the flow ages, the device deletes the flow record and exports the flow record information to the IPFIX server.

Note

Step 4 and step 5 can be reversed, or cycled multiple times. The device can filter packets first and then sample the packets. Or the device can sample the packets first and then filter the packets. After that, the device can also sample the filtered packets again.

Figure 1-3 Flowchart of IPFIX Flow Statistics



2. IPFIX Flow Aging

Flow records are deleted from the flow table under certain circumstances. Before deletion, the statistical information is exported to the server. This process is called flow aging. Flow aging prevents the flow table from increasing indefinitely, while ensuring that the IPFIX server can obtain flow statistics. An IPFIX Flow has the following three types of aging.

- Inactive flow aging
  - When no packets of a flow is detected within a certain period of time (inactive aging time), the flow should be aged and exported. The device exports the flow information to the server. This is called inactive flow aging.
- Active flow aging
  - When packets of a flow is continuously collected for a long time, its information cannot be recorded indefinitely for saving resources. You can configure a period (active aging time), after which the flow should be aged and the device exports the flow information to the server. This is called active flow aging.
- Mandatory flow aging
  - When the IPFIX mode changes or the template configuration changes, a new template is used to analyze and collect packets and create new flow records, and the existing flow records are forced to age.
  - When the flow forced aging command is executed, related flow records are forced to age. The IPFIX

flow are not automatically aged because it does not meet the aging conditions, but you want to check the latest flow statistics. Or the flow cannot be automatically aged due to an exception of the flow statistics function. In this two cases, you can use the command line to force the flow to age.

Note

When a flow record is aged and deleted, and a packet with the same seven key fields as the aged flow is immediately received, no matched flow record exists in the flow table. Therefore, the packet is processed as a new flow, and a new flow record is created.

3. IPFIX Flow Export

IPFIX flow export means that aged flow records are exported to the IPFIX server for analysis.

IPFIX uses a template to transmit flow records. A template defines the format of data records, including a series of bit fields. Each bit field includes the data type and data length of the bit field.

- When an aged flow record needs to be exported to the server, the information of the flow record is encapsulated into a UDP or TCP packet based on the format defined in IPFIX, sent to the configured server, and processed by the collector on the server.
- When sending flow records, create a template containing the record format and send the template to the server. After receiving the template, the server learns the format of the subsequent flow records based on the template, and interprets the subsequent IPFIX flow records based on the flow record format defined in the template. The template and flow record information can co-exist in the same packet, or a packet can be accompanied by multiple flow records or templates.
- The IPFIX device can send only the key attributes of a flow, and the software on the server can also extract only the useful flow attributes from flow records, and ignore other flow attributes. For example, a device may determine that the TCP flag can be used to analyze network security, and therefore include the TCP flag in the flow export information. However, some server software support only traffic analysis, but not network analysis. Therefore, the software can ignore the TCP flag sent by the device. There is no compatibility issue between the device and the server. Similarly, when the network device expands a new field (attribute) in the flow record, it only needs to recreate the corresponding template. The server can interpret the flow record information based on the new template, without upgrading the server software. This greatly improves the scalability.

4. Flexible IPFIX Flow Sampling

A template and flow record can be selected based on configurations. The concept of flexible sampling is introduced based on the basic sampling function of IPFIX. You can flexibly configure keywords of the flow record and customize the template based on your requirements.

1.1.4 Protocols and Standards

- RFC 5470: Architecture for IP Flow Information Export
- RFC 5472: IP Flow Information Export (IPFIX) Applicability
- RFC 5103: Bidirectional Flow Export Using IP Flow Information Export (IPFIX)
- RFC 5153: IP Flow Information Export (IPFIX) Implementation Guidelines
- RFC 5102: Information Model for IP Flow Information Export

- RFC 5101: Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information
- RFC 3917: Requirements for IP Flow Information Export
- RFC 3954: Cisco Systems NetFlow Services Export Version 9

1.2 Configuration Task Summary

IPFIX configuration includes the following tasks:

(1) Configuring IPFIX Flow Statistics

(2) Configuring IPFIX Flow Export

(3) (Optional) Configuring Parameters of the IPFIX Flow TableConfiguring Flexible Sampling

1.3 Configuring IPFIX Flow Statistics

1.3.1 Overview

By configuring IPFIX flow statistics, you can obtain statistics of the following types of packets in the network:

- IP packets
- IPv6 packets
- MPLS packets

1.3.2 Restrictions and Guidelines

- Assume that the sampling function is configured under a link aggregation and link aggregation loading balancing is not configured. If the member ports of a link aggregation are not on the same line card, and packets with the same seven key attributes are collected from different line cards, the number of IPFIX packets received by the server from the link aggregation within the unit aging time is the same as the total number of IPFIX packets received by the member ports from different line cards. In addition, the flow aging time of the AP is the same as that on a single line card.
- If IPFIX ACL-based sampling is enabled on a port, the ACL is used to match the IP packets; otherwise, sampling does not take effect on the port. If IPFIX sampling based on the IPv4 ACL is enabled on a port, the ACL is used to match the IPv4 packets. If IPFIX sampling based on the IPv6 ACL is enabled on a port, the ACL is used to match the IPv6 packets. For more information about ACL configuration, see *Configuring ACL*.
- When configuring statistics about the outer VXLAN packets, If the inner and outer VXLAN packets are of different types, for example, one is an IPv4 packet and the other is an IPv6 packet, the inner and outer packets cannot be sampled at the same time. In this case, you can configure either sampling of inner IPv4 and IPv6 packets or sampling of outer IPv4 and IPv6 packets.
- The **netflow-sampling-map** command is configured in global configuration mode. You can enter the sampler configuration mode, and then configure the sampler. If you enter the sampler mode, but do not configure any sampler, the sampler configuration mode is not displayed when you run **display current-config**.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Enable global sampling of VXLAN outer packets.

{ **ip** | **ipv6** } **netflow vxlan out**

VXLAN inner packets are sampled by default.

(4) Configure a sampler map (sampling rule table).

netflow-sampling-map *sampler-map-name*

No sampler map is configured by default.

(5) Configure a random sampler.

mode random one-out-of *packet-interval*

No random sampler is configured by default.

(6) Exit the IPFIX sampler configuration mode.

exit

(7) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

(8) Enable IPFIX in the ingress direction of an interface.

(IPv4 main mode)

ip netflow ingress

(IPv6 main mode)

ipv6 netflow ingress

By default, IPFIX is disabled in the ingress direction of an interface.

(9) Enable IPFIX in the egress direction of an interface.

(IPv4 main mode)

ip netflow egress

(IPv6 main mode)

ipv6 netflow egress

By default, IPFIX is disabled in the egress direction of an interface.

(10) Apply a sampler to an interface.

(IPv4 main mode)

netflow-sampling *sampler-map-name* [**egress**] [**acl** *acl-name* | **acl** *acl-number* }]

(IPv6 main mode)

ipv6 netflow-sampling *sampler-map-name* [**egress**] [**acl** { *acl-name* | **acl** *acl-number* }]

No sampler is applied to an interface by default.

1.4 Configuring IPFIX Flow Export

1.4.1 Overview

This section describes how to send the aged flow records to the server.

1.4.2 Restrictions and Guidelines

- The configured IPFIX server address must be reachable; otherwise, it cannot be displayed after the **display ip netflow export** or **display ipv6 netflow export** command is executed.
- If the type of interface used for exporting flow records is changed from port to VLAN and the interface is an SVI, you also need to configure IPFIX on the VLAN interface.
- Packets can be exported to a maximum of two IPFIX servers.
- You can configure an IPv4 or IPv6 address as the server address for IPv4 or IPv6 IPFIX. When the configured source port and the IPFIX server are not in the same VRF, the IP address of the specified source port cannot be used as the source IP address for packet transmission. The system obtains the best-matched local address as the source address based on the server address. The management port has a default VRF. If you want to use the management port to send packets, configure the source port and the management port in the same VRF.
- The **ip netnetflow-export template** and **ipv6 netnetflow-export template** commands are used to configure the template retransmission frequency. The IPFIX server parses the IPFIX packets based on the template. When the device just starts to use the IPFIX function, the template retransmission frequency can be decreased, so that the IPFIX server can receive the IPFIX template in a short time. For example, you can set the template retransmission frequency to 1 minute.
- The **ip netnetflow-export source** and **ipv6 netnetflow-export source** commands are used to configure the source IP address of an exported packet. By default, the source address of an exported IPFIX packet is the local device IP address which is used to ping the destination IP address. If the source interface is specified, the primary address (or the first global IPv6 address if any) of the interface is the source IP address of the exported IPFIX packet.
- If the source interface is not specified, the default source address is used. You can configure an IPv4 or IPv6 address, or both as the IPFIX server address. If you specify the source interface and configure an IPv6 or IPv4 address as the IPFIX server address, make sure that an IPv6 or IPv4 address is configured for the source interface. Otherwise, the system obtains the best-matched local address as the source address based on the server address. Make sure that the source interface and the IPFIX server are in the same VRF; otherwise, the system obtains the best-matched local address as the source address based on the server address.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the IPFIX server address.

(IPv4 main mode)

ip netnetflow-export target { *hostname* | *ipv4-address* | **ipv6** *ipv6-address* } *por-number* [**vrf** *vrf-name* | **oob** [**via mgmt** *mgmt-name*]

(IPv6 main mode)

ipv6 netnetflow-export target { *hostname* | *ipv4-address* | **ipv6** *ipv6-address* } *por-number* [**vrf** *vrf-name* | **oob** [**via mgmt** *mgmt-name*]

By default, no IPFIX flow record is exported to the server.

(4) (Optional) Configure the exported packet format.

(IPv4 main mode)

ip netnetflow-export version { **9** | **10** }

(IPv6 main mode)

ipv6 netnetflow-export version { **9** | **10** }

By default, the exported packet format is version 10.

(5) (Optional) Configure the source interface for packet export.

(IPv4 main mode)

ip netnetflow-export source *interface-type interface-number*

(IPv6 main mode)

ipv6 netnetflow-export source *interface-type interface-number*

By default, no destination address and port are configured, and the local interface address that best-matches the server address is adopted.

(6) (Optional) Configure parameters related to the export template.

(IPv4 main mode)

ip netnetflow-export template { **model** *model-id* | **options** { **export-stats** | **refresh-rate** *refresh-rate* | **timeout** *time* } | **refresh**-**rate** *refresh-rate* | **timeout** *timeout* }

(IPv6 main mode)

ipv6 netnetflow-export template { **model** *model-id* | **options** { **export-stats** | **refresh-rate** *refresh-rate* | **timeout** *time* } | **refresh**-**rate** *refresh-rate* | **timeout** *timeout* }

By default, the template retransmission frequency is 20 packets in 10 minutes.

1.5 Configuring Parameters of the IPFIX Flow Table

1.5.1 Overview

The parameters adjust the number of cache entries and the aging time of flow records in the IPFIX flow table.

1.5.2 Restrictions and Guidelines

- If the global IPFIX function is enabled, the modification of cache entries does not take effect immediately. To make the modification effective, you need to reconfigure the global IPFIX function or restart the device.
  a To disable the global IPFIX function, run the **unset ip netflow ingress** or **unset ip netflow egress** command.
  b To restart the IPFIX flow statistics function, run the **ip netflow ingress** command.
- IPv4 and IPv6 share the capacity for cache entries.
  - If only IPv4 or IPv6 packet sampling is enabled, the maximum number of IPv4 or IPv6 cache entries can reach the maximum capacity.
  - If both IPv4 and IPv6 packet sampling are enabled, the maximum number of IPv4 and IPv6 cache entries in total cannot exceed the maximum capacity.
- To modify the aging time of an active or inactive entry, run the **ip netflow-cache aging** command.
  - Modification of the aging time affects the duration that a flow record exists in the flow table. A longer aging time indicates that a flow record exists longer in the flow table, and it takes longer for the IPFIX server to receive the flow record information.
  - To display the flow information in real time on the server, you are advised to set the aging time to a small value. For example, set the aging time of an active entry to 1 minute.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the number of cache entries in the flow table.

(IPv4 main mode)

ip netflow-cache entries *number*

(IPv6 main mode)

ipv6 netflow-cache entries *number*

(4) (Optional) Configure the aging time of a flow record.

(IPv4 main mode)

ip netflow-cache aging { **active** *active-time* | **inactive** *inactive-time* }

(IPv6 main mode)

ipv6 netflow-cache aging { **active** *active-time* | **inactive** *inactive-time* }

By default, the aging time is **30** minutes for an active entry and **15** seconds for an inactive entry.

1.6 Configuring Flexible Sampling

Note

The Flexible Sampling command is not supported by the S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC.

1.6.1 Overview

This section describes how to flexibly create flow entries and construct a template.

1.6.2 Restrictions and Guidelines

To enable the flow-key function, ensure that the IPFIX function is already enabled.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the netflow-key configuration mode.

(IPv4 main mode)

ip netflow-key

(IPv6 main mode)

ipv6 netflow-key

By default, the flow-key function is disabled.

(4) Configure the flow table matching conditions for flexible sampling.

match { **flow direction** | **interface** { **input** | **output** } | **ip** { **protocol** | **tos** } | **ipv4** { **destination-address** | **source-address** [**mask**] } } | **transport** { **destination-port** | **source-port** } | **vlan** }*

(5) Enter the template-key configuration mode.

(IPv4 main mode)

ip template-key

(IPv6 main mode)

ipv6 template-key

By default, the template-key function is disabled.

This command can be configured only after the collect function is enabled.

(6) Configure parameters of the IPFIX template for flexible sampling.

collect { **counter** { **bytes** | **packets** } | **timestamp sys-uptime** { **first** | **last** } | **transport tcp flags** }

By default, no parameter of the IPFIX template is configured for flexible sampling.

(7) Enable the flow-key and template-key modes.

enabled

In flow-key configuration mode, this command is used to enable the flexible sampling function. In template-key configuration mode, this command is used to enable the flexible template function.

1.7 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 IPFIX monitoring

Command	Purpose
cleanup { **ip** | **ipv6** } **netflow stats**	Clears the flow table information.
display { **ip** | **ipv6** } **netflow cache**	Displays the IPFIX flow table information.
display { **ip** | **ipv6** } **netflow export**	Displays the IPFIX server information.
display ip netflow interface	Displays the IPFIX configurations of an interface.

1.8 Configuration Examples

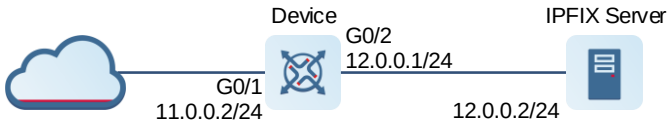
1.8.1 Configuring the IPv4 IPFIX Function

1. Requirements

The device is connected to the network through GigabitEthernet 0/1, and needs to monitor and analyze the IPv4 traffic passing through GigabitEthernet 0/1. Therefore, enable the IPv4 IPFIX function on the device, and configure the IPFIX sampling function in the ingress direction of GigabitEthernet 0/1.

2. Topology

Figure 1-4 Configuring the IPv4 IPFIX Function



3. Notes

- Configure the IP address of the device so that the device and the IPFIX server are mutually reachable at layer 3.
- Configure the sampling function.

- Export the IPv4 IPFIX flow records.
- Configure parameters of the IPv4 IPFIX flow table.

4. Procedure

(1) Configure the IP address of the device.

a. Set the IPv4 address of GigabitEthernet 0/1 to 11.0.0.2/24.

```
Device> privilege
Devcie# config terminal
Device(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# ip address 11.0.0.2 255.255.255.0
Device(cfg-if-GigabitEthernet 0/1)#exit
```

b. Set the IPv4 address of GigabitEthernet 0/2 to 12.0.0.1/24.

```
Device(cfg)# interface gigabitEthernet 0/2
Device(cfg-if-GigabitEthernet 0/2)# ip address 12.0.0.1 255.255.255.0
Device(cfg-if-GigabitEthernet 0/2)#exit
```

(2) Configure the IPFIX sampling function.

a. Configure a sampler map "zxw".

```
Device(cfg)#netflow-sampling-map zxw
```

b. Set the random sampling rate to 4096.

```
Device(cfg-sampler)#mode random one-out-of 4096
Device(cfg-sampler)#exit
```

c. Apply the sampler map to GigabitEthernet 0/1.

```
Device(cfg)# interface gigabitEthernet 0/1
Devcie(cfg-if-GigabitEthernet 0/1)#netflow-sampling zxw
```

d. Enable IPv4 packet sampling in the ingress direction of GigabitEthernet 0/1.

```
Devcie(cfg-if-GigabitEthernet 0/1)#ip netflow ingress
Devcie(cfg-if-GigabitEthernet 0/1)#exit
```

(3) Export the IPv4 IPFIX flow records to the server.

a. Export the flow records to the server with the IPv4 address 12.0.0.2 and port ID 9996.

```
Devcie(cfg)# ip netflow-export target 12.0.0.2 9996
```

b. Configure the IP address of GigabitEthernet 0/1 as the source IP address for packet export.

```
Devcie(cfg)# ip netflow-export source gigabitEthernet 0/2
```

c. Set the template retransmission frequency to 40 minutes.

```
Devcie(cfg)# ip netflow-export template timeout-rate 40
```

d. Configure the export option **export-stats**.

```
Devcie(cfg)# ip netflow-export template options export-stats
```

e. Set the export option **refresh-rate** to 35 packets.

```
Devcie(cfg)# ip netflow-export template options refresh-rate 35
```

f. Set the export option **timeout-rate** to 45 minutes.

```
Device(cfg)# ip netflow-export template options timeout-rate 45
```

g. Export packets using the standard format of Netflow V9.

```
Devcie(cfg)# ip netflow-export version 9
```

(4) Configure parameters of the IPv4 IPFIX flow table.

a. Set the number of cache entries in the flow table to 32768.

```
Device(cfg)# ip netflow-cache entries 32768
```

b. Set the aging time of an active entry to 2 minutes.

```
Device(cfg)# ip netflow-cache aging active 2
```

c. Set the aging time of an inactive entry to 20 seconds.

```
Device(cfg)# ip netflow-cache aging inactive 20
```

5. Verification

Run the **ping** command to verify L3 reachability between the device and the IPFIX server.

```
Device# ping 12.0.0.2
Sending 5, 100-byte ICMP Echoes to 12.0.0.2, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/2 ms.
```

Run the **display ip netflow export** command to display the IPFIX server information.

```
Devcie# display ip netflow export

IPV4 main Cache:
        Exporting flows to 12.0.0.2 (9996)
        Exporting using source IP address 12.0.0.1
        Version 9 flow records
        IPV4 Template ID = 266
        IPV4 Template ID = 292
                Template timeout = 40 Minute(s)
                Template refresh rate = 30
        0 flows exported in 0 udp datagrams
        0 flows failed to export
        0 messages failed to export
```

Run the **display ip netflow interface** command to display the IPFIX configurations of the interface.

```
Devcie# display ip netflow interface
interface GigabitEthernet0/1
  netflow-sampling zxw
  ip netflow ingress
```

Run the **display ip netflow cache** command to display the information about the IPv4 IPFIX flow table.

```
Devcie# display ip netflow cache
IPFIX CACHE(slot-0/1):
        IPV4 Flow Switching Cache, 0 bytes
        0 active, - inactive, 0 added
```

```
            0 flow alloc failures
            0 displayed
            Exporting flows to 12.0.0.2 (9996)
            0 ipv4 flows exported in 0 udp datagrams, 0 failed
            And Exporting flows to 0.0.0.0 (0)
            0 ipv4 flows exported in 0 udp datagrams, 0 failed
            Last clearing of statistics 0 Days 00:00:16

SrcIf                          SrcIPAddress        DstIf                          DstIPAddress
VlanId   VNI              Pr   Tos Flgs   Pkts
Port     Msk     AS                            Port     Msk     AS              NextHop
B/Pk   Active

Protocol     Total         Flows      Packets      Bytes      Packets   Active(Sec) Idle(Sec)
--------     Flows         /Sec       /Flow        /Pkt       /Sec      /Flow       /Flow
Total:         0             0          0            0          0         0           0
IPFIX CACHE(slot-0/1):
            IPV4 Flow Switching Cache, 0 bytes
            0 active, - inactive, 0 added
            0 flow alloc failures
            0 displayed
            Exporting flows to 12.0.0.2 (9996)
            0 ipv4 flows exported in 0 udp datagrams, 0 failed
            And Exporting flows to 0.0.0.0 (0)
            0 ipv4 flows exported in 0 udp datagrams, 0 failed
            Last clearing of statistics 0 Days 00:00:16

SrcIf                          SrcIPAddress        DstIf                          DstIPAddress
VlanId   VNI              Pr   Tos Flgs   Pkts
Port     Msk     AS                            Port     Msk     AS              NextHop
B/Pk   Active
GigabitEthernet 0/1          13.0.0.2              -                              11.0.0.2            1       -
6    0    -       112
1024     -       -                                 179      -       -                -
124    1

Protocol     Total         Flows      Packets      Bytes      Packets   Active(Sec) Idle(Sec)
--------     Flows         /Sec       /Flow        /Pkt       /Sec      /Flow       /Flow
TCP-BGP        1             0          112          124        7         1           0
Total:         1             0          112          124        7         1           0
```

6. Configuration Files

Device configuration file

```
hostname Device
! interface gigabitEthernet 0/1
unset switch-port
```

```
ip address 11.0.0.2 255.255.255.0
netflow-sampling zxw
ip netflow ingress
!
interface gigabitEthernet 0/2
unset switch-port
ip address 12.0.0.1 255.255.255.0

netflow-sampling-map zxw
  mode random one-out-of 4096
!
ip netflow-cache aging active 2
ip netflow-cache aging inactive 20
ip netflow-cache entries 32768
ip netflow-export source GigabitEthernet 0/1
ip netflow-export template timeout 40
ip netflow-export template options export-stats
ip netflow-export template options timeout-rate 45
ip netflow-export template options refresh-rate 35
ip netflow-export version 9
ip netflow-export target 12.0.0.2 9996
!
end
```

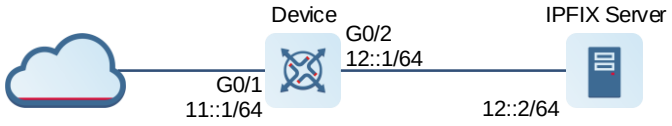
1.8.2 Configuring the IPv6 IPFIX Function

1. Requirements

The device is connected to the network through GigabitEthernet 0/1, and needs to monitor and analyze the IPv6 traffic passing through GigabitEthernet 0/1. Therefore, enable the IPv6 IPFIX function on the device, and configure the IPFIX sampling function in the ingress direction of GigabitEthernet 0/1.

2. Topology

Figure 1-5 Configuring the IPv6 IPFIX Function



3. Notes

- Configure the IP address of the device so that the device and the IPFIX server are mutually reachable at layer 3.
- Configure the sampling function.
- Export the IPv6 IPFIX flow records.
- Configure parameters of the IPv6 IPFIX flow table.

4. Procedure

(1) Configure the IP address of the device.

a. Set the IPv6 address of GigabitEthernet 0/1 to 11::1/64.

```
Device> privilege
Devcie# config-system terminal
Device(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# ipv6 address 11::1/64
Device(cfg-if-GigabitEthernet 0/1)#exit
```

b. Set the IPv6 address of GigabitEthernet 0/2 to 12::1/64.

```
Device(cfg)# interface gigabitEthernet 0/2
Device(cfg-if-GigabitEthernet 0/2)# unset switch-port
Device(cfg-if-GigabitEthernet 0/2)# ipv6 address 12::1/64
Device(cfg-if-GigabitEthernet 0/2)#exit
```

(2) Configure the IPFIX sampling function.

a. Configure a sampler map "zxw".

```
Device(cfg)#netflow-sampling-map zxw
```

b. Set the random sampling rate to 4096.

```
Device(cfg-sampler)#mode random one-out-of 4096
Device(cfg-sampler)#exit
```

c. Apply the sampler map to GigabitEthernet 0/1.

```
Device(cfg)# interface gigabitEthernet 0/1
Devcie(cfg-if-GigabitEthernet 0/1)#ipv6 netflow-sampling zxw
```

d. Enable IPv6 packet sampling in the ingress direction of GigabitEthernet 0/1.

```
Devcie(cfg-if-GigabitEthernet 0/1)#ipv6 netflow ingress
Devcie(cfg-if-GigabitEthernet 0/1)#exit
```

(3) Export the IPv6 IPFIX flow records to the server.

a. Export the flow records to the server with IPv6 address 12::2 and port ID 9996.

```
Devcie(cfg)# ipv6 netflow-export destination 12::2 9996
```

b. Configure the IP address of GigabitEthernet 0/1 as the source IP address for packet export.

```
Devcie(cfg)# ipv6 netflow-export source gigabitEthernet 0/2
```

c. Set the template retransmission frequency to 40 minutes.

```
Devcie(cfg)# ipv6 netflow-export template timeout-rate 40
```

d. Configure the export option **export-stats**.

```
Devcie(cfg)# ipv6 netflow-export template options export-stats
```

e. Set the export option **refresh-rate** to 35 packets.

```
Devcie(cfg)# ipv6 netflow-export template options refresh-rate 35
```

f. Set the export option **timeout-rate** to 45 minutes.

```
Device(cfg)# ipv6 netflow-export template options timeout-rate 45
```

g. Export packets using the standard format of Netflow V9.

```
Devcie(cfg)# ipv6 netflow-export version 9
```

(4) Configure parameters of the IPv6 IPFIX flow table.

a. Set the number of cache entries in the flow table to 32768.

```
Device(cfg)# ipv6 netflow-cache entries 32768
```

b. Set the aging time of an active entry to 2 minutes.

```
Device(cfg)# ipv6 netflow-cache timeout active 2
```

c. Set the aging time of an inactive entry to 20 seconds.

```
Device(cfg)# ipv6 netflow-cache timeout inactive 20
```

5. Verification

Run the **ping** command to verify L3 reachability between the device and the IPFIX server.

```
Device# ping 12::2
Sending 5, 100-byte ICMP Echoes to 12::2, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/2 ms.
```

Run the **display ipv6 netflow export** command to display the IPFIX server information.

```
Devcie# display ipv6 netflow export

IPV6 main Cache:
        Exporting flows to 12::2 (9996)
        Exporting using source IP address 30::1
        Version 9 flow records
        IPV6 Template ID = 5000
        IPV6 Template ID = 5026
                Template timeout = 40 Minute(s)
                Template refresh rate = 30
        0 flows exported in 0 udp datagrams
        0 flows failed to export
        0 messages failed to export
```

Run the **display ip netflow interface** command to display the IPFIX configurations of the interface.

```
Devcie# display ip netflow interface
interface GigabitEthernet0/1
  ipv6 netflownetflow-sampling zxw
  ipv6 netflow ingress
```

Run the **display Ipv6 netflow cache** command to display the information about the IPv6 IPFIX flow table.

```
Devcie# display ipv6 netflow cache
IPFIX CACHE(slot-0/1):
        Ipv6 netflow Switching Cache, 0 bytes
        0 active, - inactive, 0 added
        0 flow alloc failures
        0 displayed
        Exporting flows to 12::2 (9996)
```

```
            0 ipv6 netflows exported in 0 udp datagrams, 0 failed
            And Exporting flows to 0.0.0.0 (0)
            0 ipv6 netflows exported in 0 udp datagrams, 0 failed
            Last cleanup of statistics 0 Days 00:00:18

SrcIf                           SrcIPAddress        DstIf                           DstIPAddress
VlanId   VNI              Pr   Tos Flgs   Pkts
Port     Msk     AS                                 Port     Msk     AS              NextHop
B/Pk   Active

Protocol     Total         Flows      Packets      Bytes      Packets   Active(Sec) Idle(Sec)
--------     Flows         /Sec       /Flow        /Pkt       /Sec      /Flow       /Flow
Total:         0             0          0            0          0         0           0
IPFIX CACHE(slot-0/1):
            IPV6 Flow Switching Cache, 0 bytes
            0 active, - inactive, 0 added
            0 flow alloc failures
            0 displayed
            Exporting flows to 12::2 (9996)
            0 ipv6 flows exported in 0 udp datagrams, 0 failed
            And Exporting flows to 0.0.0.0 (0)
            0 ipv6 flows exported in 0 udp datagrams, 0 failed
            Last clearing of statistics 0 Days 00:00:18

SrcIf                           SrcIPAddress        DstIf                           DstIPAddress
VlanId   VNI              Pr   Tos Flgs   Pkts
Port     Msk     AS                                 Port     Msk     AS              NextHop
B/Pk   Active
GigabitEthernet 0/1             13::2               -                               11::2               1
-                    6    0    -       112
1024     -       -                                  179      -       -               -
124    1

Protocol     Total         Flows      Packets      Bytes      Packets   Active(Sec) Idle(Sec)
--------     Flows         /Sec       /Flow        /Pkt       /Sec      /Flow       /Flow
TCP-BGP        1             0          112          124        7         1           0
Total:         1             0          112          124        7         1           0
Protocol     Total         Flows      Packets      Bytes      Packets   Active(Sec) Idle(Sec)
--------     Flows         /Sec       /Flow        /Pkt       /Sec      /Flow       /Flow
TCP-BGP        1             0          112          124        7         1           0
Total:         1             0          112          124        7         1           0
```

6. Configuration Files

Device configuration file

```
hostname Device
!
```

```
interface gigabitEthernet 0/1
unset switch-port
ipv6 address 11::1/64
netflow-sampling zxw
ipv6 netflow ingress
!
interface gigabitEthernet 0/2
unset switch-port
ipv6 address 12::1/64
!
netflow-sampling-map zxw
 mode random one-out-of 4096
!
ipv6 netflow-cache aging active 2
ipv6 netflow-cache aging inactive 20
ipv6 netflow-cache entries 32768
ipv6 flownetflow-export source GigabitEthernet 0/1
ipv6 netflow-export template timeout-rate 40
ipv6 netflow-export template refresh-rate 30
ipv6 netflow-export template options export-stats
ipv6 netflow-export template options timeout-rate 45
ipv6 netflow-export template options refresh-rate 35
ipv6 netflow-export version 9
ipv6 netflow-export destination ipv6 12::2 9996
!
end
```

Contents

1 Configuring the Function of Displaying Device Restart Reasons 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Restrictions and Guidelines 1
1.3 Monitoring 1

1 Configuring the Function of Displaying Device Restart Reasons

1.1 Introduction

1.1.1 Overview

Diversified availability detection functions are implemented to ensure the availability of all system functions. While these functions ensure system availability by killing target processes or resetting cards, the following risks arise: The self-recovery behavior cannot be tracked or controlled in a unified manner. When a process is killed or a card is restarted, the service that triggers this self-recovery behavior cannot be located. If a process is frequently killed or a card is frequently reset due to the logic failure of a detection function, more serious system failure may occur. To solve these problems, a component is needed to track and control the service self-recovery behaviors of the system in a unified manner. The secondary protect and self-healing (PSH) module provides a method to check the device restart reason on the CLI interface. When a device restarts, you can view the device restart reasons by using the command provided by the PSH module, eventually improving the availability and quality of the device.

1.1.2 Principles

The PSH module collects information about the process that triggers restart of the device before the device restarts, and saves the restart reason locally. You can either run a command on a device to view the locally stored restart reason of the device, or run a command on the primary device of a cluster to view the restart reasons of all devices in the cluster.

1.2 Restrictions and Guidelines

The PSH module records only self-healing events of the system, and service components cannot use it as an internal logging tool to record logs of the running process in a log file.

1.3 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Table 1-1 Monitoring of the Function of Displaying Device Restart Reasons

Command	Purpose
display reboot-reason [**all**]	Displays the device restart reason.

Contents

1 Configuring NTP....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.1.3 Protocols and Standards....3
1.2 Restrictions and Guidelines....3
1.3 Configuration Task Summary....3
1.4 Configuring NTP Security Authentication....4
1.4.1 Overview....4
1.4.2 Restrictions and Guidelines....4
1.4.3 Procedure....4
1.5 Configuring NTP in Client/Server Mode....5
1.5.1 Restrictions and Guidelines....5
1.5.2 Overview....5
1.5.3 Procedure....5
1.6 Configuring NTP-Related Parameters....6
1.6.1 Procedure....6
1.7 Configuring NTP Access Control....6
1.7.1 Overview....6
1.7.2 Restrictions and Guidelines....7
1.7.3 Procedure....7
1.8 Monitoring....7

1.9 Configuration Examples....8
1.9.1 Configuring the Local Clock Reference Mode of NTP....8
1.9.2 Configuring the External Clock Reference Mode of NTP....10
1.9.3 Configuring NTP Security Authentication....12
1.9.4 Configuring NTP Access Control Rights....15

1 Configuring NTP

1.1 Introduction

1.1.1 Overview

Network Time Protocol (NTP) is an application-layer protocol that enables network devices to synchronize time. NTP enables network devices to synchronize time with their servers or clock sources and provides high-precision time correction (the difference from the standard time is smaller than one millisecond in a LAN and smaller than decades of milliseconds in a WAN). In addition, NTP can prevent attacks by using encrypted acknowledgment. NTP is implemented based on the transport protocol, User Datagram Protocol (UDP), and the UDP port used is 123.

Currently, Ruijie devices can be used both as NTP clients and NTP servers. In other words, a Ruijie device can synchronize time with a time server, or act as a time server to provide time synchronization for other devices. When a device is used as a server, it supports only the unicast server mode.

1.1.2 Principles

1. Basic Concepts

- NTP server

  A device uses a local clock as the reference clock source to provide time synchronization for other devices in the network.

- NTP client

  A device synchronizes time with an NTP server in the network.

- **Stratum**

  In NTP, "stratum" is used to describe the hops from a device to an authority clock source. An NTP server whose stratum is 1 has a directly connected atomic clock or radio controlled clock; an NTP server whose stratum is 2 obtains time from the server whose stratum is 1; an NTP server whose stratum is 3 obtains time from the server whose stratum is 2; and so on. Therefore, clock sources with lower stratum have higher clock precision.

- Hardware Clock

  A hardware clock operates based on the frequency of the quartz crystal resonator on a device and is powered by the device battery. After the device is shut down, the hardware clock continues running. After the device is started, the device obtains time information from the hardware clock as the software time of the device.

2. NTP Principles

As shown in Figure 1-1, a client exchanges NTP packets with a server to implement time synchronization.

(1) The client sends a time synchronization request packet to all servers every 64 seconds.

(2) After receiving response packets from the servers, the client filters these response packets, and synchronizes time with an optimum server.

3. NTP Working Process

As shown in Figure 1-1, device B (B for short) acts as an NTP reference clock source, and device A (A for short) acts as an NTP client that synchronizes time with B. At a time point, the local clock of A is 19:00:00 and the local clock of B is 19:30:20.

(1) A sends an NTP request packet. The local time (T0) when the packet leaves from A is 19:00:00 and is filled in **Originate Timestamp**.

(1) After a network delay of 2s, the local time (T1) when B receives the request packet is 19:30:23 and is filled in **Receive Timestamp**.

(2) B processes the NTP request and sends an NTP response packet one second later. The local time (T2) when the response packet leaves from B is 19:30:24 and is filled in **Transmit Timestamp**.

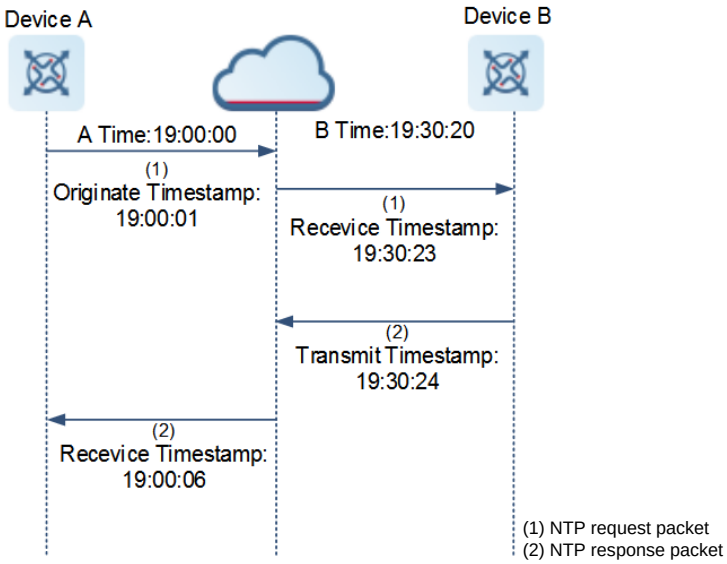
(3) After a network delay of 2s, A receives the response packet. The local time (T3) when the response packet arrives at A is 19:00:06.

The specific calculations for time synchronization are as follows:

a A obtains the time difference of 30 minutes and 20 seconds between B and A by using the formula ((T1-T0)+(T2-T3))/2.

b A obtains the packet round-trip delay of 4 seconds between A and B by using the formula (T3-T0)-(T2-T1).

Figure 1-1 NTP Principles



- NTP working mode
  - External clock reference mode: In this mode, a device acts as both a server and a client. If receiving time synchronization requests from other clients, the device must synchronize time with the specified server, and then provide the time synchronization service for the clients only after successful synchronization with the specified server.
  - Local clock reference mode: In this mode, a device uses the default local clock as the reliable clock source and provides the time synchronization service directly for other clients.

4. NTP Security Mechanism

To prevent malicious damage to the time server and improve the security of time synchronization, NTP provides security authentication and access control functions.

- Security authentication

  NTP uses the authentication mechanism to check whether the time synchronization information really comes from the announced server and check the information return path to provide an anti-interference protection mechanism.

  An NTP client and an NTP server are configured with the same key. When sending request and response packets, a device calculates the hash values of the packets by using the MD5 algorithm based on the specified key and NTP packet content, and fills the hash values into the packet authentication information. The receiving device checks whether the packets are sent by a trusted device or modified based on the authentication information.

- NTP access control using the kiss-o-death (KoD) packet

  NTP provides a minimum security measure by using an ACL. Normally, the denial-of-service packets are discarded, and there is no operation other than incrementing the statistics counter. Sometimes a more proactive response is required. For example, the client is explicitly requested to stop sending packets to the server and to leave a message for the system operator. Therefore, open-source NTP proposes a KoD packet, which can be used to respond to special processing of the client when access control is configured on the server.

  The format of a KoD packet is the same as that of a common NTP packet, except that the **Leap Indicator** field is set to **clock unsynchronized**, the **Stratum** field is set to **0**, and the **Reference ID** field is set to a 4-byte ASCII code. The code is "RATE" if the limited flag is set and the rate threshold is exceeded. After the NTP client receives the KoD packet with the code "RATE", it limits the rate according to the minimum interval specified in the packet.

1.1.3 Protocols and Standards

- RFC 1305: Network Time Protocol (Version 3)
- RFC 5905: Network Time Protocol (Version 4)

1.2 Restrictions and Guidelines

- You cannot configure both the NTP and SNTP functions on the same device.
- Only Layer 3 Ethernet interfaces or link aggregations support NTP.
- To ensure accuracy of time synchronization, avoid configuring two or more clock sources whenever possible to prevent clock flapping.
- You can run the **ntp service disable** command to disable the NTP server function so that the device acts only as a client. This command cannot be configured together with **ntp master**.

1.3 Configuration Task Summary

NTP configuration includes the following tasks:

(1) (Optional) Configuring NTP Security Authentication

(2) Configuring NTP in Client/Server Mode

(3) (Optional) Configuring NTP-Related Parameters

(4) (Optional) Configuring NTP Access Control

1.4 Configuring NTP Security Authentication

1.4.1 Overview

- Synchronizing time from a trusted reference clock source: Use a device as a client to synchronize time only from a trusted external reference clock source to the local clock.
- Providing time synchronization for a trusted device: Use the local clock of a device as the NTP reference clock source to provide time synchronization for only a trusted device.

1.4.2 Restrictions and Guidelines

- The authentication keys of the client and server must be the same.
- To provide time synchronization for a trusted device, you must specify a trusted authentication key by using the key ID. Only one trusted key can be configured. The specified authentication key must be consistent with that of the trusted device.
- To synchronize time with a trusted reference clock source, you must specify a trusted authentication key by using the key ID. Each trusted reference clock source is mapped to an authentication key. The authentication keys must be consistent with the keys of trusted reference clock sources.
- By default, a client does not use a global security authentication mechanism. If no security authentication mechanism is used, communication will not be encrypted. A global security indicator is not enough to imply that the communication between the client and server is implemented in an encrypted manner. Other global keys and an encryption key for the server must also be configured for initiating encrypted communication between the client and server.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a global security authentication mechanism for NTP.

ntp-service authenticate

(4) Configure a global authentication key for NTP.

ntp-service auth-algorithm-key *authentication-key-id* **md5** *authentication-key-string* [**0** | **7**]

The NTP global authentication mechanism is disabled by default.

(5) (Optional) Configure a globally trusted key ID for NTP.

ntp-service trust *trusted-key-id*

(6) (Optional) Configure an authentication key ID for an external reference clock source.

ntp-service server [**oob** | **vrf** *vrf-name*] { *ipv4-address* | *ipv6-address* | *server-hostname* | **ip** *domain* | **ipv6** *domain* } [**key** *keyid* | **port** *port-number* | **prefer** | **source** *interface-type interface-number* | **version** *version*]* [**via** *mgmt-name*]

By default, no authentication key ID is configured for an external reference clock source.

1.5 Configuring NTP in Client/Server Mode

1.5.1 Restrictions and Guidelines

- When configuring the NTP server, you must specify at least one external reference clock source. A maximum of 20 different external reference clock sources can be configured.
- If it is necessary to configure an NTP key, you must configure NTP security authentication before configuring the NTP server. A device that acts as a client supports interaction with up to 20 NTP servers. You can configure an authentication key for each server (after configuring global authentication and the related key) to initiate encrypted communication with the servers.
- The default version of NTP for communicating with a server is NTP version 4. In addition, you can configure the source interface for transmitting NTP packets and specify that the NTP packets from a corresponding server can be received only on the transmitting interface.
- After a local clock is configured as the master clock (especially when a clock with a lower stratum is specified), the system does not synchronize with a clock source with a higher stratum, and therefore a real effective clock source may be overwritten. If this function is used for multiple devices in a network, the clock difference between the devices may cause unstable time synchronization of the network.
- Before a local clock is configured as the master clock, if the system never synchronizes time with an external clock source, you may need to manually calibrate the system clock to ensure that there is no excessive difference. For details about how to manually calibrate the system clock, refer to the basic management section in the Basic Configuration Commands.
- By default, if NTP works in client/server mode, after the NTP device synchronizes time from an external reliable clock source, the device acts as the time server to provide the time synchronization service for other devices. If you configure an NTP device only as a client, run the **ntp service disable** command to disable the external time synchronization service provided by NTP. After this command is configured, an external device cannot synchronize time from this NTP device.

1.5.2 Overview

This section describes how to configure a device as a client to synchronize time from an external reference clock source to the local clock. After successful time synchronization, the device can act as a time synchronization server to provide the time synchronization service.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an NTP server.

ntp-service server [**oob** | **vrf** *vrf-name*] { *ipv4-address* | *ipv6-address* | *server-hostname* | **ip** *domain* | **ipv6** *domain* } [**key** *keyid* | **port** *port-number* | **prefer** | **source** *interface-type interface-number* | **version** *version*]* [**via** *mgmt-name*]

1.6 Configuring NTP-Related Parameters

1.6.1 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the interval for time synchronization with the external server.

ntp-service interval *synchronization-interval-time*

By default, the NTP synchronization interval is 64s.

(4) (Optional) Configure automatic update of the hardware clock.

ntp-service client-update-calendar

Automatic update of the hardware clock is not configured by default.

(5) (Optional) Configure the NTP master clock.

ntp-service master [*stratum*]

The NTP master clock is disabled by default.

stratum: specifies the stratum of a local clock, ranging from 1 to 15. The default value is 8.

(6) (Optional) Disable the NTP time synchronization service provided for other devices.

ntp-service global disable

The NTP time synchronization service is enabled by default.

(7) Enter the interface configuration mode.

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

(8) (Optional) Disable receiving of NTP packets on an interface.

ntp-service disable

The function of receiving NTP packets on an interface is enabled by default.

1.7 Configuring NTP Access Control

1.7.1 Overview

NTP access control provides a minimum security measure. A more secure method is to use the NTP authentication mechanism.

1.7.2 Restrictions and Guidelines

- If no access control rule is configured, all NTP services can be accessed. If access control rules are configured, only NTP services that comply with the access control rules can be accessed.
- The system currently does not support the access control query function.
- NTP service packets can only match some fields in an ACL rule, including:
  - Source IPv4 address/source IPv6 address
  - UDP number
- If the packet request interval of the NTP client is smaller than that allowed by the server, the NTP server discards the packets. This function takes effect only after the **ntp access-group limited access-list-number** | **access-list-name** [**kod**] command is configured. When the KoD-enabled NTP device acts as a client, it reduces the rate after receiving a KoD packet. A KoD-disabled device does not reduce the rate. Note that KoD cannot be triggered by specifying the interval in the **ntp interval** command.
- You can run the **ntp discard** command to configure the packet request interval allowed by the server. If the packet request interval of the NTP client is smaller than that allowed by the server, the NTP server discards the packets. This function takes effect only after the **ntp access-group limited access-list-number** | **access-list-name** [**kod**] command is configured.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure the NTP access control rights.

ntp-service access-apply { **limited** | **peer** | **serve** | **serve-only** | **query-only** } { *access-list-name* | *access-list-number* } [**kod**]

No NTP access control rule is configured by default.

(4) (Optional) Configure the minimum packet interval allowed by NTP.

ntp-service discard min-spacing *discard-min-spacing-interval* **avg-spacing** *avg-spacing-interval*

By default, the minimum packet request interval allowed by NTP is 2s, and the average packet request interval is 8s.

1.8 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 NTP Monitoring

Command	Purpose
display ntp-service status	Displays the current NTP information.
display ntp-service packets	Displays the sent and received NTP packets.
display ntp-service server	Displays the NTP server information.
display ntp-service server	Displays the NTP server list.
debugging ntp-service	Debugs the DNS function.
view ntp-service peer	Displays the NTP configuration peer list.
unset debugging ntp-service	Disables the debugging function.

1.9 Configuration Examples

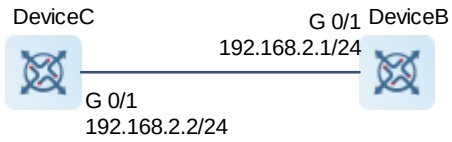
1.9.1 Configuring the Local Clock Reference Mode of NTP

1. Requirements

Use the local clock of a device as the NTP reference clock source to provide time synchronization.

2. Topology

Figure 1-2 Configuring the Local Clock Reference Mode of NTP



3. Notes

- Configure the local clock of device B as the NTP reference clock source.
- Device C synchronizes time from device B.

4. Procedure

Set the IP address of GigabitEthernet 0/1 on device B to 192.168.2.1/24.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Set the IP address of GigabitEthernet 0/1 on device C to 192.168.2.2/24.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.2 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/1)# exit
```

Configure device B as the NTP server.

```
DeviceB(cfg)# ntp-service master
```

Enable the NTP function on device C.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# ntp-service server 192.168.2.1
```

5. Verification

Run the **ping** command to verify L3 reachability between device B and device C.

```
DeviceB# ping 192.168.2.2
Sending 5, 100-byte ICMP Echoes to 192.168.2.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the **clock set** command to modify the time on device B.

```
DeviceB# clock set 11:00:00
Set system clock: 11:00:00 UTC Fri, Feb 26, 2021
```

Run the **display clock** command on device C to check whether time synchronization is successful.

```
DeviceC# display clock
11:00:01 UTC Fri, Feb 26, 2021
```

6. Configuration Files

- Device B configuration file

```
hostname DeviceB
!
interface GigabitEthernet 0/1
 ip address 192.168.2.1 255.255.255.0
ntp-service master
!
end
```

- Device C configuration file

```
hostname DeviceC
!
ntp-service server 192.168.2.1
!
interface GigabitEthernet 0/1
```

```
 ip address 192.168.2.2 255.255.255.0
!
end
```

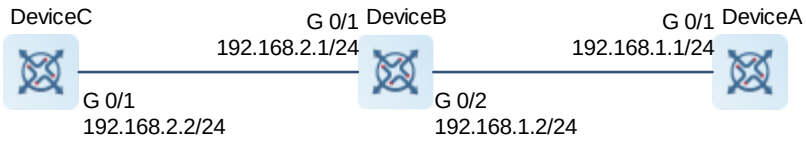
1.9.2 Configuring the External Clock Reference Mode of NTP

1. Requirements

Use a device as a client to synchronize time from an external reference clock source to the local clock. After successful time synchronization, the device can act as a time synchronization server to provide the time synchronization service.

2. Topology

Figure 1-3 Configuring the External Clock Reference Mode of NTP



3. Notes

- Configure the NTP external clock reference mode on device B.
- Configure device A as the reference clock source of device B.
- Device C synchronizes time from device B.

4. Procedure

Set the IP address of GigabitEthernet 0/1 on device A to 192.168.1.1/24.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Set the IP address of GigabitEthernet 0/1 on device B to 192.168.2.1/24, and the IP address of GigabitEthernet 0/2 to 192.168.1.2/24.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
DeviceB(cfg)# interface gigabitEthernet 0/2
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.2 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Set the IP address of GigabitEthernet 0/1 on device C to 192.168.2.2/24.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.2 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/1)# exit
```

Configure device A as the reference clock source of device B.

```
DeviceA(cfg)# ntp-service master
```

Configure the NTP external clock reference mode on device B.

```
DeviceB(cfg)# ntp-service server 192.168.1.1
```

Enable device C to synchronize time from device B.

```
DeviceC(cfg)# ntp-service server 192.168.2.1
```

5. Verification

Run the **ping** command to verify L3 reachability between device B and device A.

```
DeviceB# ping 192.168.1.1
Sending 5, 100-byte ICMP Echoes to 192.168.1.1, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the **ping** command to verify L3 reachability between device B and device C.

```
DeviceB# ping 192.168.2.2
Sending 5, 100-byte ICMP Echoes to 192.168.2.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the **display ntp-service status** command to display the NTP configuration.

```
DeviceB# sh ntp-service status
Clock is synchronized, stratum 9, reference is 192.168.1.1
nominal freq is 250.000 Hz, actual freq is 250.000 Hz, precision is 2**18
reference time is E3527947.D5254A29 (14:03:51.000 UTC Sun, Nov 8, 2020)
clock offset is -0.00012 sec, root delay is 0.00282 sec
root dispersion is 0.01153 msec, peer dispersion is 0.00025 msec
system poll interval is 64, last update was 43 sec ago
system time(GMT) is E3527973.740AB522 (14:04:35.000 GMT Sun, Nov 8, 2020)
```

Run the **clock set** command to modify the time on device A.

```
DeviceA# clock set 11:00:00
Set system clock: 11:00:00 UTC Fri, Feb 26, 2021
```

Run the **display clock** command on device B to check whether time synchronization is successful.

```
DeviceB# display clock
11:00:01 UTC Fri, Feb 26, 2021
```

Run the **display clock** command on device C to check whether time synchronization is successful.

```
DeviceC# display clock
11:00:01 UTC Fri, Feb 26, 2021
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 0/1
 ip address 192.168.1.1 255.255.255.0
ntp-service master
!
end
```

- Device B configuration file

```
hostname DeviceB
!
ntp-service server 192.168.1.1
!
interface GigabitEthernet 0/1
 ip address 192.168.2.2 255.255.255.0
!
interface GigabitEthernet 0/2
 ip address 192.168.1.2 255.255.255.0
!
end
```

- Device C configuration file

```
hostname DeviceC
!
ntp server 192.168.2.1
!
interface GigabitEthernet 0/1
 ip address 192.168.2.2 255.255.255.0
!
end
```

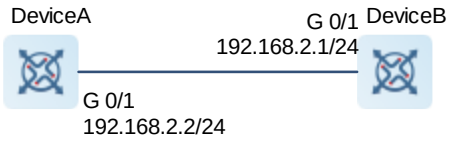
1.9.3 Configuring NTP Security Authentication

1. Requirements

Use a device as a client to synchronize time only from a trusted external reference clock source to the local clock. Use the local clock of a device as the NTP reference clock source to provide time synchronization for only a trusted device.

2. Topology

Figure 1-4 Configuring NTP Security Authentication



3. Notes

- Configure device A as the reference clock source of device B, enable device A to provide device B with the NTP service that requires security authentication, and set the authentication key to "abcd".

4. Procedure

Set the IP address of GigabitEthernet 0/1 on device A to 192.168.2.2/24.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Configure device A as the reference clock source of device B, enable device A to provide the NTP service that requires security authentication, and set the authentication key to **abcd**.

```
DeviceA(cfg)# ntp master
DeviceA(cfg)# ntp authenticate
DeviceA(cfg)# ntp authentication-key 1 md5 abcd
DeviceA(cfg)# ntp trusted-key 1
```

Set the IP address of GigabitEthernet 0/1 on device B to 192.168.2.1/24.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Configure device A as the reference clock source of device B, enable device A to provide the NTP service that requires security authentication, and set the authentication key to **abcd**.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# ntp master
DeviceA(cfg)# ntp authenticate
DeviceA(cfg)# ntp authentication-key 1 md5 abcd
DeviceA(cfg)# ntp trusted-key 1
```

Configure device B to synchronize time from device A, enable device B to provide the NTP service that requires security authentication, and set the authentication key to **abcd**.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# ntp-service trusted-key 1
DeviceB(cfg)# ntp-service server 192.168.2.2
```

5. Verification

Run the **ping** command to verify L3 reachability between device B and device A.

```
DeviceB# ping 192.168.2.2
Sending 5, 100-byte ICMP Echoes to 192.168.2.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the **clock set** command to modify the time on device A.

```
DeviceA# clock set 11:00:00
Set system clock: 11:00:00 UTC Fri, Feb 26, 2021
```

Run the **display clock** command on device B to check whether time synchronization is successful.

```
DeviceB# display clock
11:00:01 UTC Fri, Feb 26, 2021
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 0/1
 ip address 192.168.2.2 255.255.255.0
ntp-service master
!
ntp authentication-key 1 md5 $10$23e$uYC4zg==$ 7
ntp authenticate
ntp trusted-key 1
end
```

- Device B configuration file

```
hostname DeviceB
!
interface GigabitEthernet 0/1
 ip address 192.168.2.1 255.255.255.0
!
ntp authentication-key 1 md5 $10$28a$x5aToA==$ 7
ntp authenticate
ntp trusted-key 1
ntp server 192.168.2.2 key 1
!
```

```
end
```

1.9.4 Configuring NTP Access Control Rights

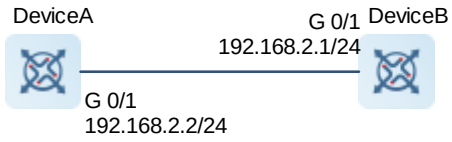
1. Requirements

Configure the access control function to prevent malicious damage to the time server and improve the security of time synchronization.

Use a device as a client to synchronize time only from the specified external reference clock source to the local clock.

2. Topology

Figure 1-5 Configuring NTP Access Control Rights



3. Procedure

Set the IP address of GigabitEthernet 0/1 on device A to 192.168.2.2/24.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.2 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

Set the IP address of GigabitEthernet 0/1 on device B to 192.168.2.1/24.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Configure device A as the reference clock source of device B. Allow only the device with the IP address of 192.168.2.2 to send a time synchronization request to the local device.

```
DeviceB(cfg)# ntp-service server 192.168.2.2
DeviceB(cfg)# access-list 1 permit host 192.168.2.2
DeviceB(cfg)# ntp access-group serve-only 1
```

Configure device A as the NTP server, limit the request packet interval of the client, and send KoD packets.

```
DeviceA(cfg)# ip access-control-list standard limited1
DeviceA(cfg-stdacl)# 10 permit any
DeviceA(cfg)# exit
DeviceA(cfg)# ntp-service access-group limited limited1 kod
```

```
DeviceA(cfg)# ntp-service discard min-spacing 5 avg-spacing 5
DeviceA(cfg)# ntp-service master
```

4. Verification

Run the **ping** command to verify L3 reachability between device B and device A.

```
DeviceB# ping 192.168.2.2
Sending 5, 100-byte ICMP Echoes to 192.168.2.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the **clock set** command to modify the time on device A.

```
DeviceA# clock set 11:00:00
Set system clock: 11:00:00 UTC Fri, Feb 26, 2021
```

Run the **display clock** command on device B to check whether time synchronization is successful.

```
DeviceB# display clock
11:00:01 UTC Fri, Feb 26, 2021
```

5. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
ip access-control-list standard limited1
 10 permit any
!
interface GigabitEthernet 0/1
 ip address 192.168.2.2 255.255.255.0
ntp-service master
ntp-service access-group limited limited1 kod
!
end
```

- Device B configuration file

```
hostname DeviceB
!
ip access-control-list standard 1
 10 permit 192.168.2.2
!
ntp-service server 192.168.2.2
ntp-service access-group serve-only 1
!
interface GigabitEthernet 0/1
 ip address 192.168.2.1 255.255.255.0
!
end
```

Contents

1 Configuring SNTP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 2
1.2 Restrictions and Guidelines 2
1.3 Configuring SNTP 2
1.3.1 Overview 2
1.3.2 Restrictions and Guidelines 2
1.3.3 Procedure 2
1.4 Monitoring 3
1.5 Configuration Examples 3
1.5.1 Configuring Basic Functions of SNTP 3

1 Configuring SNTP

1.1 Introduction

1.1.1 Overview

NTP uses a complex algorithm for time synchronization, and raises high requirements for the system. As a simplified version of NTP, Simple Network Time Protocol (SNTP), inherits the protocol specification and implementation process of NTP, and adopts the same packet format and interaction mode as NTP. By setting some fields in the packet to predetermined values, SNTP simplifies the time synchronization algorithm and lowers the requirements for system performance. Although the precision of SNTP time synchronization is compromised while the algorithm is simplified, SNTP can generally achieve a precision of about 1s. Therefore, when the time precision requirements are not very high, you can choose to use SNTP. SNTP packets are the same as NTP packets, and therefore the SNTP client is fully compatible with the NTP server.

1.1.2 Principles

Like NTP, SNTP also uses the client/server mode for SNTP/NTP message interaction. The client sends a time synchronization packet to the server at a certain interval (half an hour by default). After receiving a response packet from the server, the client synchronizes time.

As shown in Figure 1-1, device B (B for short) acts as an NTP reference clock source, device A (A for short) acts as an SNTP client that synchronizes time with B. At a time point, the local clock of A is 19:00:00 and the local clock of B is 19:30:20.

(1) A sends an SNTP/NTP request packet. The local time (T0) when the packet leaves from A is 19:00:00 and is filled in Originate Timestamp.

(2) After a network delay of 2s, the local time (T1) when B receives the request packet is 19:30:23 and is filled in Receive Timestamp.

(3) B processes the NTP request and sends an NTP response packet one second later. The local time (T2) when the response packet leaves from B is 19:30:24 and is filled in Transmit Timestamp.

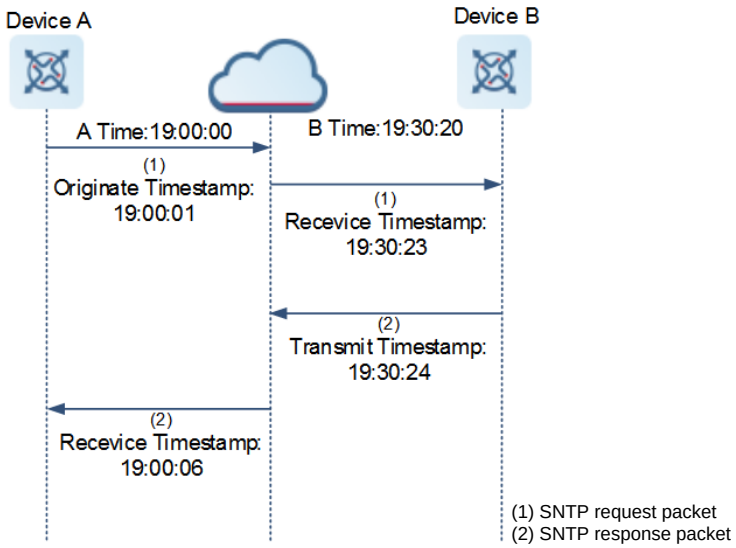
(4) After a network delay of 2s, A receives the response packet. The local time (T3) when the response packet arrives at A is 19:00:06.

The specific calculations for time synchronization are as follows:

a A obtains the time difference of 30 minutes and 20 seconds between B and A by using the formula ((T1-T0)+(T2-T3))/2.

b A obtains the packet round-trip delay of 4 seconds between A and B by using the formula (T3-T0)-(T2-T1).

Figure 1-1 SNTP Principles



1.1.3 Protocols and Standards

- RFC 2030: Simple Network Time Protocol (SNTP) Version 4 for IPv4, IPv6 and OSI

1.2 Restrictions and Guidelines

You cannot configure both the NTP and SNTP functions on the same device.

1.3 Configuring SNTP

1.3.1 Overview

An SNTP client accesses an NTP server regularly to calibrate the clock.

1.3.2 Restrictions and Guidelines

All time obtained through SNTP communication is Greenwich Mean Time (GMT). To obtain precise local time, you need to set the local time zone for alignment with GMT.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable SNTP.

sntp enable

The SNTP function is disabled by default.

(4) Configure the IP address of an SNTP server.

sntp server [**oob**] { *ipv4-address* | *domain* } [**via** *mgmt-name*] [**source** *source-ipv4-address*]

No SNTP server is configured by default.

(5) (Optional) Configure the SNTP time synchronization interval.

sntp interval *synchronization-interval-time*

The default interval for clock synchronization between an SNTP client and SNTP/NTP server is 1800s.

1.4 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 SNTP Monitoring

Command	Purpose
display sntp	Displays SNTP-related parameters.
debugging sntp	Debugs the DNS function.

1.5 Configuration Examples

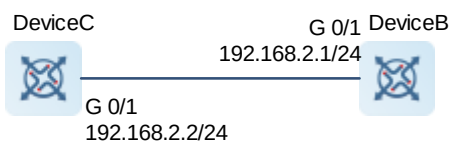
1.5.1 Configuring Basic Functions of SNTP

1. Requirements

- Use the local clock of device B as the NTP reference clock source to provide time synchronization for device C.
- Use device C as an SNTP client to synchronize time from device B.

2. Topology

Figure 1-2 Configuring Basic Functions of SNTP



3. Notes

- Configure device B as the SNTP server of device C.
- Enable the SNTP function on device C.

4. Procedure

Set the IP address of GigabitEthernet 0/1 on device B to 192.168.2.1/24.

```
DeviceB> privilege
DeviceB# config-system terminal
DeviceB(cfg)# interface gigabitEthernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceB(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.1 255.255.255.0
DeviceB(cfg-if-GigabitEthernet 0/1)# exit
```

Set the IP address of GigabitEthernet 0/1 on device C to 192.168.2.2/24.

```
DeviceC> privilege
DeviceC# config-system terminal
DeviceC(cfg)# interface gigabitEthernet 0/1
DeviceB(cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceC(cfg-if-GigabitEthernet 0/1)# ip address 192.168.2.2 255.255.255.0
DeviceC(cfg-if-GigabitEthernet 0/1)# exit
```

Configure device B as the NTP server.

```
DeviceB(cfg)# ntp master
```

Enable the SNTP function on device C.

```
DeviceC(cfg)# sntp server 192.168.2.1
DeviceC(cfg)# sntp enable
```

5. Verification

On device B, run the **ping** command to verify L3 reachability between device B and device C.

```
DeviceB# ping 192.168.2.2
Sending 5, 100-byte ICMP Echoes to 192.168.2.2, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

On device C, run the **display sntp** command to display the SNTP status and check whether the server is successfully configured.

```
DeviceC# display sntp
 SNTP state              : Enable
 SNTP oob mode           : Disable
 SNTP server             : 192.168.2.1
 SNTP sync interval      : 1800
 Time zone               : -0
```

Run the **clock set** command to modify the time on device B.

```
DeviceB# clock set 11:00:00
```

```
Set system clock: 11:00:00 UTC Fri, Feb 26, 2021
```

Run the **display clock** command on device C to check whether time synchronization is successful.

```
DeviceC# display clock
11:00:01 UTC Fri, Feb 26, 2021
```

6. Configuration Files

- Device B configuration file

```
DeviceB# diaplay current-config
hostname DeviceB
!
interface GigabitEthernet 0/1
 ip address 192.168.2.1 255.255.255.0
ntp master
!
end
```

- Device C configuration file

```
DeviceC# display current-config
hostname DeviceC
!
sntp server 192.168.2.1
sntp subscribe
!
interface GigabitEthernet 0/1
 ip address 192.168.2.2 255.255.255.0
!
end
```

Contents

1 Configuring the FTP Server 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 3
1.2 Configuration Task Summary 3
1.3 Configuring Basic Functions of the FTP Server 3
1.3.1 Overview 3
1.3.2 Procedure 3
1.4 Configuring FTP Session Management 4
1.4.1 Overview 4
1.4.2 Procedure 4
1.5 Configuring the FTP Lock Function 5
1.5.1 Overview 5
1.5.2 Restrictions and Guidelines 5
1.5.3 Procedure 5
1.6 Enabling AAA on the FTP Server 6
1.6.1 Overview 6
1.6.2 Restrictions and Guidelines 6
1.6.3 Prerequisites 6
1.6.4 Procedure 6
1.7 Configuring Read/Write Permission Control of FTP Users 6

1.7.1 Overview 6

1.7.2 Procedure 6

1.8 Monitoring 7

1.9 Configuration Examples 7

1.9.1 Providing the FTP Service in a LAN 7

1 Configuring the FTP Server

1.1 Introduction

1.1.1 Overview

File Transfer Protocol (FTP) is a standard application layer protocol formulated by the IETF Network Working Group. It implements file transfer based on Transmission Control Protocol (TCP). FTP is a major method for transferring files on the Internet.

FTP adopts the client/server model. As shown in Figure 1-1, a device can act as an FTP server to provide remote client access and program operations. By issuing commands to the server through the client, you can view the files in the FTP server directory, copy files from a remote computer to the local device, and transfer local files to the server. In addition, FTP provides functions, such as login authentication and read/write access control.

Figure 1-1 FTP Client/Server Mode



1.1.2 Principles

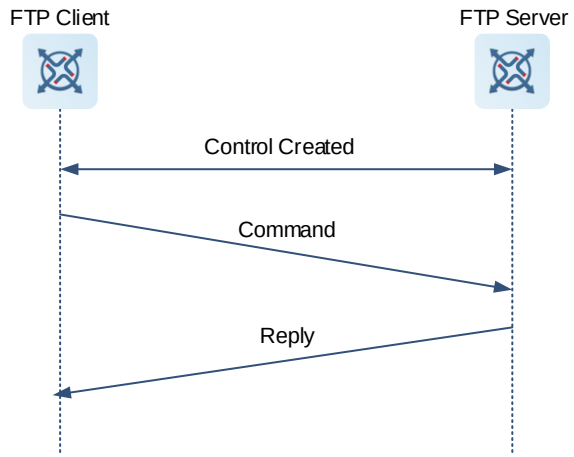
1. FTP Connection

FTP sets up two TCP connections between the client and the server, namely the control connection and the data connection.

- The control connection uses port 21 to transmit control commands, and is always in enabled state.
- The data connection uses port 20 to transmit data. The server can proactively enable or disable the data connection. It sets up a new data connection every time a file or directory is transferred.

For some simple sessions, only the control connection needs to be enabled, the client sends a command to the server, and the sever returns a response after receiving the command. The process is shown in Figure 1-2.

Figure 1-2 Control Connection



When the client sends a command for uploading or downloading data, both the control connection and data connection must be established.

2. User Login and Authentication

To access files on the FTP server, the FTP client must have a user account authorized by the FTP server. By default, a device supports configuration of up to 10 users, 2 connections per user, and 10 connections with the server.

In addition, the read and write permissions of files can be defined to effectively prevent unauthorized users from accessing internal files. When the number of a user's FTP login failures reaches the upper limit, the IP address or username used for login is locked. Within a certain period of time, all users cannot log in to the FTP server properly using this IP address or username.

3. Supported FTP Commands

After receiving an FTP connection request, the FTP server requires the client to provide the username and password for authentication.

If the client passes the authentication, the FTP client commands can be executed. The following table lists the FTP client commands supported by the FTP server.

Table 1-1 FTP Client Commands Supported by the FTP Server

ascii	close	mdelete	mput	quit	size
bin	delete	mdir	nlist	recv	system
bye	dir	mget	passive	rename	type
cd	get	mkdir	put	rmdir	user
cdup	ls	mls	pwd	send	

For how to use these FTP client commands, refer to the documentation of the FTP client software used. In addition, many FTP client tools, such as CuteFTP and FlashFXP, provide the graphic user interface (GUI), which allows users to perform operations without configuring FTP commands.

1.1.3 Protocols and Standards

- RFC 959: File Transfer Protocol (FTP)
- RFC 3659: Extensions to FTP
- RFC 2228: FTP Security Extensions
- RFC 2428: FTP Extensions for IPv6 and NATs
- RFC 1635: How to Use Anonymous FTP

1.2 Configuration Task Summary

FTP server configuration includes the following tasks:

(1) Configuring Basic Functions of the FTP Server

(2) (Optional) Configuring FTP Session Management

(3) (Optional) Configuring the FTP Lock Function

(4) (Optional) Enabling AAA on the FTP Server

(5) (Optional) Configuring Read/Write Permission Control of FTP Users

1.3 Configuring Basic Functions of the FTP Server

1.3.1 Overview

This section describes how to enable the FTP server and configure the username, password, and top-level directory to provide the FTP service for the FTP client.

1.3.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the FTP server.

ftp-server enable

The FTP server is disabled by default.

(4) Configure the top-level directory of the FTP server.

ftp-server topdir *directory*

By default, no top-level directory under which the FTP client can read and write files is configured, that is, the client is prohibited from accessing any directory on the FTP server.

(5) Configure a username and password for server login.

ftp-server username *username* [**privilege level**] **password** [*type*] *password*

By default, no username and password are configured for login to the FTP server, that is, no login users are restricted.

1.4 Configuring FTP Session Management

1.4.1 Overview

- The FTP login times refers to the number of times that the user's account and password can be verified when the user logs in to the FTP server.
- The FTP login timeout refers to the maximum time that the user can stay online after the username and password are verified. If the username and password are not verified again before login timeout, the session will be terminated to ensure that other users can log in to the FTP server.
- The idle timeout refers to the time from the completion of the last FTP operation to the start of the next FTP operation in an FTP session. After the server responds to an FTP client command (for example, after a file is completely transferred), the server starts to count the idle time again, and stops counting when the next FTP client command arrives. Therefore, the configuration of the idle timeout does not affect time-consuming file transfer operations.

1.4.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure FTP session management. Configure at least one of the following tasks.

- Configure the FTP login times.

  ftp-server login times *times*

  By default, the FTP login times is 1, that is, a session will be terminated once you enter an incorrect username or password so that other users can go online.

- Configure the FTP login timeout.

  ftp-server login timeout *time*

  By default, the FTP login timeout is 2 minutes.

- Configure the idle timeout.

  ftp-server timeout *time*

  By default, the idle timeout of an FTP session is 10 minutes.

- Configure the maximum number of FTP sessions.

  ftp-server max-session-num *number*

  By default, the maximum number of FTP sessions is 20.

1.5 Configuring the FTP Lock Function

1.5.1 Overview

- After the IP address lock function is enabled, if the cumulative number of login failures caused by entry of the incorrect username or password reaches the configured upper limit, the FTP session is terminated and the user's IP address is locked. In addition, all users cannot log in to the FTP server properly using this IP address or username.
- After the username lock function is enabled, if the cumulative number of login failures caused by entry of the incorrect password reaches the configured upper limit, the FTP session is terminated and the username is locked. The user cannot log in to the FTP server, but other users are not affected.
- The FTP lock time refers to the duration that a user needs to wait for automatic unlocking after the user is locked because the number of the user's login failures reaches the upper limit. The locked user can log in to the FTP server only after the login silence time expires.
- The maximum number of login verification times allowed by the FTP lock function refers to the maximum number of times that the username and password of a user can be verified after the FTP lock function is enabled. The default value is 3. If the number of verification failures exceeds the configured upper limit, the session is terminated and the user is locked.

1.5.2 Restrictions and Guidelines

After the IP address lock function is enabled, if the number of locked IP addresses reaches the upper limit, the full lock function is enabled. Then, the FTP server no longer accepts the connection request from any user until the number of locked IP addresses is smaller than the upper limit.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the FTP lock function. Configure at least one of the following tasks.

- Disable the IP address lock function.

  ftp-server login ip-block disable

  The IP address lock function is enabled by default.

- Configure the username lock function.

  ftp-server login username-block disable

  The username lock function is enabled by default.

- Configure the FTP lock time.

  ftp-server login silence-time *time*

  By default, the FTP lock time is 10 minutes.

- Configure the maximum number of IP addresses that can be locked after the IP address lock function is enabled.

ftp-server login max-block-limit *limit*

By default, up to 30 IP addresses can be locked.

- Configure the maximum number of login verification times allowed by the FTP lock function.

  ftp-server login block failed-times *times*

  By default, the maximum number of login verification times allowed by the FTP lock function is 3.

1.6 Enabling AAA on the FTP Server

1.6.1 Overview

By default, FTP does not support AAA login authentication. However, you can enable AAA on the FTP server through configuration.

1.6.2 Restrictions and Guidelines

This command takes effect only after FTP is enabled.

1.6.3 Prerequisites

Enable the AAA function before configuring this command. For details about how to enable AAA, see *Configuring AAA*.

1.6.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable AAA on the FTP server.

ftp-server authentication { **default** | *name* }

By default, FTP does not support AAA login authentication.

1.7 Configuring Read/Write Permission Control of FTP Users

1.7.1 Overview

After the user read/write permission control function is enabled, the configured FTP user levels or AAA username levels can be used to control read/write permissions. By default, users of all levels have the read/write permissions. After permission control is enabled, if the user level is not configured, it is set to 1 by default and the user has only the read permission, that is, the user can only download data. You can configure FTP users of different levels as required for read/write permission control.

1.7.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable read/write permission control of FTP users.

ftp-server login permission enable

By default, read/write permission control of FTP users is disabled.

1.8 Monitoring

Run the display commands to check the running status of a configured function to verify the configuration effect.

Run the debugging command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Run the cleanup commands to clear information.

Caution

Running the **cleanup** command may lose vital information and thus interrupt services.

Table 1-2 FTP Server Monitoring

Command	Purpose
cleanup ftp-server block-list { **all** | **ip-address** { *ipv4-address* | *ipv6-address* } [**vrf** *vrf-name*] | **username** *user* }	Clears the user lock entries in the FTP lock list.
display ftp-server	Displays the FTP server status information.
display ftp-server { **ip-block** | **username-block** } **list**	Displays the user lock entries in the FTP lock list.
debugging ftp-server err	Debugs the FTP server error events.
debugging ftp-server pro	Debugs the FTP server message events.

1.9 Configuration Examples

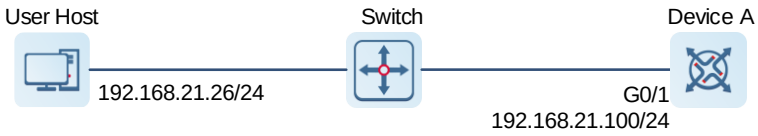
1.9.1 Providing the FTP Service in a LAN

1. Requirements

As shown in Figure 1-3, a device in a LAN acts as the FTP server to provide the upload and download services for users. The switch is an access device. The required session idle timeout is 5 minutes.

2. Topology

Figure 1-3 Configuring the FTP Service on a LAN



3. Notes

Configure device A as follows:

(1) Enable the FTP server function.

(2) Configure the top-level directory **/syslog**.

(3) Set the username to **user** and password to **password**.

(4) Set the session idle timeout to **5** minutes.

4. Procedure

(1) Configure device A.

a. Configure an IP address.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.168.21.100 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

b. Enable the FTP server function.

```
DeviceA(cfg)# ftp-server enable
```

c. Configure the top-level directory **/syslog**.

```
DeviceA(cfg)# ftp-server topdir /syslog
```

d. Set the username to **user** and password to **password**.

```
DeviceA(cfg)# ftp-server user-account user password password
```

e. Set the session idle timeout to **5** minutes.

```
DeviceA(cfg)# ftp-server timeout 5
```

(2) Configure the user host.

Set the IP address of the user host to 192.168.21.26/24.

5. Verification

Run the **display ftp-server** command to view the FTP server configurations.

```
DeviceA# display ftp-server
      ftp-server information
==================================
enable : Y
```

```
topdir : /syslog
timeout: 5min
total connect: 0
username:user
password:(PLAINT)password
connect num[0]
```

6. Configuration Files

- Device A configuration file

```
hostname DeviceA
!
interface GigabitEthernet 0/1
ip address 192.168.21.100 255.255.255
!
ftp-server enable
ftp-server topdir /syslog
ftp-server timeout 5
ftp-server username user password password
!
```

7. Common Errors

- No username is configured.
- No password is configured.
- No top-level directory is configured.

Contents

1 Configuring FTP Client 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 3
1.2 Configuration Task Summary 3
1.3 Enabling the FTP Client to Upload Files 4
1.3.1 Overview 4
1.3.2 Restrictions and Guidelines 4
1.3.3 Prerequisites 4
1.3.4 Procedure 4
1.4 Enabling the FTP Client to Download Files 4
1.4.1 Overview 4
1.4.2 Restrictions and Guidelines 4
1.4.3 Prerequisites 4
1.4.4 Procedure 5
1.5 Configuring the FTP Connection Mode 5
1.5.1 Overview 5
1.5.2 Restrictions and Guidelines 5
1.5.3 Procedure 5
1.6 Configuring the Transmission Mode of the FTP Client 5
1.6.1 Overview 5

1.6.2 Restrictions and Guidelines 6
1.6.3 Procedure 6
1.7 Configuring the Source IP Address for FTP Connection 6
1.7.1 Overview 6
1.7.2 Restrictions and Guidelines 6
1.7.3 Procedure 6
1.8 Monitoring 6
1.9 Configuration Examples 7
1.9.1 Configuring the FTP Client 7

1 Configuring FTP Client

1.1 Introduction

1.1.1 Overview

File Transfer Protocol (FTP) is an application of TCP/IP. By setting up a connection-oriented and reliable TCP connection between the FTP client and server, a user can access a remote computer that runs the FTP server program.

An FTP client enables file transfer between a device and the FTP server over FTP. A user uses the client to send a command to the server. The server responds to the command and sends the execution result to the client. By means of command interaction, the user can view files in the server directory, copy files from a remote computer to a local computer, and transfer local files to a remote computer.

FTP facilitates sharing of program/data files and remote use of computers. Users can transfer data in an efficient and reliable manner without caring about differences between files systems on different hosts.

You enter the copy command in the CLI to automatically run commands related to the control connection, such as **open**, **user**, and **pass**. After a control connection is established, the file transfer process starts. Then, a data connection is established to upload or download files.

Note

For the principles of the connection between the FTP client and server, see *Configuring FTP Server*.

The following sections introduce only the FTP client.

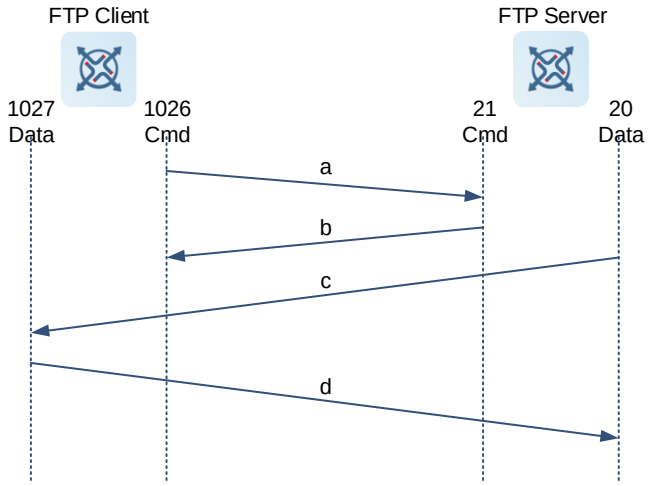
1.1.2 Principles

1. FTP Connection Mode

FTP supports two data connection modes: active (PORT) and passive (PASV). The two modes differ only in the way that a data connection is established, and the control connection is the same.

- In PORT mode, when a data connection is established, the FTP server initiates a request to establish a new connection with the client for data transmission. This mode cannot be used if the firewall is deployed on the FTP client. The connection establishment process is as follows:
  a The FTP client establishes a control channel through port 1026 and port 21 of the FTP server, and sends control commands through this channel.
  b The client sends the **PORT** command through this channel when it needs to receive data. The **PORT** command contains the port information (1027) of the client data channel.
  c When transmitting data, the server connects to port 1027 of the client through port 20 on the server to establish a data channel for data transmission and reception.
  d The client sends a response. Data transmission ends.

Figure 1-1 PORT Mode



- In PASV mode, the FTP client initiates a request when a data connection is established. The connection establishment process is as follows:

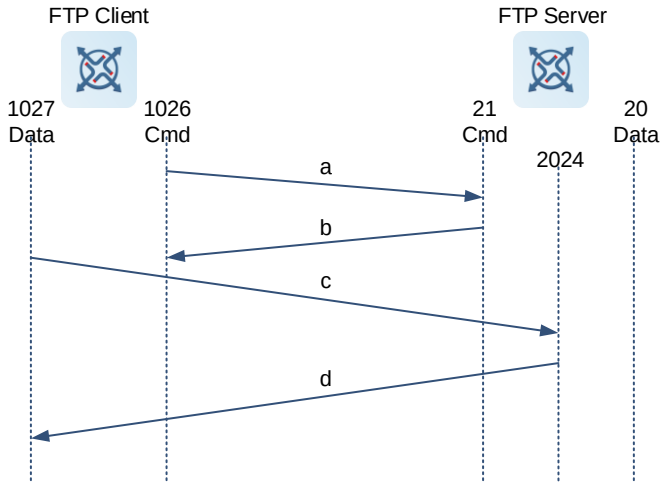
  a The FTP client establishes a control channel through port 1026 and port 21 of the FTP server, and sends control commands through this channel.

  b The client sends the **PASV** command through this channel when it needs to receive data.

  c After receiving the **PASV** command, the FTP server randomly opens a high-end port 2024 and notifies the client of data transmission to this port.

  d The client uses port 1027 to connect to port 2024 of the FTP server, and then sends or receives data on this channel. The FTP server does not need to establish a new connection with the client.

Figure 1-2 PASV Mode



Note

The control connection for command and feedback transmission is always present whereas the data connection is established as required. Only the FTP client can select the PASV or PORT mode. The FTP mode supported by an FTP client depends on the actual FTP client software. A Ruijie FTP client uses the PASV mode by default.

2. FTP File Transmission Modes

FTP provides two file transmission modes:

- Binary transmission mode: It is used to transfer program files (for example, .app, .bin and .btm files), executable files, compressed files, image files, and all ASCII values. In this transmission mode, data is not processed, and therefore the transmission rate is faster than that of the text transmission mode.
- Text transmission mode (ASCII mode): It is used to transfer text files (such as .txt, .bat, and .cfg files). This mode is different from the binary mode in carriage return and line feed processing. In ASCII mode, carriage return and line feed are converted to local CRC characters. For example, \n in Unix is \r\n in Windows or \r in Mac. Assume that a file copied by a user contains ASCII text. If a remote computer does not run Unix, FTP automatically converts the file format accepted by the remote computer during file transfer.

3. Specifying the Source IP Address for FTP Transmission

You can configure a source IP address for the FTP client and use it to communicate with an FTP server. With this IP address, the FTP client can connect to the server and share files with the server.

4. Checking Sizes of Files Downloaded from an FTP Server

When the FTP client downloads a file, it checks the file size by default to detect file transfer errors (if any). You can also disable the file size check when downloading files from FTP servers that cannot reply to FTP clients with file sizes.

1.1.3 Protocols and Standards

- RFC 959: FILE TRANSFER PROTOCOL (FTP)

1.2 Configuration Task Summary

All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring basic functions of the FTP client
  - Enabling the FTP Client to Upload Files
  - Enabling the FTP Client to Download Files
- Configuring extended functions of the FTP client
  - Configuring the FTP Connection Mode
  - Configuring the Transmission Mode of the FTP Client
  - Configuring the Source IP Address for FTP Connection

1.3 Enabling the FTP Client to Upload Files

1.3.1 Overview

You can use the FTP client to upload files to the FTP server.

1.3.2 Restrictions and Guidelines

During file transfer, do not insert or remove a storage medium or transmission medium to or from the device to avoid transmission errors.

When you transfer the file through the FTP client:

(1) The username is up to 32 characters long and cannot contain special characters such as : / @ or spaces. The username is required.

(2) The password is up to 32 characters long and cannot contain special characters such as : / @ or spaces. The password is required.

1.3.3 Prerequisites

The user must have an account with the read and write permissions for the FTP server, and can use this account to access the target folder on the FTP server.

1.3.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Upload files.

copy flash:*local-file* { **ftp: | oob_ftp:** }***//****account-information@remote-file*

1.4 Enabling the FTP Client to Download Files

1.4.1 Overview

You can use the FTP client to download files from the FTP server.

1.4.2 Restrictions and Guidelines

During file transfer, do not insert or remove a storage medium or transmission medium to or from the device to avoid transmission errors.

When you transfer the file through the FTP client:

(1) The username is up to 32 characters long and cannot contain special characters such as : / @ or spaces. The username is required.

(2) The password is up to 32 characters long and cannot contain special characters such as : / @ or spaces. The password is required.

1.4.3 Prerequisites

The user must have an account with the read permission for the FTP server, and can use this account to access the target files on the FTP server.

1.4.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) (Optional) Disable the size check of files downloaded from the FTP server.

- Enter the global configuration mode.

  config-system terminal

- Disable the size check of files downloaded from the FTP server.

  ftp-client disable-size-check

  By default, the sizes of files downloaded from the FTP server are checked.

- Exit the global configuration mode.

  exit

(3) Download files.

copy { **ftp:** | **oob_ftp:** }***//****account-information@remote-file* **flash:***local-file*

1.5 Configuring the FTP Connection Mode

1.5.1 Overview

You can download or upload files in the specified FTP connection mode. The default FTP connection mode is PASV. You can set the FTP connection mode to PORT, in which the FTP server initiates a connection request to the client.

1.5.2 Restrictions and Guidelines

If you need to specify the **vrf** *vrf-name* parameter when configuring the FTP client, configure the VRF first. For details about the VRF, see *Configuring VRF*.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Set the connection mode to **PORT**.

ftp-client [**vrf** *vrf-name*] **port**

The default FTP connection mode is PASV.

1.6 Configuring the Transmission Mode of the FTP Client

1.6.1 Overview

You can download or upload files in the specified transmission mode of the FTP client. The default transmission mode is Binary. You can also set the transmission mode to ASCII according to your requirements.

1.6.2 Restrictions and Guidelines

If you need to specify the **vrf** *vrf-name* parameter when configuring the FTP client, configure the VRF first. For details about the VRF, see *Configuring VRF*.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Set the transmission mode to ASCII.

ftp-client [**vrf** *vrf-name*] **ascii**

The default transmission mode is Binary.

1.7 Configuring the Source IP Address for FTP Connection

1.7.1 Overview

You can download or upload files using the specified IP address of the FTP client. The FTP client is bound to the IP address of a port so that the client can use this IP address to connect to the server.

1.7.2 Restrictions and Guidelines

If you need to specify the **vrf** *vrf-name* parameter when configuring the FTP client, configure the VRF first. For details about the VRF, see *Configuring VRF*.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the source IP address for FTP connection.

ftp-client [**vrf** *vrf-name*] **source** { *ipv4-address* | *ipv6-address* | *interface-type interface-number* }

By default, no source IP address is bound to the FTP client. Instead, an IP address is selected for the client based on the route.

1.8 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 Monitoring

Command	Purpose
display ftp-client	Displays the FTP client configurations.
debugging ftp-client	Debugs the FTP client.

1.9 Configuration Examples

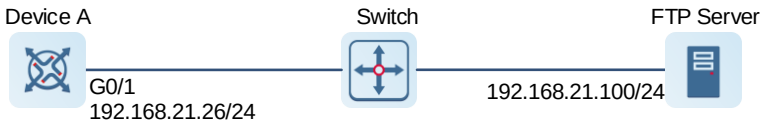
1.9.1 Configuring the FTP Client

1. Requirements

As shown in Figure 1-3, in a LAN, a local device needs to share files with a remote FTP server, and FTP is required to work in the specified connection mode and transmission mode.

2. Topology

Figure 1-3 Configuring the FTP Client



3. Notes

- Set the FTP connection mode to PORT.
- Set the transmission mode to ASCII.
- Use the username **user** and password **pass** to log in to the FTP server, upload the **local-file** file in the **home** directory of the local device to the **root** directory of the FTP server, and rename the file **remote-file**.
- Use the username **user** and password **pass** to log in to the FTP server, download the **remote-file** file from the **root** directory of the FTP server to the **home** directory of the local device, and rename this file **local-file**.

4. Procedure

(1) Configure the FTP server.

Set the IP address of the FTP server to 192.168.21.100/24.

(2) Configure device A.

a. Configure the IP address of device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.168.21.26 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

b. Set the transmission mode to ASCII.

```
DeviceA(cfg)# ftp-client ascii
```

c. Set the FTP connection mode to PORT.

```
DeviceA(cfg)# ftp-client port
DeviceA(cfg)# exit
```

d. Upload the file.

```
DeviceA# copy flash: home/local-file ftp://user:pass@192.168.21.100/root/remote-file
```

e. Download the file.

```
DeviceA# copy ftp://user:pass@192.168.21.100/root/remote-file flash: home/local-file
```

5. Verification

Run the **display ftp-client** command to view the FTP client configurations.

```
DeviceA# display ftp-client
        ftp-client information
==================================
type: ASCII
mode: PORT
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
 interface GgigabitEethernet 0/1
 ip address 192.168.21.26 255.255.255
!
 ftp-client ascii
 ftp-client port
!
```

7. Common Errors

- The command formats for uploading and downloading are incorrect.
- The username or password is incorrect.

Contents

1 Configuring TFTP Server 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 3
1.2 Configuring Basic Functions of the TFTP Server 3
1.2.1 Overview 3
1.2.2 Procedure 3
1.3 Monitoring 4
1.4 Configuration Examples 4
1.4.1 Configuring the TFTP Server 4

1 Configuring TFTP Server

1.1 Introduction

1.1.1 Overview

Trivial File Transfer Protocol (TFTP) is a standard application layer protocol formulated by the IETF Network Working Group. TFTP uses User Datagram Protocol (UDP) as its transport protocol, and the well known UDP port for TFTP traffic is 69.

TFTP provides the uncomplicated and inexpensive file transfer service. Many commonly used FTP features are unavailable in TFTP. TFTP can only upload or download files, but not list directories, perform authentication, or provide a security mechanism. TFTP employs a timeout and retransmission strategy to ensure data transmission. TFTP is described in RFC1350.

A TFTP client can conveniently download files (such as upgrade packages) on the device from the TFTP server, or upload files directly to the file system of the TFTP server.

1.1.2 Principles

1. TFTP Packet

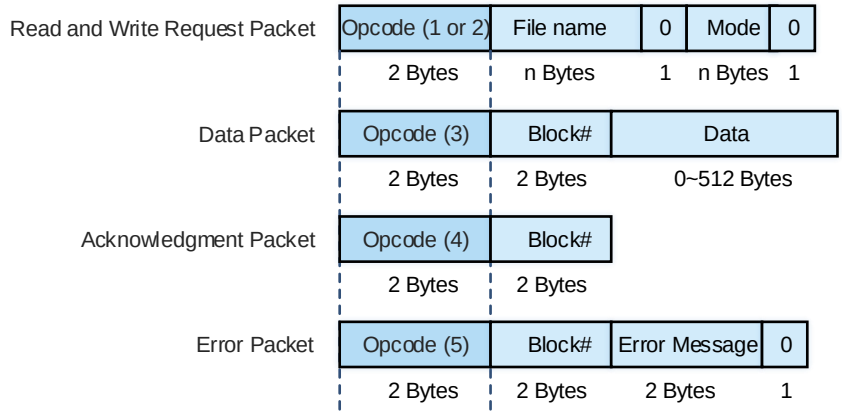
After the TFTP client initiates a file read or write request, and the TFTP server approves the request, TFTP transmits the data in 512-byte packets. A packet with less than 512 bytes indicates that the transmission is completed.

Each packet contains a data block, and the next packet can be sent only after an acknowledgment (ACK) packet is received from the peer. If the ACK packet is not received within the specified time, the last packet sent is retransmitted.

The header of a TFTP packet contains an Opcode field, which indicates the packet type. Opcodes 1 to 5 represent five types of packets:

- 1: Read Request (RRQ)
- 2: Write Request (WRQ)
- 3: Data (DATA)
- 4: Acknowledgment (ACK)
- 5: Error (ERROR)

Figure 1-1 TFTP Packet

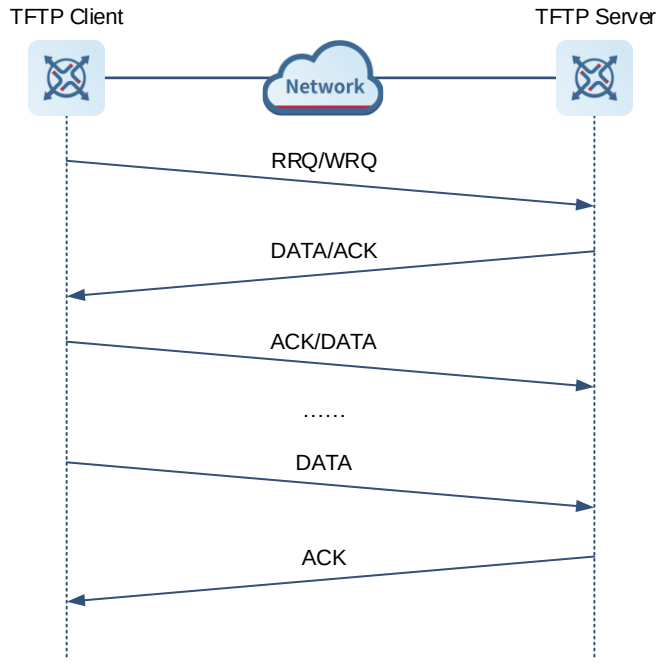


The Error Message in an error packet contains 2 bytes in ASCII code. TFTP provides the following error codes:

- 0: Not defined;
- 1: File not found;
- 2: Access violation;
- 3: Disk full;
- 4: Illegal TFTP operation;
- 5: Unknown port;
- 6: File already exists;
- 7: No such user.

2. TFTP Working Modes

Figure 1-2 TFTP Working Modes



(1) The TFTP client initiates a read request (RRQ) or write request (WRQ) to the TFTP server.

(2) If the TFTP server receives an RRQ, it checks whether the read conditions are met (whether the file exists and whether the TFTP client has the access permission). If yes, the TFTP server sends a data packet (DATA) to the TFTP client. If the TFTP server receives a WRQ, it checks whether the write conditions are met (whether there is enough space and whether the FTP client has the write permission). If yes, the TFTP server sends an ACK packet to the TFTP client.

(3) If the TFTP client receives a DATA packet when preparing to download a file, it replies with an ACK packet. If the TFTP client receives an ACK packet when preparing to upload a file, it continues to send the DATA packet.

(4) The process of sending the ACK packet repeats until the last DATA packet with less than 512 bytes is received and indicates the end of the transmission.

(5) If an error is encountered in the transmission process, an ERROR packet is sent to the peer.

3. TFTP File Transmission Modes

TFTP supports the following transmission modes:

- Binary mode for transmitting program files
- ASCII mode for transmitting text files

1.1.3 Protocols and Standards

- RFC 1350: The TFTP Protocol (revision 2)
- RFC 2347: TFTP Option Extension
- RFC 2348: TFTP Blocksize Option
- RFC 2349: TFTP Timeout Interval and Transfer Size Options

1.2 Configuring Basic Functions of the TFTP Server

1.2.1 Overview

This section describes how to configure the TFTP server to provide the upload and download service for the TFTP client.

1.2.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the top-level directory of the TFTP server.

tftp-server topdir *directory*

By default, the TFTP client can read and write files in the Flash directory.

(4) Enable the TFTP server function.

tftp-server enable

The TFTP server is disabled by default. After the TFTP server is enabled, files can be transmitted only through the MGMT port.

1.3 Monitoring

Run the **display** commands to check the running status of a configured function to verify the configuration effect.

Run the **debugging** command to output debugging information.

Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 Monitoring

Command	Purpose
display tftp-server	Displays the TFTP server configurations.
display tftp-server updating-list	Displays the file download progress on the client.
debugging tftp-server	Debugs the server.

1.4 Configuration Examples

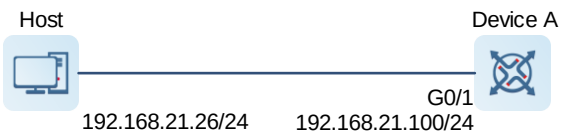
1.4.1 Configuring the TFTP Server

1. Requirements

Configure a TFTP server in a LAN so that users can upload and download files.

2. Topology

Figure 1-3 Configuring the TFTP Server



3. Notes

Configure device A as follows:

(1) Enable the TFTP server.

(2) Configure the top-level directory.

4. Procedure

(1) Configure device A.

Configure the IP address of the Ethernet interface.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-gigabitethernet 0/1)# ip address 192.168.21.100 255.255.255.0
DeviceA(cfg-if-gigabitethernet 0/1)# exit
```

Configure the top-level directory of the TFTP server.

```
DeviceA(cfg)# tftp-server topdir /tmp
```

Enable the FTP server function.

```
DeviceA(cfg)# tftp-server enable
```

(2) Configure the host.

Set the IP address of the host to 192.168.21.26/24.

5. Verification

Run the **display tftp-server** command to view the FTP server configurations.

```
DeviceA# display tftp-server
      tftp-server information
==================================
enable : Y
topdir : /tmp
```

6. Configuration Files

TFTP server configuration file

```
hostname DeviceA
!
interface gigabitethernet 0/1
ip address 192.168.21.100 255.255.255.0
!
tftp-server enable
tftp-server topdir /tmp
!
```

7. Common Errors

- No top-level directory is configured.

Contents

1 Configuring TFTP Client....1
1.1 Introduction....1
1.1.1 Overview....1
1.1.2 Principles....1
1.1.3 Protocols and Standards....1
1.2 Configuration Task Summary....1
1.3 Enabling the TFTP Client to Upload Files....2
1.3.1 Overview....2
1.3.2 Restrictions and Guidelines....2
1.3.3 Procedure....2
1.4 Enabling the TFTP Client to Download Files....2
1.4.1 Overview....2
1.4.2 Restrictions and Guidelines....2
1.4.3 Procedure....2
1.5 Binding the Source IP Address to the TFTP Client....2
1.5.1 Overview....2
1.5.2 Procedure....2
1.6 Configuring a Port Number for TFTP Connection....3
1.6.1 Overview....3
1.6.2 Restrictions and Guidelines....3
1.6.3 Procedure....3
1.7 Monitoring....3

1.8 Configuration Examples....4

1.8.1 Configuring the TFTP Client....4

1 Configuring TFTP Client

1.1 Introduction

1.1.1 Overview

Trivial File Transfer Protocol (TFTP) is a specific application of Transmission Control Protocol/Internet Protocol (TCP/IP). File transfer between TFTP client and server is implemented based on User Datagram Protocol (UDP), and the default port number is **69**. Compared with the TCP-based FTP protocol, TFTP does not require authentication or have complex packets. It is suitable for a stable network environment and small file transfer.

Note

TFTP is suitable for small file transfer, and FTP supports transfer of large files.

1.1.2 Principles

1. Specifying the Source Interface IP Address for TFTP Transmission

You can configure a source IP address for the TFTP client and use it to communicate with a TFTP server. With this IP address, the TFTP client can connect to the server and share files with the server. Specifying the egress interface of packets facilitates interface management.

2. Specifying a Port Number for Connection with the TFTP Server

You can specify a port number on the TFTP client to connect with the TFTP server and transfer files. Specifying a port number also prevents attacks.

1.1.3 Protocols and Standards

- RFC 783: Trivial FILE TRANSFER PROTOCOL (TFTP)

1.2 Configuration Task Summary

All the configuration tasks below are optional. Select the configuration tasks as required.

- Basic functions of the TFTP client
  - Enabling the TFTP Client to Upload Files
  - Enabling the TFTP Client to Download Files
- Extended functions of the TFTP client
  - Binding the Source IP Address to the TFTP Client
  - Configuring a Port Number for TFTP Connection

1.3 Enabling the TFTP Client to Upload Files

1.3.1 Overview

You can upload files from the TFTP client to the TFTP server.

1.3.2 Restrictions and Guidelines

During file transfer, do not insert or remove a storage medium or transmission medium to or from the device to avoid transmission errors.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Upload files.

copy flash:*local-file* { **tftp: | oob_tftp:** }***//****remote-file*

1.4 Enabling the TFTP Client to Download Files

1.4.1 Overview

You can download files from the TFTP server to the TFTP client.

1.4.2 Restrictions and Guidelines

During file transfer, do not insert or remove a storage medium or transmission medium to or from the device to avoid transmission errors.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Download files.

copy { **tftp:** | **oob_tftp:** }***//****remote-file* **flash:***local-file*

1.5 Binding the Source IP Address to the TFTP Client

1.5.1 Overview

You can bind a source IP address to the TFTP client so that the client can use this IP address to communicate with the server for file upload or download.

1.5.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Bind a source IP address to the TFTP client.

tftp-client source { **ip** *ipv4-address* | **ipv6** *ipv6-address* | interface-type interface-number }

By default, no source IP address is bound to the TFTP client. Instead, an IP address is selected for the client based on the route.

1.6 Configuring a Port Number for TFTP Connection

1.6.1 Overview

You can specify the port number used by the TFTP client to connect with the TFTP server for file upload and download.

1.6.2 Restrictions and Guidelines

The range of the port number configured for the TFTP client is from 20000 to 65534.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the port number used by the TFTP client to connect with the TFTP server.

tftp-client port *port-number*

The default port number used by the TFTP client to connect with the TFTP server is **69**.

1.7 Monitoring

Run the **debugging** command to output debugging information.

⚠ Caution

The output debugging information occupies system resources. Therefore, disable the debugging function immediately after use.

Table 1-1 Monitoring

Command	Purpose
debugging tftp	Debugs the TFTP client.

1.8 Configuration Examples

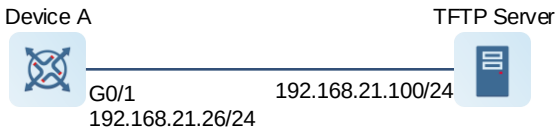
1.8.1 Configuring the TFTP Client

1. Requirements

A local device needs to upload a file to a remote TFTP server, and download another file from the TFTP server.

2. Topology

Figure 1-1 Configuring the TFTP Client



3. Notes

- Upload the **local-file** file in the **Flash** directory on device A to the **root** directory of the TFTP server with an IP address 192.168.21.100, and rename the file **remote-file**.
- Download the **remote-file** file from the **root** directory of the TFTP server with the IP address 192.168.21.100 to the **Flash** directory on device A, and save the file as **download-file**.

4. Procedure

(1) Configure the TFTP server.

Set the IP address of the TFTP Server to 192.168.21.100/24.

(2) Configure device A.

a. Configure an IP address.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitethernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# unset switch-port
DeviceA(cfg-if-GigabitEthernet 0/1)# ip address 192.168.21.26 255.255.255
DeviceA(cfg-if-GigabitEthernet 0/1)# exit
```

b. Upload the file.

```
DeviceA# copy flash:local-file tftp://192.168.23.69/root/remote-file
```

c. Download the file.

```
DeviceA# copy tftp://192.168.23.69/root/remote-file flash: download-file
```

5. Verification

Check whether the **remote-file** file is configured on the TFTP server.

Run the **dir** command on device A to check whether the **download-file** file is configured in the **home** directory of the Flash.

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
 interface GigabitEthernet 0/1
 ip address 192.168.21.26 255.255.255.0
!
```

7. Common Errors

- The format of the uploaded or downloaded file is incorrect.

Contents

1 Configuring SNMP 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 SNMP Versions 1
1.1.3 Principles 1
1.1.4 Protocols and Standards 6
1.2 Configuration Task Summary 6
1.3 Configuring Basic SNMP Features 7
1.3.1 Overview 7
1.3.2 Restrictions and Guidelines 7
1.3.3 Procedure 7
1.4 Configuring the Trap Function 9
1.4.1 Overview 9
1.4.2 Restrictions and Guidelines 9
1.4.3 Procedure 9
1.5 Configuring the Agent Shielding Function 11
1.5.1 Overview 11
1.5.2 Restrictions and Guidelines 11
1.5.3 Procedure 11
1.6 Configuring SNMP Control Parameters 11
1.6.1 Overview 11
1.6.2 Restrictions and Guidelines 12

1.6.3 Procedure....12
1.7 Monitoring....13
1.8 Configuration Examples....13
1.8.1 Configuring Basic SNMPv2c Function....13
1.8.2 Configuring Basic SNMPv3 Function....15
1.8.3 Configuring Basic Trap Function....18

1 Configuring SNMP

1.1 Introduction

1.1.1 Overview

The Simple Network Management Protocol (SNMP) is used for network monitoring and management. Because many vendors support SNMP, SNMP has become a network management standard and applies to the interconnected environment of systems of multiple vendors. By using SNMP, the network administrator can implement information query for network nodes, network configuration, fault locating, and capacity planning.

1.1.2 SNMP Versions

- SNMPv1

  SNMPv1 is the first official version of SNMP and is defined in RFC 1157. Based on community names, SNMPv1 has simple structure of management information (SMI) and simple management information base (MIB) and lower security.

- SNMPv2C

  SNMPv2 is a community-based SNMP management architecture and is defined in RFC 1901. SNMPv2 is compatible with SNMPv1. Compared with SNMPv1, SNMPv2 has two more protocol operations Get-Bulk and Inform and supports more data types and error code information.

- SNMPv3

  SNMPv3 defines security extension capability and provides the following security features by identifying and encrypting data.

  - Ensuring that data is not tampered during transmission.
  - Ensuring that data is transmitted from legal data sources.
  - Encrypting packets and ensuring data confidentiality.

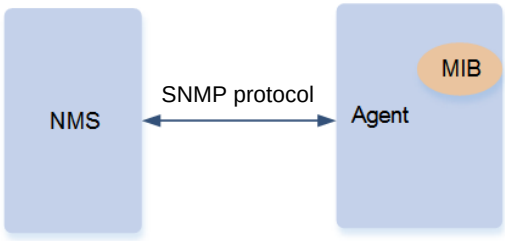
1.1.3 Principles

1. Basic Concepts

The SNMP management system has the following components:

- SNMP network manager
- SNMP agent
- MIB

Figure 1-1 Relationship Between the NMS and Network Management Agent



- SNMP network manager

  The SNMP network manager is a system that controls and monitors the network based on SNMP and is also called the network management system (NMS).

- SNMP agent

  The SNMP agent (" agent") is software running on the managed devices. It is responsible for receiving, processing, and responding to monitoring and control packets from the NMS. The agent may also actively send messages to the NMS.

- MIB

  The MIB is a virtual network management information base. The managed network devices contain lots of information. To uniquely identify a specific management unit among SNMP packets, the MIB adopts the tree hierarchical structure. Nodes in the tree indicate specific management units. A string of digits are used to uniquely identify a management unit system among network devices. The MIB is a set of unit identifiers of network devices.

- Operation types

  Six operation types are defined for information exchange between the NMS and the agent based on SNMP:

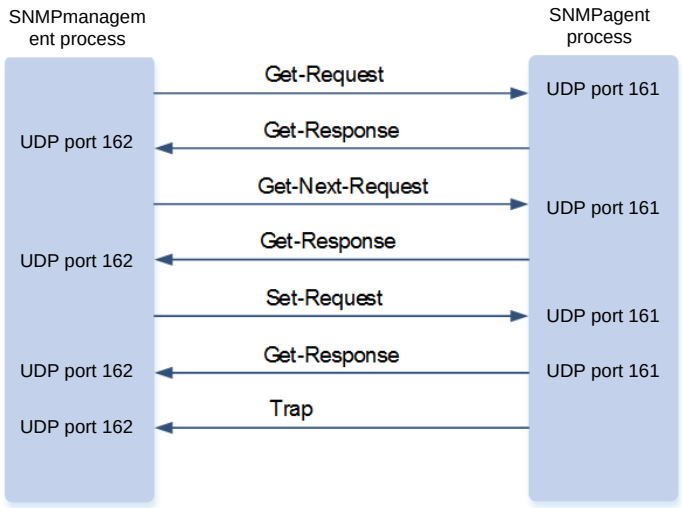
  - Get-Request: The NMS extracts one or more parameter values from the agent.
  - Get-Next-Request: The NMS extracts the parameter value next to one or more parameters from the agent.
  - Get-Bulk: The NMS extracts a batch of parameter values from the agent.
  - Set-Request: The NMS sets one or more parameter values of the agent.
  - Get-Response: The agent returns one or more parameter values, which are the operations of the agent in response to the three operations performed by the NMS.
  - Trap: The agent actively sends a message to notify the NMS of the situation happening in the agent.

  The first four packets are sent by the NMS to the agent and the last two packets are sent by the agent to the NMS. The first three operations of the NMS and the response operations of the agent are based on UDP port 161. The Trap operation performed by the agent is based on UDP port 162. For details, see Figure 1-2.

Caution

SNMPv1 does not support the Get-Bulk operation.

Figure 1-2 SNMP Packet Interaction



2. Basic SNMP Functions

SNMP interaction is response interaction (for interaction of packets, see Figure 1-2). The NMS actively sends requests to the agent, including Get-request, Get-next-request, Get-bulk, and Set-request. The agent receives the requests, completes operations, and returns a Get-response. Sometimes, the agent actively sends a Trap message and an Inform message to the NMS. The NMS does not need to respond to the Trap message but needs to return an Inform-response to the agent. Otherwise, the agent re-sends the Inform message.

3. Principles of SNMPv1 and SNMPv2C

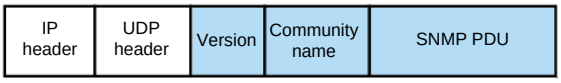
SNMPv1 and SNMPv2C adopt the community-based security architecture. The administrator who can perform operations on the MIB of the agent is limited by defining the host address and authentication name (community string).

SNMPv1 and SNMPv2 determine whether the administrator has the right to use MIB objects by using the authentication name. To be able to manage devices, the NMS must ensure that its authentication name is the same as an authentication name defined in devices.

In SNMPv2C, the Get-bulk operation enables more detailed error message types to be returned to the NMS. The Get-bulk operation allows all information from a table or lots of data to be obtained at a time, so as to reduce the request-response times. The enhanced error handling capabilities of SNMPv2C include extension of error codes to differentiate error types. In SNMPv1, however, only one error code is provided for the errors. In SNMPv2C, the error codes can be used to differentiate error types. Because management workstations on the network may support SNMPv1 and SNMPv2C at the same time, the SNMP agent must be able to identify SNMPv1 packets and SNMPv2C packets and return packets of the corresponding versions.

SNMPv1 and SNMPv2c packets are composed of version, community name, and SNMP protocol data unit (PDU). Figure 1-3 shows the structure of the packets.

Figure 1-3 Structure of SNMPv1 and SNMPv2c Packets



- The main fields in packets are as follows:
  - Version: Indicates an SNMP version. For SNMPv1 packets, the corresponding field value is 0. For SNMPv2c packets, the field value is 1.
  - Community name: It is used for authentication between the agent and NMS. The field value is a string that can be customized. The community names include readable and writable community names. During Get-Request and Get-Next-Request operations, a readable community name is used for authentication. During Set operations, a writable community name is used for authentication.
  - SNMP PDU: Includes information such as PDU type, request identifier, and variable binding list. SNMPv1 PDUs include GetRequest PDU, GetNextRequest PDU, SetRequest PDU, Response PDU, and Trap PDU. Compared with SNMPv1, SNMPv2c has GetBulkRequest PDU and InformRequest PDU.
- Security

  An authentication name has the following attributes:
  - Read-only: Provides the read permission of all MIB variables for authorized management workstations.
  - Read-write: Provides the read/write permission of all MIB variables for authorized management workstations.

4. Principles of SNMPv3

SNMPv3 redefines the SNMP architecture and includes functions of SNMPv1 and SNMPv2 into the SNMPv3 system.

The NMS and SNMP agent are SNMP entities. In the SNMPv3 architecture, SNMP entities consist of the SNMP engine and SNMP applications. The SNMP engine is used to send and receive messages, identify and encrypt information, and control access to managed objects. SNMP applications refer to internal applications of SNMP, which work by using the services provided by the SNMP engine.

SNMPv3v determines whether a user has the right to use MIB objects by using the user-based security model (USM). The security level of the NMS user must be the same as that of a specified SNMP user defined in devices so as to manage devices.

SNMPv3 requires the NMS to obtain the SNMP agent engine IDs on devices when the NMS manages devices. SNMPv3 defines the Discover and Report operation mechanisms. When the NMS does not know the agent engine ID, the NMS may first send a Discover message to the agent and the agent returns a Report message carrying an engine ID. Later, management operations between the NMS and the agent must carry the engine ID.

SNMPv3 defines a new packet format, which is shown in Figure 1-4.

Figure 1-4 Structure of SNMPv3 Packets



- The main fields in SNMP packets are as follows:
  - Version: Indicates an SNMP version. For SNMPv3 packets, the field value is 2.

- Header data: Mainly includes description information such as the maximum message size supported by the message sender and the security model of the message.
- Security parameter: Includes such security information as SNMP engine-related information, username, authentication parameter, and encryption parameter.
- Context EngineID: A unique identifier of SNMP. This field and the PDU type decide the application to which the packets are sent.
- Context name: Decides the MIB view of the managed devices based on the Context EngineID.
- SNMP PDU: Includes information such as PDU type, request identifier, and variable binding list. SNMPv3 PDUs include GetRequest PDU, GetNextRequest PDU, SetRequest PDU, Response PDU, Trap PDU, GetBulkRequest PDU, and InformRequest PDU.

- Security

  SNMPv3 determines the data security mechanism based on the security model and security level. At present, security models include SNMPv1, SNMPv2C, and SNMPv3. SNMPv3 includes SNMPv1 and SNMPv2C into the security model.

Table 1-1 Security Models and Security Levels of SNMPv1 and SNMPv2C

Security Model	Security Level	Authentication	Encryption
SNMPv1	noAuthNoPriv	Authentication name	N/A
SNMPv2c	noAuthNoPriv	Authentication name	N/A

Table 1-2 Security Model and Security Level of SNMPv3

Security Model	Security Level	Authentication	Encryption	Description
SNMPv3	noAuthNoPriv	Username	N/A	Data validity is confirmed through username.
SNMPv3	authNoPriv	MD5 or SHA	N/A	The data authentication mechanism based on HMAC-MD5 or HMAC-SHA is provided.
SNMPv3	authPriv	MD5 or SHA	DES	The data authentication mechanism based on HMAC-MD5 or HMAC-SHA and the data encryption mechanism based on CBC-DES are provided.

- Engine ID

  An engine ID is used to uniquely identify an SNMP engine. Because each SNMP entity includes only one SNMP engine, the SNMP engine uniquely identifies an SNMP entity in a management domain. Therefore, the SNMPv3 agent as an entity must has a unique engine ID, that is, SnmpEngineID.

  An engine ID is an octet string that consists of 5 to 32 bytes. RFC 3411 defines the format of an engine ID:

  - The first four bytes indicate the private enterprise ID (allocated by the Internet Assigned Numbers Authority (IANA)) of a vendor, which is expressed in hexadecimal.

- The fifth byte indicates remaining bytes.
- 0: Reserved.
- 1: The later four bytes indicate an IPv4 address.
- 2: The later 16 bytes indicate an IPv6 address.
- 3: The later six bytes indicate a MAC address.
- 4: Text consisting of up to 27 bytes, which is defined by the vendor.
- 5: Hexadecimal value consisting of up to 27 bytes, which is defined by the vendor.
- 6-127: Reserved.
- 128-255: Formats specified by the vendor.

1.1.4 Protocols and Standards

- RFC 1157, Simple Network Management Protocol (SNMP)
- RFC 1901, Introduction to Community-based SNMPv2
- RFC 2578, Structure of Management Information Version 2 (SMIv2)
- RFC 2579, Textual Conventions for SMIv2
- RFC 3411, An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks
- RFC 3412, Message Processing and Dispatching for the Simple Network Management Protocol (SNMP)
- RFC 3413, Simple Network Management Protocol (SNMP) Applications
- RFC 3414, User-based Security Model (USM) for version 3 of the Simple Network Management Protocol (SNMPv3)
- RFC 3415, View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)
- RFC 3416, Version 2 of the Protocol Operations for the Simple Network Management Protocol (SNMP)
- RFC 3417, Transport Mappings for the Simple Network Management Protocol (SNMP)
- RFC 3418, Management Information Base (MIB) for the Simple Network Management Protocol (SNMP)

RFC 3419, Textual Conventions for Transport Addresses

1.2 Configuration Task Summary

SNMP configuration includes the following tasks:

(1) Configuring Basic SNMP Features

(2) (Optional) Configuring the Trap Function

(3) (Optional) Configuring the Agent Shielding Function

(4) (Optional) Configuring SNMP Control Parameters

1.3 Configuring Basic SNMP Features

1.3.1 Overview

After the basic SNMP function is configured, users can access the agent through the NMS.

1.3.2 Restrictions and Guidelines

- When an authentication name is configured, the default access permission is read-only if no access permission is specified.
- When the view-based access control model (VACM) is used, an SNMP user group must be configured. By associating specified users with a group and associating the group with a view, you can ensure that the users in the group have the same access permission. In this way, you can determine whether managed objects associated with an operation are in the allowable range of a view. Only managed objects in the range of a view can be accessed.
- When you configure users in SNMPv3, you can specify the security level, authentication algorithm (MD5 or SHA), authentication password, encryption algorithm (DES at present), and encryption password.
- After the SNMP attack prevention and detection function is configured, the corresponding action can be taken after authentication fails consecutively.
- The permanently forbidden source IP addresses can be authenticated for access again only after the administrator manually unlocks the IP addresses. The source IP addresses that are forbidden in a period of time can be authenticated for access again after the period expires or after the administrator manually unlocks the IP addresses.
- Password dictionary check for communities and users is disabled by default. This function must be used with the **password policy info** command in global configuration mode.
- The SNMP logging function records the Get, Get-Next, and Set operations performed by the NMS on the SNMP Agent. When the Get and Get-Next operations are performed, the agent records the IP address of the NMS user, operation type, and OID of the operation node. When the Set operation is performed, the agent records the IP address of the NMS user, operation type, OID of the operation node, and set value. These logs are sent to the information center of devices. The level of these logs is informational, that is, the logs are used as prompt information of devices. A large number of logs will affect device performance. In normal conditions, you are advised to disable the SNMP logging function.
- SNMP ACL will be associated with the ACL-based filtering rule to prevent unauthorized users from sampling SNMP information.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure an SNMP view.

snmp-agent view *view-name oid-tree* { **include** | **exclude** }

An SNMP view needs to be configured when the View-based Access Control Model (VACM) function is used.

The default view allows access to all MIB objects.

(4) (Optional) Configure an SNMP user group.

snmp-agent group *groupname* { **snmpv1** | **snmpv2c** | **snmpv3** { **auth** | **noauth** | **priv** } } [**read** *readview*] [**write** *writeview*] [**access-apply** {

[**ipv6** *ipv6-acl-name*]

acl-name | *acl-number* }]

No user group is configured by default.

(5) Configure an authentication name and access permission.

snmp-agent community [**0** | **7** | **secret** [**0** | **8**]] *community* [**view** *view-name*] [**ro** | **rw**] [**host** *ipv4-address* | **host** *ipv6-address*]

[**ipv6** *ipv6-acl-name*]

[*acl-name* | *acl-number*]

No authentication name is configured and the access permission of all communities is read-only by default.

When SNMPv1 and SNMPv2C are used to manage network devices, an authentication name must be configured on the agent.

(6) Configure an SNMP user.

snmp-agent user *username groupname* { **snmpv1** | **snmpv2c** | **snmpv3** [**crypto** | **interactive**] [**authentication** { **md5** | **sha** | **sha2-256** | **sha2-512** } *auth-password*] [**priv** { **des56 | acs128** } *priv-password*] } [**access-apply** {

[**ipv6** *ipv6-acl-name*]

| *acl-name* | *acl-number* }]

No SNMP user is configured by default.

When SNMPv3 is used to manage network devices, a user must be configured.

(7) (Optional) Enable the agent function.

enable-server service snmp-agent

The SNMP agent function is enabled by default.

(8) (Optional) Enable the SNMP attack prevention and detection function.

snmp-agent authentication attempt *attempt-times* **exceed** { **lock** | **lock-time** *lock-time* | **unlock** }

SNMP attack defense is enabled by default. The maximum number of consecutive SNMP authentication failures is 3 by default. The account is locked for 10 minutes after three consecutive SNMP authentication failures.

(9) (Optional) Configure password dictionary check for communities.

snmp-agent subscribe secret-dictionary-check

Password dictionary check for communities and users is disabled by default.

(10) (Optional) Configure the SNMP logging function to record the Get, Get-Next, and Set Operations performed by the NMS on the SNMP Agent.

snmp-agent logging { **authentication** { **fail | success** } | **get-operation** | **set-operation** | **trap-info** }

By default, the function of logging the Get and Get-Next Operations performed by the NMS on the SNMP agent is disabled and the function of logging the Set Operation is enabled.

(11) (Optional) Configure a global ACL filtering rule for the SNMP server.

snmp-agent ip access-apply {*aclnum* | *aclname*}

snmp-agent ipv6 access-apply *aclv6name*

No ACL filtering rule is configured for the SNMP server by default.

1.4 Configuring the Trap Function

1.4.1 Overview

After the Trap function is configured, the agent actively sends Trap messages to the NMS.

1.4.2 Restrictions and Guidelines

- The host address of the NMS must be configured when the agent actively sends Trap messages to the NMS. That is, the **snmp-agent host** command is used with the **snmp-agent subscribe traps** command in the global configuration mode.
- Multiple SNMP hosts can be configured to receive Trap messages. A host can combine different types of Trap messages, ports, and VRF forwarding tables. If a host is configured with the same port and VRF in multiple configurations, the last configuration is combined with the previous configurations. To send different Trap messages to the same host, configure different types of Trap messages each time. These configurations are finally combined.
- The system reboot notification function must be enabled on the agent so that the RGOS system sends Trap messages to the NMS to notify system reboot before reloading or rebooting of the device.
- After the Link Trap message sending function is enabled on an interface, the SNMP sends a Link Trap message if the link status on the interface changes. Otherwise, the SNMP does not send the message.
- By default, the IP address of an interface that sends SNMP packets is used as the source address of the SNMP packets. If you want to use a fixed local IP address as the source address to facilitate management, configure the source address for sending Trap messages on the agent.
- The **snmp-agent trap**-**format private** command can be run to include private fields in Trap messages. The supported private field is the alarm generation time. For the specific data types and data ranges of the fields, see the Ruijie-TRAP-FORMAT-MIB.mib file.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) (Optional) Configure an NMS host address.

snmp-agent host [**oob**] { *ipv4-address* | **ipv**6 *ipv6-address* **| domain** *domain-name* } [**vrf** *vrf-name*] [**traps** | **informs**] [**version** { { **1** | **2c** } [**0** | **7**] *community* | **3** { **auth** | **noauth** | **priv** } *username* }] [**udp-port** *port-number*] [**via** *mgmt-name*] [*notification-type*]

No SNMP host address is configured by default.

(4) (Optional) Enable the agent to actively send Trap messages to the NMS.

snmp-agent subscribe traps [*notification-type*]

The SNMP agent is forbidden to send Trap messages to the NMS by default.

(5) (Optional) Configure the system reboot trap sending function.

snmp-agent reload-trap

The SNMP system reboot notification function is disabled by default.

(6) (Optional) Configure a source address for Trap messages.

snmp-agent trap-source-interface *interface-type interface-number*

The IP address of the interface that sends SNMP packets is used as the source address by default.

(7) (Optional) Include private fields in sent Trap messages.

snmp-agent trap-format private

SNMP Trap messages do not include private fields by default.

(8) Configure Inform message sending attempts and timeout time.

snmp-agent inform { **retries** *retry-number* | **timeout** *timeout* }

The number of default Inform message sending attempts is **3** and the default Inform message timeout time is **15** seconds.

(9) Enter the interface configuration mode.

- Enter the Layer 2 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 2 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the Layer 3 Ethernet interface configuration mode.

  interface *ethernet-type interface-number*

- Enter the Layer 3 link aggregation configuration mode.

  interface aggregateport *interface-number*

- Enter the switch virtual interface (SVI) configuration mode.

  interface vlan *interface-number*

(10) (Optional) Enable the Link Trap message sending function on an interface.

snmp-agent enable linktrap

The Link Trap message sending function is enabled on an interface by default.

(11) (Optional) Enable heartbeat Trap.

snmp heartbeat on

Heartbeat Trap is enabled by default.

(12) (Optional) Enable the heartbeat Trap interval.

snmp heartbeat period *time*

The default heartbeat Trap interval is 300 seconds.

1.5 Configuring the Agent Shielding Function

1.5.1 Overview

After the agent shielding function is enabled, the port access frequency and network attack probability are reduced.

1.5.2 Restrictions and Guidelines

The SNMP agent function is disabled by default. When SNMP agent parameters (for example, NMS host address, authentication name, and access permission) are configured, the SNMP agent service is automatically enabled. The **enable-server service snmp-agent** command must be run so that the SNMP agent service can take effect. If the SNMP agent service is disabled or the **enable-server service snmp-agent** command is not run, the SNMP agent service does not take effect. The **unset snmp-agent** command is used to shield all SNMP agent service configurations. In this case, running the **display current-config** command will not display the configurations. The configurations can be restored after the SNMP agent service is enabled again. The **unset enable-server service snmp-agent** command is used to disable the SNMP agent service. In this case, the SNMP agent function is disabled and packets and trap messages are not sent and received, but the SNMP agent configurations are not shielded.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the SNMP agent shielding function.

unset snmp-agent

The SNMP agent function is enabled by default.

(4) Disable the SNMP agent function of a device.

unset enable-server service snmp-agent

The SNMP agent function is enabled by default.

1.6 Configuring SNMP Control Parameters

1.6.1 Overview

After basic parameters such as device contact mode, device location, serial number, and Trap message sending are configured, the NMS can obtain information such as the contact person and physical location of the device by accessing the parameters.

1.6.2 Restrictions and Guidelines

- When too many SNMP request packets cause SUMP tasks to occupy a large CPU space, you can configure SNMP traffic control to limit the number of request packets processed per second in each SNMP task to control the CPU usage.
- SNMP contains three versions: v1, v2C, and v3. If only a specified version is used for MIB interaction management in practice, you can disable unused SNMP versions by running the **unset snmp-agent subscribe version** [**snmpv1** | **snmpv2c** | **snmpv3**] command.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the contact mode of the system.

snmp-agent contact *contact-text*

The contact mode of the system is empty by default.

(4) Configure the system location.

snmp-agent location *location-text*

The system location is empty by default.

(5) Configure the system serial number.

snmp-agent chassis-id *chassis-id-text*

The default system serial number is 60FF60.

In general, the device serial number is used as the SNMP serial number to facilitate identification of the device.

(6) Configure NE information of a device.

snmp-agent net-id *net-id-text*

The NE code information of a device is empty by default.

(7) Configure the maximum packet length of the SNMP agent.

snmp-agent packetsize *packetsize*

The maximum packet length of the SNMP agent is 1,472 bytes by default.

(8) Configure the UDP port ID of the SNMP service.

snmp-agent udp-port *port-number*

The default UDP port ID of the SNMP service is **161**.

(9) Configure the source port of the device on which the SNMP service is deployed.

snmp-agent source-interface *interface-type interface-number*

The source port of a device with a valid IP address is used to receive SNMP packets by default.

(10) Configure the queue length of Trap messages.

snmp-agent trap queue-length *queue-length*

The default queue length of the Trap messages is **100**.

(11) Configure the sending interval of Trap messages.

snmp-agent trap-send-interva *trap-timeout-time*

The Trap messages are resent with a timeout time of **300** milliseconds by default.

(12) Configure SNMP traffic control.

snmp-agent flow-control pps *packet-count*

About 300 SNMP request packets are processed every second by default.

(13) Configure the status of SNMP versions.

unset snmp-agent subscribe version { **snmpv1** | **snmpv2c** | **snmpv3** }

SNMPv1 and SNMPv2c are disabled and SNMPv3 is enabled by default.

1.7 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

⚠ Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-3 Monitoring

Command	Purpose
cleanup snmp locked-ip [**ipv4** *ipv4-address* | **ipv6** *ipv6-address*] **cleanup snmp locked-ip ipv4** *ipv4-address*	Clears the list of source IP addresses that are locked after authentication fails consecutively.
display snmp [**group** | **host** | **locked-ip** | **process-mib-time** | **mib** | **user** | **version** | **view**]	Displays the SNMP status information.

1.8 Configuration Examples

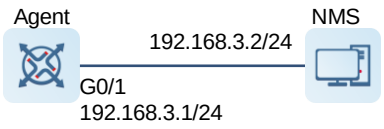
1.8.1 Configuring Basic Functions of SNMPv2c

1. Requirements

Network devices are managed and monitored through the SNMP network manager, as shown in Figure 1-5.

2. Topology

Figure 1-5 Configuring Basic Functions of SNMPv2c



3. Procedure

Enable the SNMP agent function.

```
Device> subscribe
Device# config-system terminal
Device(cfg)# enable-server service snmp-agent
```

Configure a read/write community with the name comm_v1.

```
Device(cfg)# snmp-agent community comm_v2 rw
```

Configure an IP address for the agent. Set the address of GigabitEthernet 0/1 to 192.168.3.1/24.

```
Device(cfg)# interface gigabitethernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
Device(cfg-if-gigabitEthernet 0/1)# ip address 192.168.3.1 255.255.255.0
Device(cfg-if-gigabitEthernet 0/1)# exit
```

4. Verification

Run the **ping** command to verify that the agent and NMS are mutually reachable via L3 routes.

```
Device# ping 192.168.3.2
Sending 5, 100-byte ICMP Echoes to 192.168.3.2, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the **display service** command to check whether the SNMP agent function is enabled.

```
Device# display service
snmp-agent       : enabled
ssh-server       : disabled
telnet-server : enabled
```

5. Configuration Files

Agent configuration file

```
interface GigabitEthernet 0/1
  ip address 192.168.2.1 255.255.255.0
!
snmp-agent community comm_v2 rw
!
end
```

1.8.2 Configuring Basic Functions of SNMPv3

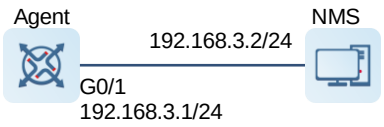
1. Requirements

The network management station is connected to the managed network devices. On the network management station, users access the MIB on the network devices through the SNMP network manager and receive messages actively sent by the network devices to manage and monitor the network devices.

- The management workstation (NMS) manages network devices (agents) based on the authentication and encryption mode of users.
- Network devices can control the operation permission of users to access MIB objects.
- Network devices can actively send authentication and encryption messages to the NMS.

2. Topology

Figure 1-6 Configuring Basic Functions of SNMPv3



3. Notes

- Configure an MIB view and an SNMP user group.
- Configure an SNMP user.
- Configure an SNMP host.
- Configure an IP address for the agent.

4. Procedure

Create an MIB view with the name view1, which includes the associated MIB object 1.3.6.1.2.1.1. Create an MIB view with the name view2, which includes the associated MIB object 1.3.6.1.2.1.1.4.0.

```
Device> subscribe
Device# config-system terminal
Device(cfg)# snmp-agent view view1 1.3.6.1.2.1.1 include
```

Create an MIB view with the name view2, which includes the associated MIB object 1.3.6.1.2.1.1.4.0.

```
Device(cfg)# snmp-agent view view2 1.3.6.1.2.1.1.4.0 include
```

Create an SNMP user group with the name g1, select v3 as the version, and configure an authentication and encryption mode with the security level priv. Configure view1 as a readable view and view2 as a writable view.

```
Device(cfg)# snmp-agent group g1 v3 priv read view1 write view2
```

Configure an SNMP user. Create a user with the name **user1** under the group g1, select v3 as the version, and set the authentication mode to md5, authentication password to 123, encryption mode to DES56, and encryption password to 321.

```
Device(cfg)# snmp-agent user user1 g1 v3 authentication md5 123 priv des56 321
```

Configure the SNMP host: Set the host address to 192.168.3.2, and the version to v3, configure an authentication and encryption mode with the security level priv, and associate the username **user1**.

```
Device(cfg)# snmp-agent host 192.168.3.2 traps version 3 priv user1
```

Enable the agent to actively send Trap messages to the NMS.

```
Device(cfg)# snmp-agent subscribe traps
```

Configure an IP address for the agent. Set the address of GigabitEthernet 0/1 to 192.168.3.1/24.

```
Device(cfg)# interface gigabitethernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
Device(cfg-if-gigabitEthernet 0/1)# ip address 192.168.3.1 255.255.255.0
Device(cfg-if-gigabitEthernet 0/1)# exit
```

5. Verification

Run the **ping** command to verify that the agent and NMS are mutually reachable via L3 routes.

```
Device# ping 192.168.3.2
Sending 5, 100-byte ICMP Echoes to 192.168.3.2, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the **display snmp user** command to display the SNMP user.

```
Device# display snmp user
User name: user1
Engine ID: 800013110300d0f8221120
storage-type: permanent         active
Security level: auth priv
Auth protocol: MD5
Priv protocol: DES
Group-name: g1
```

Run the **display snmp view** command to display the SNMP view.

```
Device# display snmp view
view1(include) 1.3.6.1.2.1.1
view2(include) 1.3.6.1.2.1.1.4.0
default(include) 1.3.6.1
```

Run the **display snmp group** command to display the SNMP group.

```
Device# display snmp group
groupname: g1
securityModel: v3
securityLevel:authPriv
readview: view1
writeview: view2
notifyview:
```

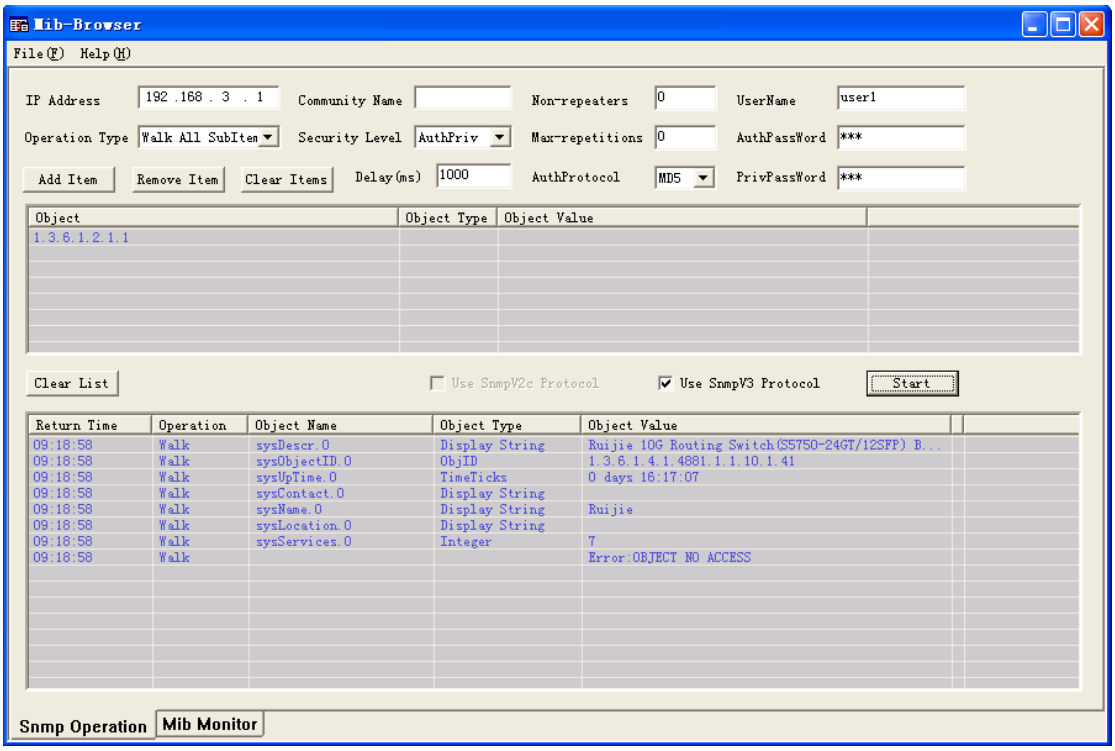
Run the **display snmp host** command to display the host information configured by the user.

```
Device# display snmp host
Notification host: 192.168.3.2
udp-port: 162
type: trap
```

```
user: user1
security model: v3 authPriv
```

Install the MIB browser query. Enter the IP address 192.168.3.1 in IP Address and **user1** in UserName, select AuthPriv for Security Level, enter 123 in AuthPassWord, select MD5 for AuthProtocol, and enter 321 in PrivPassWord. Click **Add Item** and select an MIB management unit to be queried, for example, System. Click **Start** to query the MIB of the network devices. The query results are displayed in the lowest pane of the dialog box, as shown in Figure 1-7.

Figure 1-7 MIB-browser Query Example



6. Configuration Files

Agent configuration file

```
interface GigabitEthernet 0/1
 ip address 192.168.2.1 255.255.255.0
!
snmp-agent view view1 1.3.6.1.2.1.1 include
snmp-agent view view2 1.3.6.1.2.1.1.4.0 include
snmp-agent user user1 g1 v3 crypto authentication md5 DE8E9D1158A057A69EF73D1C12C51CC5 priv
des56 C767B705F9B261E6D359D4BFBDAF9CF4
snmp-agent group g1 v3 priv read view1 write view2
snmp-agent host 192.168.3.2 traps version 3 priv user1
snmp-agent subscribe traps
!
end
```

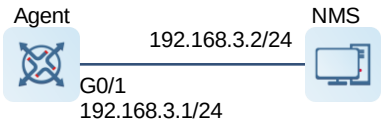
1.8.3 Configuring Basic Functions of Trap

1. Requirements

The network management station (NMS) manages network devices (agents) based on the community authentication mode, and the network devices can actively send messages to the network management station.

2. Topology

Figure 1-8 Configuring Basic Functions of Trap



3. Procedure

Set the agent and NMS to be reachable via L3 routes, and set the IP address of GigabitEthernet 0/1 on the agent to 192.168.3.1.

```
Device> subscribe
Device# config-system terminal
Device(cfg)#interface gigabitethernet 0/1
Device(cfg-if-gigabitEthernet 0/1)# unset switch-port
Device(cfg-if-GigabitEthernet 0/1)#ip address 192.168.3.1 255.255.255.0
Device(cfg-if-GigabitEthernet 0/1)#exit
```

Set the SNMP host address to 192.168.3.2, the message format to Version 2c, and the authentication name to user1. Enable the agent to actively send Trap messages to the NMS.

```
Device> subscribe
Device# config-system terminal
Device(cfg)# snmp-agent host 192.168.3.2 traps version 2c user1
Device(cfg)# snmp-agent subscribe traps
```

4. Verification

Run the **ping** command to verify that the agent and NMS are mutually reachable via L3 routes.

```
Device# ping 192.168.3.2
Sending 5, 100-byte ICMP Echoes to 192.168.3.2, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/1 ms.
```

Run the **display snmp** command to display the SNMP status.

```
Device# display snmp
Chassis: 1234567890
0 SNMP packets input
        0 Bad SNMP version errors
        0 Unknown community name
        0 Illegal operation for community name supplied
```

```
        0 Encoding errors
        0 Number of requested variables
        0 Number of altered variables
        0 Get-request PDUs
        0 Get-next PDUs
        0 Set-request PDUs
0 SNMP packets output
        0 Too big errors (Maximum packet size 1472)
        0 No such name errors
        0 Bad values errors
        0 General errors
        0 Response PDUs
        0 Trap PDUs
SNMP global trap: enabled
SNMP logging: disabled
SNMP agent: enabled
```

5. Configuration Files

Agent configuration file

```
interface GigabitEthernet 0/1
  ip address 192.168.2.1 255.255.255.0
!
snmp-agent host 192.168.3.2 traps version 2c user1
snmp-agent subscribe traps
!
end
```

Contents

1 Configuring RMON 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 3
1.2 Configuration Task Summary 3
1.3 Configuring the Ethernet Statistics Function 3
1.3.1 Overview 3
1.3.2 Restrictions and Guidelines 3
1.3.3 Procedure 4
1.4 Configuring the History Statistics Function 4
1.4.1 Overview 4
1.4.2 Restrictions and Guidelines 4
1.4.3 Procedure 4
1.5 Configuring the Alarm Function 5
1.5.1 Overview 5
1.5.2 Restrictions and Guidelines 5
1.5.3 Procedure 5
1.6 Monitoring 5
1.7 Configuration Examples 6
1.7.1 Configuring the Ethernet Statistics Function 6
1.7.2 Configuring the History Statistics Function 7

1.7.3 Configuring the Alarm Function 10

1 Configuring RMON

1.1 Introduction

1.1.1 Overview

The Remote Network Monitoring (RMON) protocol is a network management protocol that is built on the Simple Network Management Protocol (SNMP) architecture. RMON provides statistics and alarm functions for a managing device to remotely monitor and manage devices. he statistics function means that a managed device can trace and calculate different types of traffic information on a network segment that is connected to the port of the device regularly or continuously. For example, the device can trace and calculate the total number of packets or the total number of excessively long packets received on a network segment in a period. The alarm function means that a managed device can monitor specified management information base (MIB) variables and automatically record an event and send Trap messages to the managing device when one of the variables reaches the alarm threshold, for example, the port rate or the ratio of broadcast packets reaches a specified value.

Both RMON and SNMP are used for remote network management. Compared with SNMP, RMON has the following advantages:

- RMON notifies the managing device of alarm variable abnormality in Trap messages by using the SNMP Trap message sending mechanism. Though the Trap function is also defined in SNMP, this function is used to notify the managing device of the function running status of a managed device and physical state change of an interface. SNMP and RMON have different monitored objects, trigger conditions, and report content.
- Compared with SNMP, RMON can monitor remote network devices more effectively and actively, thus providing an efficient means of monitoring subnet running. According to RMON, a managed device can automatically send Trap messages when a value reaches the alarm threshold. This frees the managing device from frequently obtaining and comparing values of MIB variables and thus reduces the communication between the managing device and managed device. As a result, large interconnected networks are being managed simply yet effectively.

1.1.2 Principles

RMON defines different RMOM groups, and the company’s devices support the statistics group, history group, alarm group, and event group. The four groups are as follows:

1. Statistics Group

A statistics group is used to monitor and calculate traffic data on an Ethernet interface. The traffic data consists of cumulative values and is calculated from the time of entry creation to the current time. Statistical objects include discarded packets, broadcast packets, CRC errors, over-small or over-large packets, and conflicts. The statistical results are saved in the Ethernet statistical table for the administrator to view at any time.

2. History Group

A history group is used to collect network traffic periodically and records cumulative values of the traffic and bandwidth usage in each period. The data is saved in the history control table for the administrator to process. A history group includes two sub-groups:

- The HistoryControl sub-group is used to configure control information such as sampling interval and sampling data sources.
- The EthernetHistory sub-group provides the administrator with history data of a specific network segment, including traffic, error packets, broadcast packets, usage, and hit times.

3. Alarm Group

The alarm group is used to monitor specified MIB objects. When the value of an MIB object exceeds the set upper limit or lower limit, an alarm is triggered and processed as an event.

4. Event Group

An event group is used to define processing methods of events. When a monitored MIB object reaches the alarm condition, an event is triggered. The event can be processed with one of the following methods:

- None: The event is ignored.
- Log: The event related information is recorded in a log table for the administrator to view at any time.
- Snmp-trap: The Trap message is sent to the network management station to notify the occurrence of the event.
- Log-and-trap: The event-related information is recorded in a log table and the Trap message is sent to the network management station.

5. Working Process of RMON

RMON allows multiple monitor authorities to collect data with the following two methods: use the special RMON probe to collect data so that the NMS can obtain all MIB information from the RMON probe; embed the RMON agent into a network device (for example, switch or router) so that the device possesses the function of an RMON probe. The NMS uses basic commands of SNMP to exchange data with the network device and collect network management information. Limited by the device resources, this method does not obtain all data of the RMON MIB. Generally, only four groups of information are collected.

Through the RMON agent running on the managed network device, the NMS can obtain information such as traffic, error statistics, and performance statistics on the network segment that is connected to the port of this network device, to remotely manage this network device. Figure 1-1 displays the communication network between the NMS and RMON agent.

- RMON Ethernet statistics function

  Make the statistics of network traffic on the Ethernet interface from the time of entry creation to the current time.

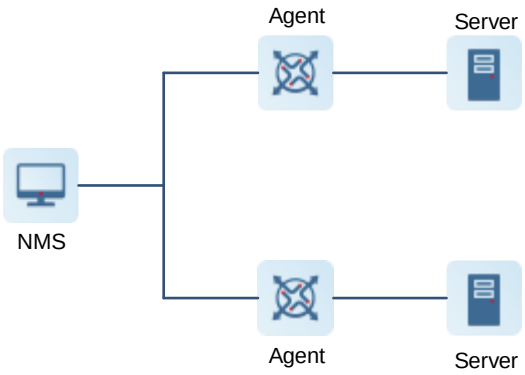
- RMON history statistics function

  Record cumulative statistical values of traffic on the Ethernet interface in each period.

- RMON alarm function

  Monitor changes of alarm variable values periodically. If an alarm variable value exceeds the specified upper or lower limit, corresponding event is triggered, for example, a Trap message is sent or a LogTable entry record is created. Continuous excess of an upper or a lower limit triggers an event only once and the event can be triggered again by excess of the opposite limit.

Figure 1-1 Communication Network Between the NMS and RMON Agent



1.1.3 Protocols and Standards

- STD 0059 / RFC 2819: Remote Network Monitoring Management Information Base
- RFC 4502: Remote Network Monitoring Management Information Base Version 2
- RFC 3919: Remote Network Monitoring (RMON) Protocol Identifiers for IPv6 and Multi Protocol Label Switching (MPLS)
- RFC 3737: IANA Guidelines for the Registry of Remote Monitoring (RMON) MIB Modules
- RFC 3434: Remote Monitoring MIB Extensions for High Capacity Alarms
- RFC 3395: Remote Network Monitoring MIB Protocol Identifier Reference Extensions
- RFC 3287: Remote Monitoring MIB Extensions for Differentiated Services
- RFC 3273: Remote Network Monitoring Management Information Base for High Capacity Networks
- RFC 2896: Remote Network Monitoring MIB Protocol Identifier Macros
- RFC 2895: Remote Network Monitoring MIB Protocol Identifier Reference

1.2 Configuration Task Summary

RMON configuration includes the following tasks:

- Configuring the Ethernet Statistics Function
- Configuring the History Statistics Function
- Configuring the Alarm Function

1.3 Configuring the Ethernet Statistics Function

1.3.1 Overview

With the Ethernet statistics function of RMON, you can learn the cumulative statistical values of traffic generated on the monitored Ethernet interface from the time of entry creation to the current time.

1.3.2 Restrictions and Guidelines

- To calculate and monitor a specified interface, you must configure the Ethernet statistics function on this interface.

- After the statistics entry is created on the specified interface, the statistics group calculates the traffic of this interface. That is, make the statistics of variables defined by the Ethernet statistics table and record the cumulative values of variables that are generated from the creation time of the RMON statistics table to the current time.
- This function cannot be configured in the batch interface configuration mode.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 or Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

(4) Configure statistical items of RMON.

rmon collection statistics *collection-stats-table-index* [**owner** *owner-name*]

The Ethernet statistics function is not configured on an Ethernet interface by default.

1.4 Configuring the History Statistics Function

1.4.1 Overview

With the history statistics function of RMON, you can learn the cumulative statistical values of traffic generated in each period and the bandwidth utilization on the monitored interface.

1.4.2 Restrictions and Guidelines

- The history group makes the statistics of variables defined by the history statistics function and records cumulative values of variables in each period.
- To collect network statistics on a specified interface, you must configure the history control function on the interface. This function cannot be configured in the batch interface configuration mode.
- It is not allowed to modify parameters of the configured history statistics function.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 or Layer 3 Ethernet interface configuration mode.

interface *ethernet-type interface-number*

(4) Configure the history statistics function of an Ethernet interface.

rmon collection history-stats *collection-history-table-index* [**buckets** *bucket-number*] [**interval** *period-time*] [**owner** *owner-name*]

The history statistics function is not configured on an Ethernet interface by default.

1.5 Configuring the Alarm Function

1.5.1 Overview

The alarm function of RMON monitors whether the values of alarm variables vary within the valid range periodically.

1.5.2 Restrictions and Guidelines

- If alarm variables are MIB variables defined by the RMON statistical group or history group, the RMON statistics function or history statistics function must be configured on the monitored Ethernet interface. Otherwise, the alarm table fails to be created.
- The alarm function is implemented by the alarm table and event table. After an alarm event is triggered, the SNMP agent must be configured so that Trap messages can be sent to the managing device. For more information about SNMP agent configuration, see *Configuring SNMP*.
- Configured event table parameters, including event type, Trap community name, event description, and event creator, can be modified.
- Configured alarm table parameters, including alarm variable, sampling type, table creator, sampling interval, upper/lower limit, and event, can be modified.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an event table.

rmon event-instance *event-table-index* [**description** *description-string* | **log** | **owner** *owner-name* | **trap** *community*]*

No RMON event is configured by default.

(4) Configure the alarm function.

rmon alarm-instance *alarm-table-index alarm-variable sampling-interval* { **absolute** | **delta** } **upper-threshold** *sampling-rising-threshold-value* [*event-number*] **lower-threshold** *falling-threshold-value* [*event-number*] [**owner** *owner-name*]

The RMON alarm function is not configured by default.

1.6 Monitoring

Run the **displaydisplay** command to check the running status of a configured function to verify the configuration effect.

Table 1-1 Monitoring

Command	Purpose
display rmon-instance	Displays all RMON configurations.
display rmon-instance statistics	Displays the Ethernet statistics table information.
display rmon-instance history	Displays the history control table information.
display rmon-instance alarm	Displays the alarm table information.
display rmon-instance event	Displays the event table information

1.7 Configuration Examples

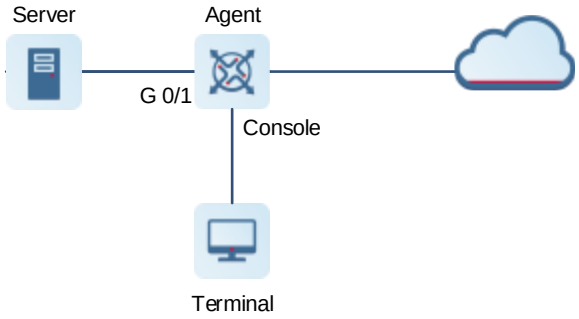
1.7.1 Configuring the Ethernet Statistics Function

1. Requirements

The RMON agent is connected to a server. The network administrator uses the RMON statistics group to calculate the performance of received packets on GigabitEthernet 0/1. Thus, the network administrator can understand the data of packets received on an interface at any time by viewing data, and take timely actions to process network abnormalities.

2. Topology

Figure 1-2 Configuring the Ethernet Statistics Function



3. Procedure

Configure a statistics table instance on GigabitEthernet 0/1 of the device on which the RMON agent is deployed to make the statistics of traffic on this interface.

```
Device> privilege
Device# config-system terminal
Device(cfg)# interface gigabitEthernet 0/1
```

```
Device(cfg-if-GigabitEthernet 0/1)# rmon collection stats 1 owner admin
```

4. Verification

Run the **display rmon-instance statistics** command to display Ethernet statistics information.

```
Device# display rmon-instance statistics
ether statistic table:
                    index = 1
                    interface = GigabitEthernet 0/1
                    owner = admin
                    status = 1
                    dropEvents = 0
                    octets = 25696
                    pkts = 293
                    broadcastPkts = 3
                    multiPkts = 0
                    crcAlignErrors = 0
                    underSizePkts = 0
                    overSizePkts = 0
                    fragments = 0
                    jabbers = 0
                    collisions = 0
                    packets64Octets = 3815
                    packets65To127Octets = 1695
                    packets128To255Octets = 365
                    packets256To511Octets = 2542
                    packets512To1023Octets = 152
                    packets1024To1518Octets = 685
```

5. Configuration Files

Agent configuration file

```
hostname Device
!
interface GigabitEthernet 0/1
 rmon collection stats 1 owner admin
!
end
```

6. Common Errors

The parameters of configured statistics entries are reconfigured or modified.

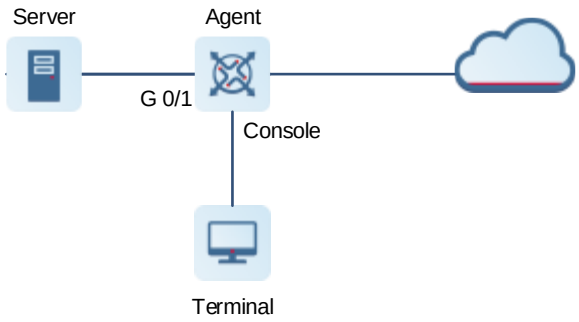
1.7.2 Configuring the History Statistics Function

1. Requirements

The RMON agent is connected to a server. The network administrator uses the history group to make the statistics of received packets on GigabitEthernet 0/1 every minute so as to monitor the network and master emergencies.

2. Topology

Figure 1-3 Configuring the History Statistics Function



3. Procedure

Configure a history control table on GigabitEthernet 0/1 of the device on which the RMON agent is deployed to make the statistics of traffic on this interface periodically.

```
Device> privilege
Device# config-system terminal
Device(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# rmon collection history 1 buckets 5 interval 300 owner admin
```

4. Verification

Run the **display rmon-instance history** command to display history group statistics.

```
Device# display rmon-instance history
rmon history control table:
                        index = 1
                        interface = GigabitEthernet 0/1
                        bucketsRequested = 5
                        bucketsGranted = 5
                        interval = 60
                        owner = admin
                        stats = 1

rmon history table:
                        index = 1
                        sampleIndex = 786
                        intervalStart = 6d:18h:37m:38s
                        dropEvents = 0
                        octets = 2040
                        pkts = 13
                        broadcastPkts = 0
                        multiPkts = 0
                        crcAlignErrors = 0
                        underSizePkts = 0
```

```
overSizePkts = 0
fragments = 0
jabbers = 0
collisions = 0
utilization = 0

index = 1
sampleIndex = 787
intervalStart = 6d:18h:38m:38s
dropEvents = 0
octets = 1791
pkts = 16
broadcastPkts = 1
multiPkts = 0
crcAlignErrors = 0
underSizePkts = 0
overSizePkts = 0
fragments = 0
jabbers = 0
collisions = 0
utilization = 0

index = 1
sampleIndex = 788
intervalStart = 6d:18h:39m:38s
dropEvents = 0
octets = 432
pkts = 6
broadcastPkts = 0
multiPkts = 0
crcAlignErrors = 0
underSizePkts = 0
overSizePkts = 0
fragments = 0
jabbers = 0
collisions = 0
utilization = 0

index = 1
sampleIndex = 789
intervalStart = 6d:18h:40m:38s
dropEvents = 0
octets = 432
pkts = 6
broadcastPkts = 0
multiPkts = 0
```

```
                    crcAlignErrors = 0
                    underSizePkts = 0
                    overSizePkts = 0
                    fragments = 0
                    jabbers = 0
                    collisions = 0
                    utilization = 0

                    index = 1
                    sampleIndex = 790
                    intervalStart = 6d:18h:41m:38s
                    dropEvents = 0
                    octets = 86734
                    pkts = 934
                    broadcastPkts = 32
                    multiPkts = 23
                    crcAlignErrors = 0
                    underSizePkts = 0
                    overSizePkts = 0
                    fragments = 0
                    jabbers = 0
                    collisions = 0
                    utilization = 0
```

5. Configuration Files

Agent configuration file

```
hostname Device
!
interface GigabitEthernet 0/1
 rmon collection history-stats 1 buckets 5 interval 300 owner admin
!
end
```

6. Common Errors

The parameters of configured history control entries are reconfigured or modified.

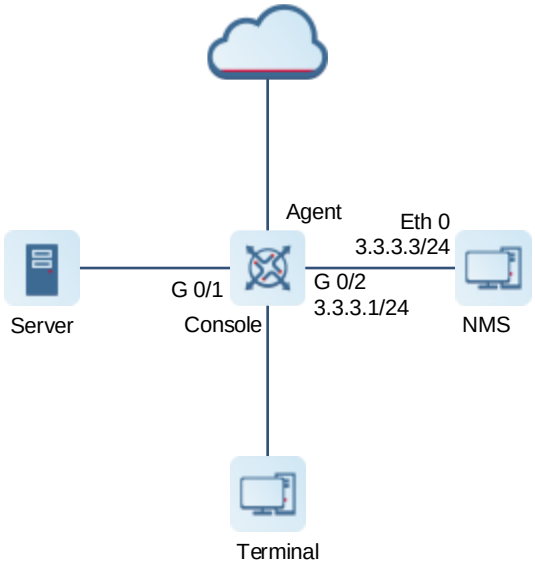
1.7.3 Configuring the Alarm Function

1. Requirements

The RMON agent is configured on a terminal device and connected to the NMS device for communication. The port GigabitEthernet 0/1 is connected to the server and must be monitored the number of received packets with unknown protocols. The sampling interval is 60 seconds. When the absolute collection value is smaller than 10, only a log record is generated. When the absolute sampling value is greater than 100, a log record is generated and a notification is sent to the NMS in a Trap message.

2. Topology

Figure 1-4 Configuring the Alarm Function



3. Notes

- Run SNMPv1 on the NMS, configure a community name public1, set the attribute to RW, and set the Trap message receiving address to 3.3.3.3.
- Monitor the number of packets with unknown protocols, received on GigabitEthernet0/1, and check whether the OID value is 1.3.6.1.2.1.2.2.1.15.3, sampling method is relative sampling, and sampling interval is 60 seconds. Check whether event 1 or event 2 is triggered when the relative sampling value is greater than 100 or smaller than 10. For event 1, a Trap message is sent and a log record is generated. For event 2, a log record is generated.

4. Procedure

Set the IP address of GigabitEthernet0/2 on the agent to 3.3.3.1.

```
Device> enabel
Device# config-system terminal
Device(cfg)# interface gigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# ip address 3.3.3.1 255.255.255.0
Device(cfg-if-GigabitEthernet 0/1)# exit
```

Configure an event group to process alarms.

```
Device(cfg)# snmp-server community public1 rw
Device(cfg)# snmp-server host 3.3.3.3 trap public1
Device(cfg)# rmon event 1 description rising-threshold-event log trap public owner admin
Device(cfg)# rmon event 2 description falling-threshold-event log owner admin
```

Configure the alarm function.

```
Device(cfg)# rmon alarm 1 1.3.6.1.2.1.2.2.1.15.3 60 delta rising-threshold 100 1 falling-threshold 10 2 owner admin
```

5. Verification

Run the **ping** command to verify that the agent and NMS are mutually reachable via L3 routes.

```
Device# ping 3.3.3.3
Sending 5, 100-byte ICMP Echoes to 3.3.3.3, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/3 ms.
```

Run the **display rmon-instance event** command to display the event table information.

```
Device# display rmon-instance event
rmon event table:
                        index = 1
                        description = rising-threshold-event
                        type = 4
                        community = public
                        lastTimeSent = 0d:0h:0m:0s
                        owner = admin
                        status = 1

                        index = 2
                        description = falling-threshold-event
                        type = 2
                        community =
                        lastTimeSent = 6d:19h:21m:48s
                        owner = admin
                        status = 1

rmon log table:
                        eventIndex = 2
                        index = 1
                        logTime = 6d:19h:21m:48s
                        logDescription = falling-threshold-event
```

Run the **display rmon-instance alarm** command to display the alarm table information.

```
Device# display rmon-instance alarm
rmon alarm table:
            index: 1,
            interval: 60,
            oid = 1.3.6.1.2.1.2.2.1.15.3
            sampleType: 2,
            alarmValue: 0,
            startupAlarm: 3,
            risingThreshold: 100,
            fallingThreshold: 10,
            risingEventIndex: 1,
            fallingEventIndex: 2,
```

```
        owner: admin,
        stauts: 1
```

6. Configuration Files

Agent configuration file

```
hostname Device
!
interface GigabitEthernet 0/1
 ip address 3.3.3.1 255.255.255.0
!
rmon alarm 1 1.3.6.1.2.1.2.2.1.15.3 60 delta rising-threshold 100 1 falling-threshold 10 2 owner admin
rmon event 1 description rising-threshold-event log trap public1 owner admin
rmon event 2 description falling-threshold-event log owner admin
!
snmp-server host 3.3.3.3 traps public1
snmp-server community public1 rw
!
end
```

7. Common Errors

- The entered object OID is improper. For example, the variable corresponding to this OID does not exist or the OID type is not an integer or unsigned integer.
- The upper limit is smaller than or equal to the lower limit.

Contents

1 Configuring NETCONF 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Structure of NETCONF 1
1.1.3 NETCONF Session Interaction Procedure 3
1.1.4 Authentication Mechanism 9
1.1.5 Protocols and Standards 11
1.2 Configuration Task Summary 11
1.3 Configuring Communication Between the NETCONF Server and Client 11
1.3.1 Overview 11
1.3.2 Restrictions and Guidelines 11
1.3.3 Prerequisites 12
1.3.4 Procedure 12
1.4 Configuring NETCONF Authentication 13
1.4.1 Overview 13
1.4.2 Restrictions and Guidelines 13
1.4.3 Procedure 14
1.5 Monitoring 14
1.6 Configuration Examples 15
1.6.1 Configuring NETCONF 15
1.6.2 Configuring NETCONF Authentication 20

1 Configuring NETCONF

Note

NETCONF is not supported by only the S6520-64CQ, and S6920-4C.

1.1 Introduction

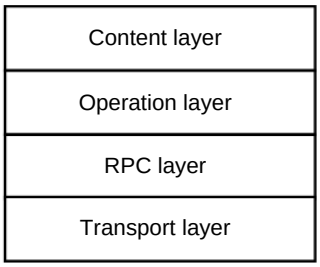
1.1.1 Overview

As network scale, complexity, and heterogeneity increase, it becomes more and more difficult for the traditional IP network management protocol Simple Network Management Protocol (SNMP) to manage the complicated network. In particular, SNMP cannot meet the requirements for configuration management and application system development of a large network. To overcome the shortcomings of SNMP, the Internet Engineering Task Force (IETF) formulates an extensible markup language (XML) based network management protocol – Network Configuration Protocol (NETCONF). NETCONF provides a programmable method to configure and manage network devices. This protocol configures parameters and obtains parameter values and statistics in XML packets and supports scalability, making network device configuration and management easy and efficient.

1.1.2 Structure of NETCONF

NETCONF supports the client/server (C/S) communication mode, in which a device runs the server of the protocol and a client runs the client program of the protocol. NETCONF packets, including all configuration data and protocol messages, follow the XML format. With a hierarchical similarity to the ISO/OSI system, NETCONF consists of four layers from down to up: transport layer, remote procedure call (RPC) layer, operation layer, and content layer, as shown in Figure 1-1.

Figure 1-1 Structure of NETCONF



1. Transport Layer

The transport layer provides secure transmission channels for NETCONF and supports security protocols such as Secure Shell (SSH), Simple Object Access Protocol (SOAP), and Blocks Extensible Exchange Protocol (BEEP). Now, SSH is preferred.

2. RPC Layer

The RPC layer provides a simple transmission-irrelevant mechanism. The client uses the <rpc> element to encapsulate the NETCONF request message, and the server uses the <rpc-reply> element to encapsulate the NETCONF reply message. RPC defines three message types:

- Hello: The hello messages are used to complete capability set exchange when a session is created between the NETCONF client and server.
- RPC and RPC-Reply: RPC are request messages sent by the NETCONF client to the NETCONF server, and RPC-Reply are reply messages returned by the NETCONF server to the NETCONF client. The reply messages are returned only when the NETCONF server receives RPC request messages; the reply messages and the received RPC request messages must contain the same message ID.
- Notification: The notifications are sent by means of subscription.

3. Operation Layer

Based on the basic primitive operation sets applied in RPC, 9 types of basic operation methods are defined:

- Data acquisition: Get and Get-Config
- Configuration: Edit-Config, Copy-Config, and Delete-Config
- Protection for critical resource (for example, configuration file) concurrency: Lock and Unlock
- Session ending: Close-Session and Kill-Session

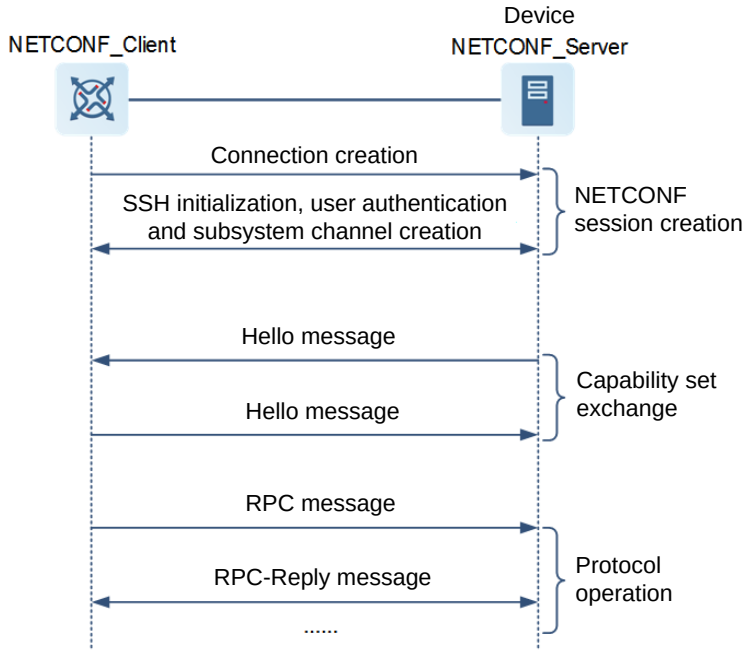
4. Content Layer

The content layer is a set of managed data objects. The content layer is the only layer that is not standardized and does not have a unified data model in NETCONF.

Note

The products use YANG for data modeling and XML for data transmission.

Figure 1-2 NETCONF Session Message Interaction



1.1.3 NETCONF Session Interaction Procedure

Figure 1-2 shows the NETCONF session interaction procedure, which falls into three parts.

1. Session Creation

(1) The transport layer of NETCONF relies on SSH. After the NETCONF server on a device starts the NETCONF process, a monitoring port is created.

(2) The NETCONF server monitors port 830 and creates an SSH channel. The SSH channel is created after a series of transport algorithm negotiations (including key agreement, compression algorithm, hash algorithm, encryption algorithm, and signature algorithm) and client authentication.

(3) After a session is created on the transport layer, the NETCONF client can exchange with the NETCONF server through this session.

Note

According to NETCONF, the default SSH TCP port ID is 830, which is customized based on the particular situation.

2. Capability Set Exchange

After the session is created, the NETCONF server and client send hello messages mutually to provide respective realized capability sets and ignore capabilities that are not understandable or realizable. The basic capability (urn:ietf:params:netconf:base:1.1) must be supported by the NETCONF server and client. To be compatible with earlier versions of NETCONF, the NETCONF server and client must support the basic capability (urn:ietf:params:netconf:base:1.0) of the earlier versions of NETCONF The NETCONF server and client finally obtain the intersected capabilities to perform subsequent data operation and management.

Caution

- In addition to the exchange capabilities defined in NETCONF RFC, developers can design other capabilities based on the specifications and formats described in NETCONF RFC.
- Capability exchange messages sent by the NETCONF client to the NETCONF server do not contain the session ID node (<session-id>).

3. Protocol Operation

- Get: Obtains device status or configuration data.

The packets sent by the client follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc message-id="xxx " xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<get>
<filter type="subtree">
        Configuration data (or status data) filtering rule
</filter>
</get>
</rpc>
```

The replies returned by the server follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="xxx " xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<data xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
Obtained configuration data (or status data)
</data>
</rpc-reply>
```

If all subsets of status data on the device fail to match the filtering rule, an empty data node is returned as follows:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="Message ID " xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<data xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"/>
</rpc-reply>
```

- Get-config: Obtains configuration data based on the filtering node.

The packets sent by the client follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc message-id="xxx" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<get-config>
<source>
<running/>
</source>
<filter type="subtree">
        Protocol-based filtering rule
</filter>
</get-config>
</rpc>
```

The replies returned by the server follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="xxx" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<data xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
Obtained configuration data
</data>
</rpc-reply>
```

If all subsets of configuration data on the device fail to match the filtering rule, an empty data node is returned as follows:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="xxx " xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<data xmlns="urn:ietf:params:xml:ns:netconf:base:1.0"/>
</rpc-reply>
```

Note

The Get-config operation can obtain configuration data subsets through different subtree filtering rules but cannot obtain device status.

- Edit-config: Configures a device based on the model definitions and operation attributes.

  In the edit-config packet, there are five operation attributes which are given in the description of the operation attribute of the XML packet. These five attributes are as follows:

  - Merge: Merges the configuration data that includes this attribute in the Edit-config packets to the specified configuration file or database. If this configuration data does not exist, create this configuration data.
  - Replace: Replaces the configuration data node in the specified configuration file or database with the configuration data that includes this attribute in the Edit-config packets. If this configuration data does not exist, create this configuration data based on the delivered content.
  - Create: Creates configuration data that includes this attribute in the Edit-config packets in the specified configuration file or database. If the configuration data does not exist, deliver content to create this configuration data. If the configuration data exists, a reply, corresponding to the RPC-error packet, is returned and contains the value of error-tag "data-exists".
  - Delete: Deletes configuration data that includes this attribute in the Edit-config packets from the specified configuration file or database. If the configuration data does not exist, a reply, corresponding to the RPC-error packet, is returned and contains the value of error-tag "data-missing". If the configuration data exists, corresponding configuration is deleted.
  - Remove: Removes configuration data that includes this attribute in the Edit-config packets from the specified configuration file or database. If the configuration data does not exist, this operation is ignored and OK is returned. If the configuration data exists, corresponding configuration is removed.

  The packets sent by the client follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc message-id="xxx " xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<edit-config>
```

```
<target> <running/> </target>
<config xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
        Configuration data
</config>
</edit-config>
</rpc>
```

The replies returned by the server follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="Message ID " xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<ok/>
</rpc-reply>
```

The Edit-config packet includes the Error-option node. The values of the Error-option node are of the enumerated type, including:

- **Continue-on-error**: During an Edit-config operation, if a configuration error occurs, the current error node is recorded and configuration continues. But an error message is returned finally (In the case of any configuration error, the final reply message is RPC-error).
- **Stop-on-error**: This is the default value of the Error-option. During an Edit-config operation, if a first configuration error occurs, the Edit-config operation stops immediately. The data configured before this error occurs takes effect.

The packets generally follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc message-id="xxx" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
<edit-config>
<target> <running/> </target>
Acton option when a configuration error occurs
<config xmlns:xc="urn:ietf:params:xml:ns:netconf:base:1.0">
        Configuration data
</config>
</edit-config>
</rpc>
```

Specification

Currently, Ruijie devices do not support the Replace operation. If this operation is delivered, it is processed as a Merge operation.

Note

- If the Edit-config packet does not include the error-option node, the default value of the error-option node is stop-on-error. Once a configuration error occurs on a node, the remaining configuration in the Edit-config packet stops immediately and an RPC-error is returned.
- If the Edit-config packet does not include the test-option node, the default value of the test-option node is **test-then-set**.
- If the Edit-config packet does not include the default-operation node, the default value of the default-operation node is **merge**.

- The digit type of configuration data in edit-config packets does not support the octal or hexadecimal format. For data in octal format, the 0 before the digit will be removed by default and processed as decimal data. If hexadecimal data carries 0x, an error is returned to the client.

- Copy-config: Copies the configuration file. For example, candidate configuration is copied to the configuration file, startup configuration is copied to the running configuration file, and running configuration is written into the startup configuration file. The copies can be successful if the target files support the write capability.

The packets sent by the client follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc message-id="xxx"    xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <copy-config>
      <target>
         <startup/>
      </target>
      <source>
         <running/>
      </source>
   </copy-config>
</rpc>
```

The replies returned by the server follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="xxx"    xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <ok/>
</rpc-reply>
```

- Delete-config: Deletes the configuration file. It is not allowed to delete the running file.

The packets sent by the client follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc message-id="xxx " xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <delete-config>
     <target>
        <startup/>
     </target>
  </delete-config>
</rpc>
```

The replies returned by the server follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="xxx " xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
   <ok/>
</rpc-reply>
```

Specification

Now, only the startup configuration can be synchronized to the running configuration file.

- Lock: Locks the configuration file. The configuration file can be accessed or modified through the current client, but cannot be accessed or modified by other clients or non-NETCONF clients (for example, SNMP or CLI). The packets sent by the client follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc message-id="xxx " xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
 <lock>
    <target>
       <running/>
    </target>
 </lock>
</rpc>
```

The replies returned by the server follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="xxx "    xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <ok/>
</rpc-reply>
```

Note

According to RFC 6241, the Lock operation is used to lock the configuration database (configuration files) and prevent multiple sources (for example, CLI, SNMP, and NETCONF sessions) from concurrently revising the configuration files. This avoids introducing other irrelevant configuration revisions. The device removes some functions of the operation, and can only prevent concurrent modifications (running configuration) from multiple NETCONF.

- Unlock: Unlocks the configuration database (configuration file, which indicates the running configuration file in a device). This operation is opposite to the Lock operation.

The packets sent by the client follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc message-id="xxx "    xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
 <unlock>
    <target>
       <running/>
    </target>
 </unlock>
</rpc>
```

The replies returned by the server follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="xxx "    xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <ok/>
</rpc-reply>
```

- Close-session: Closes the current session, including resource and lock release and disconnection.

The packets sent by the client follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
```

```
<rpc message-id="xxx" xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
    <close-session/>
</rpc>
```

The replies returned by the server follow the format below:

```
<?xml version="1.0" encoding="utf-8"?>
<rpc-reply message-id="xxx"    xmlns="urn:ietf:params:xml:ns:netconf:base:1.0">
  <ok/>
</rpc-reply>
```

Note

Before the Close-session operation closes the current session, ensure that proceeding business has been completed and no new business request is received.

- Kill-session: Forcibly closes the current session, including resource and lock release and disconnection.

Caution

- Before using the Kill-session operation to forcibly close the current session, ensure that proceeding business has been stopped and unfinished business is rolled back to original state.
- The Kill-session operation cannot close the current session.

1.1.4 Authentication Mechanism

The NETCONF authentication mechanism is used to manage the NETCONF operations of specific clients and the access to NETCONF resources, allowing the clients to perform operations and access data nodes stipulated by NETCONF.

1. Access Permissions Supported by the NETCONF Authentication Mechanism

The NETCONF authentication function is not configured by default. Therefore, authorized clients have all NETCONF permissions. Access permissions supported by the NETCONF authentication mechanism are as follows:

- Create: Permits or denies new data nodes to be added.
- Read: Permits or denies data nodes to be read.
- Update: Permits or denies data nodes to be updated.
- Delete: Permits or denies data nodes to be deleted.
- Execute: Permits or denies all protocol operations.

2. Content Supported by the NETCONF Authentication Mechanism

The NETCONF authentication function supports RPC method authentication and data node authentication.

- RPC method authentication

  RPC method authentication of NETCONF is used to control permissions of the operation methods stipulated by NETCONF. Table 1-1 lists the RPC methods and default permissions.

Table 1-1 RPC Methods and Default Permissions

RPC Method	Function	Default Permission (R: read; W: write; X: execute)
get	Obtains data.	X→R
get-config	Obtains configuration.	X→R
edit-config	Edits configuration.	X→W
copy-config	Replaces the target configuration with the source configuration.	X
delete-config	Deletes configuration.	X
validate	Performs syntax check.	X
commit	Validates the candidate configuration and runs the configuration.	X
cancel-commit	Cancels the confirmed-commit operation.	X
discard-changes	Discards candidate configuration that has not taken effect.	X
lock/unlock	Locks or unlocks configuration.	X
close-session	Closes the current session.	X
kill-session	Closes other sessions.	X
get-schema	Obtains the YANG file.	X

Note

- All RPC methods require the execution permission by default. The Get, Get-config, and Edit-config methods require the check of read/write permission for the data node.
- The Close-session method is permitted by default and cannot be denied.
- The Delete-config and Kill-session methods are denied by default, and can be configured as Permitted. Other RPC methods are permitted by default.
- The Lock and Unlock methods must be paired. If the Lock permission is configured, the Unlock permission is available as well.

- Data node authentication

  Data node authentication of NETCONF is used to control permissions of NETCONF data nodes. It can control the permissions of all modules, single module, and specific data nodes.

  - If only the slash (/) is configured, the authentication of NETCONF can control the permissions of the data nodes of all modules.

- If a module name is configured, for example, / snmp:snmp, the authentication of NETCONF can control the permissions of the data nodes of the single module.
- In other configuration conditions, permission control is applied to the data nodes of specific modules.

1.1.5 Protocols and Standards

- RFC 4741: NETCONF Configuration Protocol
- RFC 4742: Using the NETCONF Configuration Protocol over Secure Shell (SSH)
- RFC 4743: Using NETCONF over the Simple Object Access Protocol (SOAP)
- RFC 4744: Using the NETCONF Protocol over the Blocks Extensible Exchange Protocol (BEEP)
- RFC 5277: NETCONF Event Notifications
- RFC 5381: Experience of Implementing NETCONF over SOAP
- RFC 5539: NETCONF Over Transport Layer Security (TLS)
- RFC 5717: Partial Lock RPC for NETCONF
- RFC 6022: NETCONF Monitoring Schema
- RFC 6241: Network Configuration Protocol
- RFC 6242: Using the Network Configuration Protocol over Secure Shell
- RFC 6243: With-defaults capability for NETCONF
- RFC 6470: NETCONF Notification Events
- RFC 6536: NETCONF Access Control Model (NACM)

Note

RFC 4741 and RFC 4742 have been replaced by RFC 6241 and RFC 6242 respectively.

1.2 Configuration Task Summary

NETCONF configuration includes the following tasks:

(1) Configuring Communication Between the NETCONF Server and Client

(2) (Optional) Configuring NETCONF Authentication

1.3 Configuring Communication Between the NETCONF Server and Client

1.3.1 Overview

After the NETCONF function is enabled on the NETCONF server, the NETCONF server can communicate with the NETCONF client to manage network devices.

1.3.2 Restrictions and Guidelines

- The **netconf yang multi-revision** command must be configured before the capability packet (Hello packet) of the NETCONF server is advertised.

- The **unset netconf yang multi-revision** command must be configured before the capability packet (Hello packet) of the NETCONF server is advertised and one YANG module can advertise only its current latest version in the capability notification packet.
- In strict check mode, some XML packets that can be delivered in lightweight check mode may be intercepted.

1.3.3 Prerequisites

The NETCONF protocol is based on the SSH protocol. Before the NETCONF protocol is enabled, the SSH protocol must be configured. For the configuration procedure, see *Configuring SSH*.

1.3.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the attribute parameters of the NETCONF server. The configuration tasks below are optional. Select tasks as required.

- (Optional) Enable the NETCONF service.

  netconf enable

  The NETCONF function is enabled by default.

- (Optional) Configure the maximum number of session connections supported by NETCONF.

  netconf netconf max-session-num *max-sessions-numbers*

  The maximum number of session connections supported by NETCONF is **1** by default.

- (Optional) Configure the timeout time of the Edit-config operation in a NETCONF session.

  netconf timeout *timeout*

  The default timeout time of the Edit-config operation in a NETCONF session is **120** seconds.

- (Optional) Enable the YANG module multi-version notification function.

  netconf yang multi-revision

  The YANG module multi-version notification function of NETCONF is enabled by default.

- (Optional) Configure the NETCONF capabilities.

  netconf capability{ **candidate** | **rollback** | **validate** }

  The NETCONF-related capabilities are not configured by default.

- (Optional) Configure the NETCONF check mode.

  netconf calibration-mode { **lightweight** | **strict** }

  The default NETCONF check mode is lightweight check.

- (Optional) Enable the NETCONF logging function.

  netconf log { **capability** | **edit** | **get** | **session** }

  The NETCONF logging function is disabled by default.

- (Optional) Configure the port ID of the NETCONF server.

netconf port *port-number*

The default port ID of the NETCONF server is 830.

- (Optional) Configure the timeout time of a NETCONF session.

  netconf session-idle-time *num*

  No timeout time is configured for a NETCONF session by default. That is, a session will not be disconnected due to a timeout time by default.

- (Optional) Configure an ACL filtering rule for NETCONF.

  - **netconf access-apply** { **ipv4** *acl-name* | **ipv4** *acl-number* | **ipv6** *acl-name* }

  No ACL filtering rule is configured for NETCONF by default.

- (Optional) Configure the device to save configuration automatically upon the NETCONF commit operation.

  netconf commit auto-save delay-time *time* **delay-count** *count*

  The device does not save configuration automatically upon the NETCONF commit operation by default.

- (Optional) Configure the DSCP value of NETCONF packets.

  netconf dscp { ip | ipv6 } ssh dscp-value

1.4 The default DSCP value of NETCONF packets is 2.Configuring NETCONF Authentication

1.4.1 Overview

The NETCONF authentication mechanism is configured to manage the NETCONF operations of specific clients and the access to NETCONF resources, allowing the clients to perform operations and access data nodes stipulated by NETCONF.

1.4.2 Restrictions and Guidelines

- A user group cannot be named with the asterisk (*), because the symbol indicates all user groups.
- For a single rule list, the rules are successively matched based on the rule configured by users. When an operation matches a permit rule, the operation is permitted. When an operation matches a deny rule or does not match a rule, the operation is denied in the current rule list.
- For multiple rule lists, the rules in the lists are matched based on the rule configuration sequence. When an operation matches a permit rule in a rule list, the operation is permitted. When an operation matches a deny rule, the operation is denied. When an operation does not match any rule in the list, the operation matches the rules of the next list.
- RPC operation methods include Get, Get-Config, Edit-Config, Copy-Config, Delete-Config, Validate, Commit, Cancel-Commit, Discard-Changes, Lock, Kill-Session, and Get-Schema.
- If no rule type is configured, permission control is applied to both RPC operation methods and data node rules.
- Rule control for enabling RPC operation methods can be configured. If the Execute permission is included in the access operation permissions configured under the rules, the rules are matched during RPC operation method check. Otherwise, such rule matching does not occur.

- Rule control for enabling data nodes can be configured. If at least one of the Create, Read, Update and Delete permissions is included in the access operation permissions configured under the rules, the rules are matched during data node check. Otherwise, such rule matching does not occur.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the authentication configuration mode.

netconf acm

(4) Configure a user group and enter the user group mode.

user-group-name *name*

No user group is configured by default.

(5) Associate users with the user group.

user-name *name*

No user is associated with a user group by default.

(6) Quit the authentication configuration mode.

exit

(7) Configure a rule list and enter the rule list mode.

rule-list *name*

No rule list is configured by default.

(8) Associate the user group with the rule list.

user-group { *name* | ***** }

No user group is associated with a rule list by default.

(9) Configure rules and enter the rule mode.

rule *name* { **action deny** | **action permit** }

No rule is configured by default.

(10) Configure rule types.

rule-type { **path** *path* | **rpc-name** { *name* | ***** }

No rule type is configured by default.

(11) Configure access permissions of the rules.

access-operation { * | { create | delete | exec | read | update } * }

No access permission of a rule is configured by default.

1.5 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Table 1-2 Monitoring

Command	Purpose
display netconf-server authorization user-name *name* { **rule-list** [**detail**] | **user-group** }	Displays a rule list that users are associated with, list details, and user group information.
display netconf-server session-info	Displays all session information of NETCONF.
display netconf-server statistics	Displays global statistics of NETCONF.
display netconf-server yang file	Displays all YANG files supported currently.
display netconf-server yang node-path [*module-name*]	Displays all node paths supported currently.
display netconf-server yang tree [*module-name*]	Displays all YANG model trees supported currently.

1.6 Configuration Examples

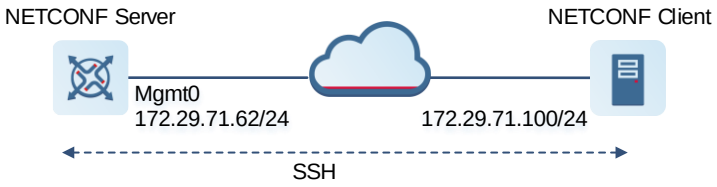
1.6.1 Configuring NETCONF

1. Requirements

NETCONF can be used as a network management tool to meet the high security and scalability requirements for network device management. NETCONF is based on SSH. As a security protocol of the application layer, SSH ensures security of NETCONF. As shown in Figure 1-3, the NETCONF network management software is used to manage and monitor network devices.

2. Topology

Figure 1-3 Configuring NETCONF



3. Notes

- Configure an IP address for the network management interface of the NETCONF server to ensure that the NETCONF server and client are mutually reachable via L3 routes.
- Enable the SSH function on the NETCONF server.
- Configure NETCONF related parameters on the NETCONF server to meet requirements of actual scenarios.
- Connect the NETCONF server to the NETCONF client through SSH to manage and monitor network

devices.

4. Procedure

Configure an IP address for the network management interface of the NETCONF server.

```
Hostname> privilege
Hostname# config-system terminal
Hostname(cfg)# interface mgmt 0
Hostname(cfg-if-Mgmt 0)# ip address 172.29.71.62 255.255.255.0
Hostname(cfg-if-Mgmt 0)# gateway 172.29.71.1
```

Enable the SSH function and generate a local key. For more information about the rule of selecting a key type, see the Usage Guidelines in the **crypto key generate** command in *SSH Commands*.

```
Hostname> privilege
Hostname# config-system terminal
Hostname(cfg)# enable-server service ssh-server
Hostname(cfg)# ssh-crypto key generate rsa
% You already have RSA keys.
% Do you really want to replace them? [yes/no]:y
Choose the size of the rsa key modulus in the range of 512 to 2048
and the size of the dsa key modulus in the range of 360 to 2048 for your
Signature Keys. Choosing a key modulus greater than 512 may take
a few minutes.
Choose the size of the ecc key modulus from (256, 384, 521)

How many bits in the modulus [1024]:2048
% Generating 2048 bit RSA1 keys ...[ok]
% Generating 2048 bit RSA keys ...[ok]
Hostname(cfg)#
```

Create an SSH user with the username **netconf** and set the password to netconf_1234.

```
Hostname(cfg)# username netconf privilege 15 password netconf_1234
```

Caution

To ensure security, you are advised to regularly modify the username and password.

Configure local user authentication of a VTY line.

```
Hostname(cfg)# line vty 0 35
Hostname(cfg-line)# login local
```

Enable the NETCONF service on the NETCONF server.

```
Hostname(cfg)# netconf enable
```

Set the maximum number of session connections supported by the NETCONF server to 6.

```
Hostname(cfg)# netconf max-session-num 6
```

Set the timeout time of the Edit-config operation in a NETCONF session to 100 seconds on the NETCONF server.

```
Hostname(cfg)# netconf timeout 100
```

Enable the YANG module multi-version notification function on the NETCONF server.

```
Hostname(cfg)# netconf yang multi-revision
```

Enable the candidate capability on the NETCONF server.

```
Hostname(cfg)# netconf capability candidate
```

Configure the check mode as lightweight check on the NETCONF server.

```
Hostname(cfg)# netconf calibration lightweight
```

Enable the syslog function related to the NETCONF sessions on the NETCONF server.

```
Hostname(cfg)# netconf log session
```

Set the port ID of the NETCONF service to 5000 on the NETCONF server.

```
Hostname(cfg)# netconf port 5000
```

Set the timeout time of a NETCONF session to 120 seconds on the NETCONF server.

```
Hostname(cfg)# netconf session-idle-time 120
```

Configure IPv4 ACL 1 on the NETCONF server.

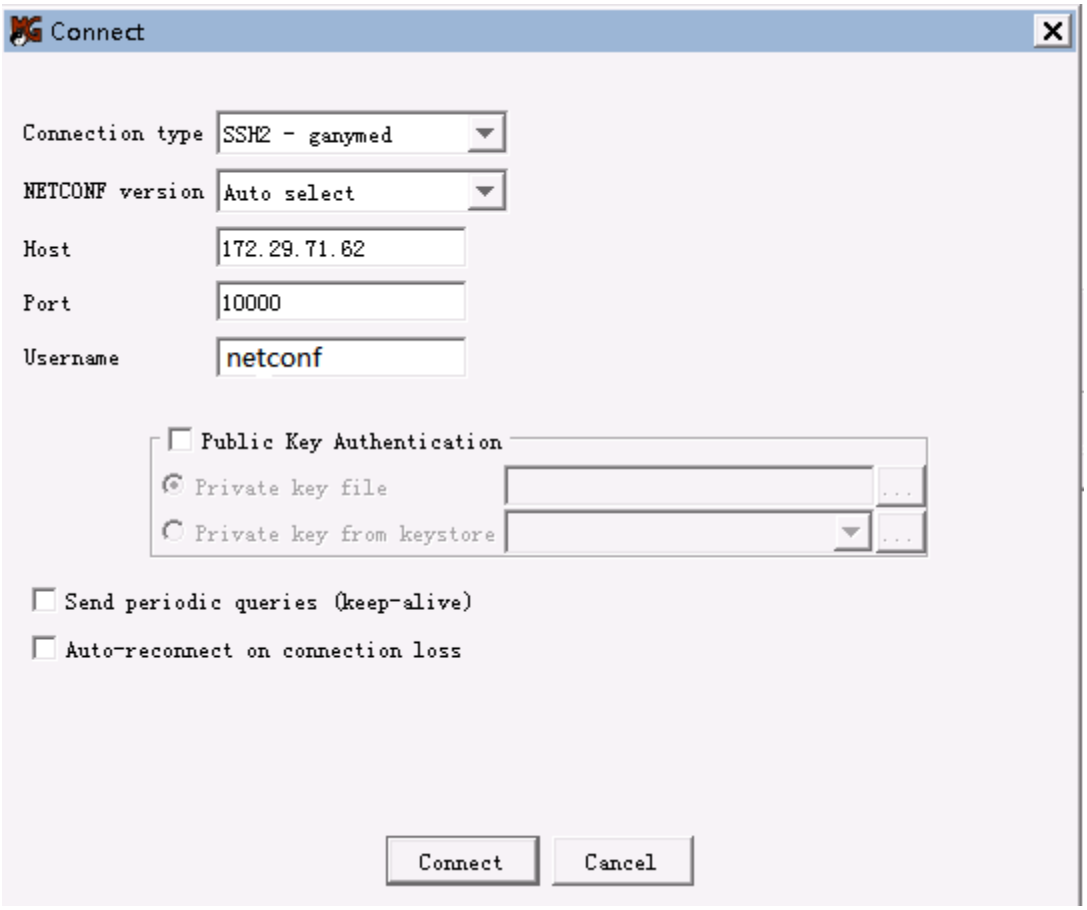
```
Hostname(cfg)# netconf access-class ipv4 1
```

Configure the NETCONF server to save configuration automatically after the NETCONF commit operation.

```
Hostname(cfg)# netconf commit auto-save
```

Log in to the NETCONF server through the SSH protocol by using the NETCONF client software and manage and monitor network devices, as shown in Figure 1-4.

Figure 1-4 Device Login Example Through the NETCONF Client Software



5. Verification

Run the **ping** command to check whether the NETCONF server and client are mutually reachable via L3 routes.

```
Hostname# ping oob 172.29.71.100
Sending 5, 100-byte ICMP Echoes to 172.29.71.100, timeout is 2 seconds:
   < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/3 ms.
Hostname#
```

Run the **display service** command to check whether the SSH service is enabled.

```
Hostname# display service
snmp-agent       : enabled
ssh-server       : enabled
telnet-server : enabled
```

Run the **display netconf session** command to check whether the NETCONF server is connected to the NETCONF client software through the SSH protocol.

```
Hostname# display netconf-server session-info
**********session information**********
                                Session count: 1
```

```
**********************************************************************
Session ID                    : 20
Session version               : 1.1
Session transport             : netconf-ssh
Session login IP              : 172.29.69.21
Session login time            : 2020-12-18T08:26:30Z
Session in rpcs               : 0
Session in bad rpcs           : 0
Session out rpc errors    : 0
Session out notification: 0
Session out rpcs              : 0
Session out send fail     : 0
Session get                   : 0
Session get config            : 0
Session edit config           : 0
Session copy config           : 0
Session delete config     : 0
Session close session     : 0
Session unsupport             : 0
Session lock_or_unlock    : 0
======================================================================
```

6. Configuration Files

NETCONF_Server configuration file

```
hostname Hostname
!
username netconf privilege 15 password netconf_1234
!
netconf enable
netconf log session
netconf capability candidate
netconf calibration-mode lightweight
netconf port 5000
netconf access-apply ipv4 1
netconf dscp ip ssh 30
netconf netconf max-session-num 6
netconf session-idle-time 120
netconf yang multi-revision
!
enable-server service ssh-server
!
interface GigabitEthernet 0
unset switch-port
  ip address 172.29.71.62 255.255.255.0

!
```

```
line vty 0 35
 line-login local
!
end
```

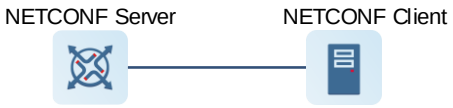
1.6.2 Configuring NETCONF Authentication

1. Requirements

The user **user1** has the permission to get all modules, and the user **user2** has the permission to create SNMP modules.

2. Topology

Figure 1-5 Configuring NETCONF



3. Notes

Configure the Get permission for all modules for **user1**.

Configure the Create permission for SNMP modules for **user2**.

4. Procedure

(1) Configure the Get permission for all modules for **user1**.

Configure a user **user1**.

```
Device> privilege
Device# config-system terminal
Device(cfg)# username user1 password test12345
```

Enter the authentication mode.

```
Device(cfg)# netconf acm
```

Configure a user group **guser1**.

```
Device(cfg-nacm)# user-group-name guser1
```

Associate **user1** with **guser1**.

```
Device(cfg-nacm-user-group)# user-name user1
Device(cfg-nacm-user-group)# exit
```

Configure a rule list rule-list1.

```
Device(cfg-nacm)# rule-list rule-list1
```

Associate **guser1** with rule-list1.

```
Device(cfg-nacm-rule-list)# user-group guser1
```

Configure a permit rule rule1.

```
Device(cfg-nacm-rule-list)# rule rule1 action permit
```

Configure the RPC method as Get.

```
Device(cfg-nacm-rule-list-rule)# rule-type rpc-name get
Device(cfg-nacm-rule-list-rule)# end
```

(2) Configure the Create permission for SNMP modules for **user2**.

Configure a user **user2**.

```
Device# config-system terminal
Device(cfg)# username user2 password test12345
```

Enter the authentication mode.

```
Device(cfg)# netconf acm
```

Configure a user group **guser2**.

```
Device(cfg-nacm)# user-group-name guser2
```

Associate a user group **user2** with a user **guser2**.

```
Device(cfg-nacm-user-group)# user-name user2
```

Configure a rule list rule-list2.

```
Device(cfg-nacm-user-group)# exit
```

Associate a user group **guser2** with a rule list rule-list2.

```
Device(cfg-nacm)# rule-list rule-list2
Device(cfg-nacm-rule-list)# user-group guser2
```

Configure a permit rule rule2.

```
Device(cfg-nacm-rule-list)# rule rule2 act permit
```

Configure a data node type and configure the data node as an SNMP module name.

```
Device(cfg-nacm-rule-list-rule)# rule-type path /snmp:snmp
```

Configure the access permission as Create.

```
Device(cfg-nacm-rule-list-rule)# access-operation create
```

5. Verification

Run the **display netconf-server authorization user-name user1 user-group** command to display user group information of a user **user1**.

```
Device # display netconf-server authorization user-name user1 user-group
UserGroup: guser1
```

Run the **display netconf-server authorization user-name user2 user-group** command to display user group information of a user **user2**.

```
Device # display netconf-server authorization user-name user2 user-group
UserGroup: guser2
```

Run the **display netconf-server authorization user-name user1 rule-list detail** command to display details of the rule list associated with a user **user1**.

```
Device# display netconf-server authorization user-name user1 rule-list detail
RuleList: rule-list1
  ------------------------------------------------------------------
  Rule                 Action    AccessOperation    Type        Entity
  ------------------------------------------------------------------
```

```
  rule1                permit     RCUDX                RPC          get
  [R:Read C:Create U:Update D:Delete X:Exec]
```

Run the **display netconf-server authorization user-name user2 rule-list detail** command to display details of the rule list associated with a user **user2**.

```
Device# display netconf-server authorization user-name user2 rule-list detail
RuleList: rule-list2
  ------------------------------------------------------------------
  Rule                 Action     AccessOperation     Type         Entity
  ------------------------------------------------------------------
  rule2                permit     -C---               PATH         /snmp:snmp
  [R:Read C:Create U:Update D:Delete X:Exec]
```

6. Configuration Files

Device configuration file

```
hostname Device
!
username user1 password test12345
username user2 password test12345
!
netconf acm
 user-group-name guser1
  user-name user1
 user-group-name guser2
  user-name user2
 rule-list rule-list1
  user-group guser1
  rule rule1 action permit
   rule-type rpc-name get
 rule-list rule-list2
  user-group guser2
  rule rule2 action permit
   access-operation create
   rule-type path /snmp:snmp
!
end
```

Contents

1 Configuring gRPC 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 9
1.3 Configuring Basic Functions of gRPC 9
1.3.1 Overview 9
1.3.2 Restrictions and Guidelines 9
1.3.3 Procedure 9
1.4 Configuring the gRPC Response Timeout Time 10
1.4.1 Overview 10
1.4.2 Restrictions and Guidelines 10
1.4.3 Procedure 10
1.5 Enabling Stream-based RPC for Events 10
1.5.1 Overview 10
1.5.2 Restrictions and Guidelines 10
1.5.3 Procedure 10
1.6 Configuring the DSCP Value of gRPC Packets 11
1.6.1 Overview 11
1.6.2 Restrictions and Guidelines 11
1.6.3 Procedure 11
1.7 Configuring Event Types Supported by gRPC 11

1.7.1 Overview 11
1.7.2 Restrictions and Guidelines 12
1.7.3 Procedure 12
1.8 Configuring gRPC Server Attributes 13
1.8.1 Configuration Tasks 13
1.8.2 Loading an SSL/TLS Certificate 13
1.8.3 Loading a CA Certificate 13
1.8.4 Configuring an ACL Filtering Rule 14
1.8.5 Configuring GZIP Format for Compressed Data Transmission 14
1.9 Configuring gRPC to Support Pre-configuration Functions 15
1.9.1 Configuration Tasks 15
1.9.2 Configuring gRPC to Pre-configure a Subscription Server and Event 15
1.9.3 Configuring gRPC to Support User Login Information of the Functional Server 16
1.10 Configuring a YANG Model Path 17
1.10.1 Overview 17
1.10.2 Restrictions and Guidelines 17
1.10.3 Procedure 17
1.11 Configuring the gRPC Subscription Function 17
1.11.1 Overview 17
1.11.2 Restrictions and Guidelines 18
1.11.3 Configuration Tasks 18
1.11.4 Configuring gRPC to Specifying a Source Interface to Send Data 18
1.11.5 Configuring gRPC Destination Server 18
1.11.6 Configuring gRPC Subscription 19

1.12 Configuring gRPC to Support the Login and Logout Function 20
1.12.1 Overview 20
1.12.2 Restrictions and Guidelines 20
1.12.3 Procedure 20
1.13 Configuring a Third-party Defined RPC Method 21
1.13.1 Overview 21
1.13.2 Procedure 21
1.14 Configuring a Dial-out RPC Method 21
1.14.1 Overview 21
1.14.2 Procedure 21
1.15 gRPC Packet Retransmission Timeout 22
1.15.1 Overview 22
1.15.2 Procedure 22
1.16 Monitoring 22
1.17 Configuration Examples 23
1.17.1 Configuring Basic Functions of gRPC 23

1 Configuring gRPC

1.1 Introduction

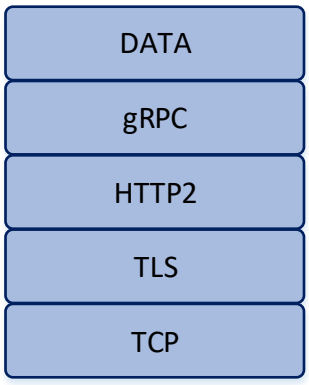
1.1.1 Overview

The Google Remote Procedure Calls (gRPC) is a high-performance open source software framework developed by Google using the Hyper Text Transfer Protocol 2 (HTTP2) as a data transmission protocol. gRPC provides a network configuration and management method that supports multiple programming languages. On the gRPC framework, secondary development can be performed.

1.1.2 Principles

1. Protocol Structure

Figure 1-1 Protocol Stack Layers



As shown in Figure 1-1, gRPC stack has five layers. The function description of each layer is as follows.

- DATA

  The layer is also known as the data content layer. It contains data from both sides of the communication. Based on specific service definitions, the layer consists of the following two elements.

  - Encoding format: The layer usually uses GPB and JSON as the encoding format. Currently the device only supports JSON.
  - Data abstraction: The YANG model is usually used to define the data format. The device also supports defining the data format based on the event ID.

- gRPC

  RPC interaction format is defined based on gRPC protocol. RPC is defined in the **proto** file, such as **gnmi.proto**, **dial-out.proto**, and so on.

- HTTP/2

  gRPC runs on HTTP/2, mainly using HTTP/2 features such as the bidirectional flow, flow control, header compression, multiplexing requests on a single connection.

- TLS

A gRPC protocol that uses TLS encryption.

- TCP

  An underlying communication protocol based on TCP connections.

2. Telemetry Mode

Figure 1-2 A gRPC Connection



(1) As shown in Figure 1-2, the network management system establishes connection with the device through gRPC. The network management system collects device data through gRPC to monitor and manage the device remotely.

The device supports the following two telemetry modes:

- Dial-in mode

  The device acts as a gRPC server and the network management system acts as a gRPC client. The network management system initiates connection to the device to subscribe to device data or deliver configurations.

  Dial-in mode applies to small networks and the scenarios where the network management system needs to deliver configurations through gRPC connections.

- Dial-out mode

  The device acts as a gRPC client and the network management system acts as a gRPC server. The device initiates connection to the network management system and pushes subscribed data to the network management system.

  Dial-out mode applies to larger networks where there are multiple devices.

3. Data Abstraction

The device supports the following two methods to define the data format.

- Yang data modeling

  The device can get, set and subscribe Yang model data. Yang model defines the data grammar and displays the data using the JSON format. The device supports both OpenConfig Yang and customized private RGOS Yang. See gRPC data format description document for Yang models supported by current devices. You are advised to use Yang model to define the data format because it is common and comprehensive.

- Event data modeling

  The device supports formatting data base on events. One event corresponds to a JSON data group. See gRPC data format description document for events supported by current devices. The event data modeling has been applied to multiple customer scenarios. It is easier to use the event data modeling than Yang data modeling. The customers who want to establish and use gRPC connections fast are advised to adopt the event data modeling.

(1) Proto files based on Yang data modeling

Yang data consists of Yang paths and JSON data. The following describes the basic data format in the proto definition.

```
message Path {
   repeated string element = 1 [deprecated=true]; // Abandoned.
   string origin = 2;                                  // A reserved field.
   repeated PathElem elem = 3;                         // Node names of a path.
   string target = 4;                                  // A reserved field.
}

message PathElem {
   string name = 1;                                    // Node names of a path.
   map<string, string> key = 2;                        // A keyword can be added if it is a list. node.
}

message Update {
   Path path = 1;                                      // A YANG path.
   Value value = 2 [deprecated=true];                  // Abandoned.
   TypedValue val = 3;                                 // JSON data of a YANG path.
   uint32 duplicates = 4;                              // A reserved field.
}

message TypedValue {
   oneof value {
      string string_val = 1;
      int64 int_val = 2;
      uint64 uint_val = 3;
bool bool_val = 4;
      bytes bytes_val = 5;
      float float_val = 6;
      Decimal64 decimal_val = 7;
      ScalarArray leaflist_val = 8;
      google.protobuf.Any any_val = 9;
      bytes json_val = 10;                             // Only supports json_val.
      bytes json_ietf_val = 11;
      string ascii_val = 12;
      bytes proto_bytes = 13;
   }
}
```

- Proto files based on Yang data modeling are all open source files. Proto files of the dial-in mode are gnmi.proto files and those of the dial-out mode are dial-out.proto files.

Note

gRPC Network Management Interfac (gnmi) is an interface developed by Google based on OpenConfig.

- Proto files based on the dial-in mode of Yang data modeling

The device can use the following three RPC types to implement the gnmi.proto definition.

- Get RPC: It is used to obtain current device data. The implementation and content description of request and response messages are as follows.

```
message GetRequest {
  Path prefix = 1;                                  // A shared prefix of multiple Yang paths.
  repeated Path path = 2;                           // Multiple Yang paths.
  enum DataType {
    ALL = 0;                                        // All data elements.
    CONFIG = 1;                                     // Config (rw) only elements.
    STATE = 2;                                      // State (ro) only elements.
    OPERATIONAL = 3;
  }
  DataType type = 3;                                // Only supports ALL.
  Encoding encoding = 5;                            // Only supports json.
  repeated ModelData use_models = 6;                // A reserved field.
  repeated gnmi_ext.Extension extension = 7;        // A reserved field.
}

message GetResponse {
  repeated Notification notification = 1;     // Data values.
  Error error = 2 [deprecated=true];          // Errors that occurred in the Get.
  repeated gnmi_ext.Extension extension = 3;
}

message Notification {
  int64 timestamp = 1;              // A nanosecond timestamp.
  Path prefix = 2;                  // A shared prefix of multiple Yang paths.
  string alias = 3;
  repeated Update update = 4;       // Pushed data.
  repeated Path delete = 5;         // A reserved field.
  bool atomic = 6;
}
```

- Set RPC: It is used to deliver configurations to devices. The implementation and content description of request and response messages are as follows.

```
message SetRequest {
  Path prefix = 1;                      // A shared prefix of multiple Yang paths.
  repeated Path delete = 2;             // A Yang path to be deleted.
  repeated Update replace = 3;          // A Yang path to be updated with path and val to be added into.
  repeated Update update = 4;           // A Yang path to be updated with path and val to be added into.
  repeated gnmi_ext.Extension extension = 5;          // A reserved field.
}

message SetResponse {
  Path prefix = 1;                              // A shared prefix of multiple Yang paths.
```

```
  repeated UpdateResult response = 2;     // If the delivery succeeds, the device sends the response
       message. If not, the device sends a gRPC error message.
  Error message = 3 [deprecated=true];                 // Abandoned. If the delivery fails, the device uses the
 failure mechanism of gRPC.
  int64 timestamp = 4;                                 // A nanosecond timestamp.
  repeated gnmi_ext.Extension extension = 5;       // A reserved field.
}
```

- Subscribe RPC: It is used to subscribe to a device to collect data. The implementation and content description of request and response messages are as follows.

```
message SubscribeRequest {
  oneof request {
    SubscriptionList subscribe = 1;                // Subscription information.
    Poll poll = 3;                                 // A reserved field.
    AliasList aliases = 4;                         // A reserved field.
  }
  repeated gnmi_ext.Extension extension = 5;     // A reserved field.
}
message SubscriptionList {
  Path prefix = 1;                                 // A shared prefix of multiple Yang paths.
  repeated Subscription subscription = 2;          // Subscription information.
  bool use_aliases = 3;                            // A reserved field.
  QOSMarking qos = 4;                              // A reserved field.
  enum Mode {
    STREAM = 0;
    ONCE = 1;
    POLL = 2;
  }
  Mode mode = 5;                                   // A reserved field.
  bool allow_aggregation = 6;                      // A reserved field.
  repeated ModelData use_models = 7;               // A reserved field.
  Encoding encoding = 8;                           // Only supports JSON.
  bool updates_only = 9;                           // A reserved field.
}

message Subscription {
  Path path = 1;                                   // A YANG path.
  SubscriptionMode mode = 2;                       // Interval or triggered subscription.
  uint64 sample_interval = 3;                      // Intervals of interval subscription. The field is
unnecessary for triggerd subscription.
```

```
  bool suppress_redundant = 4;                  // A reserved field.
  uint64 heartbeat_interval = 5;                // A reserved field.
}

message SubscribeResponse {
  oneof response {
    Notification update = 1;                    // Notification, same to the returned Get messages.
    bool sync_response = 3;                     // A reserved field.
    Error error = 4 [deprecated=true];          // Abandoned.
  }
  repeated gnmi_ext.Extension extension = 5;    // A reserved field.
```

- Proto files based on the dial-out mode of Yang data modeling

  Dial-out.proto is a proto provided by SONIC based on gnmi and has only one publish RPC. The request message is SubscribeRespone. The collected data should be the same to the data collected by Get and Subscribe PRC. The response message is defined as follows.

```
message PublishResponse {
  int64 timestamp = 1;          // A nanosecond timestamp. Currently the device does not process the collected data. The field can be omitted.
  Path prefix = 2;              // A shared prefix of multiple Yang paths. Currently the device does not process the collected data. The field can be omitted.
  string alias = 3;             // An alias. Currently the device does not process the collected data. The field can be omitted.
  repeated Path path = 4;       // A YANG path. Currently the device does not process the collected data. The field can be omitted.
}
```

(2) Proto files based on the event data modeling

- Proto files based on the dial-in mode of the event data modeling

  Proto files based on the dial-in mode of the event data modeling are private definitions. The file name is Ruijie-subscr.proto. The method supports SubscrAuthSend PRC for authentication, SubscrGetSend PRC for collecting data and SubscrSetSend PRC for setting. The definition and description of proto are as follows.

```
syntax = "proto3";

package Ruijie_subscr;

service Subscr {
  rpc SubscrAuthSend (SubscrAuthRequest) returns (SubscrAuthReply) {}
  rpc SubscrGetSend (SubscrGetRequest) returns (stream SubscrGetReply) {}
  rpc SubscrSetSend (SubscrSetRequest) returns (stream SubscrSetReply) {}
}

message SubscrDeviceMsg {
```

```
  string producer_name = 1;                  // Manufacture name.
  string device_name = 2;                    // Device name.
  string device_model = 3;                   // Device model.
}

message SubscrAuthInfo {
  string username = 1;                       // Username of the subscribed server.
  string password = 2;                       // Password of the subscribed server.
}

message SubscrGetInfo {
  string username = 1;                       // Username of the subscribed server.
}

message SubscrAuthRequest {
  SubscrAuthInfo info = 1;                   // Authentication information of the subscribed server.
  bool is_login = 2;                         // The field of login is true and the field of logout is false.
}

message SubscrAuthReply {
  SubscrDeviceMsg device_info = 1;           // Device information.
  sint32 ret = 2;                            // Indicates whether the authentication succeeds. 0 stands for
success and 1 stands for failure.
  string detail = 3;                         // An authentication information string.
}

message SubscrGetRequest {
  SubscrGetInfo info = 1;                    // Authentication information of the subscribed server.
  uint32 json_event = 2;                     // A specific event.
  string json_string = 3;                    // Event parameters in the standard JSON format.
}

message SubscrGetReply {
  SubscrDeviceMsg device_info = 1;           // Device information.
  sint32 ret = 2;                            // Indicates whether the data processing succeeds. 0 stands for
success and 1 stands for failure.
  uint32 json_event = 3;                     // A specific event.
  string json_string = 4;                    // Reported information strings in the standard JSON format.
}

message SubscrSetRequest {
  SubscrGetInfo info = 1;                    // Authentication information of the subscribed server.
  uint32 json_event = 2;                     // A specific event.
  string json_string = 3;                    // Event parameters in the standard JSON format.
}
```

```
message SubscrSetReply {
  SubscrDeviceMsg device_info = 1;          // Device information.
  sint32 ret = 2;                           // Indicates whether the data processing succeeds. 0 stands for
success and 1 stands for failure.
  uint32 json_event = 3;                    // A specific event.
  string json_string = 4;                   // The reason for the result of data processing in the standard
JSON format.
}
```

- Proto files based on the dial-out mode of the event data modeling

  Proto files based on the dial-in mode of the event data modeling are private definitions. The file name is Ruijie-subscr.proto. The method supports SubscrAuthSend PRC for authentication, SubscrGetSend PRC for collecting data and SubscrSetSend PRC for setting. The definition and description of proto are as follows.

```
syntax = "proto3";

package Ruijie_json;

service Json {
  rpc JsonSend (JsonRequest) returns (JsonReply) {}
  rpc JsonStreamSend (stream JsonRequest) returns (stream JsonReply) {}
}

message JsonDeviceMsg {
  string producer_name = 1;                 // Manufacture name.
  string device_name = 2;                   // Device name.
  string device_model = 3;                  // Device model.
  string device_ip = 4;                     // Device IP
  string device_vendor = 5;                 // Device vendor
  string device_version = 6;                // Device version
}

message JsonRequest {
  JsonDeviceMsg device_info = 1;            // Device information.
  uint32 json_event = 2;                    // A specific event.
  string json_string = 3;                   // Reported information strings in the standard JSON format.
  int64 timestamp = 4 [deprecated=true];
  string sensor_path = 5 [deprecated=true];
}

message JsonReply {
  sint32 ret = 1;                           // Indicates whether the data processing succeeds. 1 stands for
success and 0 stands for failure.
}
```

1.2 Configuration Task Summary

gRPC configuration includes the following tasks:

(1) Configuring Basic Functions of gRPC

(2) Configuring the gRPC Response Timeout Time

(3) Enabling Stream-based RPC for Events

(4) Configuring the DSCP Value of gRPC Packets

(5) Configuring Event Types Supported by gRPC

(6) Configuring gRPC Server Attributes

- Configuring gRPC to Pre-configure a Subscription Server and Event
- Configuring gRPC to Support User Login Information of the Functional Server

(7) Configuring a YANG Model Path

(8) Configuring the gRPC Subscription Function

- (Optional) Configuring gRPC to Specifying a Source Interface to Send Data
- Configuring gRPC Destination Server
- Configuring gRPC Subscription

(9) (Optional) Configuring gRPC to Support the Login and Logout Function

(10) Configuring a Third-party Defined RPC Method

(11) Configuring a Dial-out RPC Method

1.3 Configuring Basic Functions of gRPC

1.3.1 Overview

The gRPC function is enabled, and the gRPC Client and gRPC Server are created. The gRPC Client uploads data to the server, and the gRPC Server parses authentication messages sent from the server and obtains the message content.

1.3.2 Restrictions and Guidelines

- The connection between the device and server must be normal.
- The port number of the local gRPC Server is 50051.
- A gRPC instance is created by running a **gRPC** command, the gRPC function is enabled, and the gRPC process configuration mode is enabled.

1.3.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the gRPC function.

grpc

The gRPC function is disabled by default.

1.4 Configuring the gRPC Response Timeout Time

1.4.1 Overview

The data response timeout time is configured for the gRPC Server.

1.4.2 Restrictions and Guidelines

- After the data response timeout time is configured for the gRPC process and the collector delivers a Get or Set operation, a timeout message is returned if the server does not obtain corresponding data within the timeout time.
- The value of the timeout time should be configured based on the actual scenario. If the configured timeout time is too short, part of the required data may not be returned in the timeout period during a Get operation.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure the gRPC data response timeout time.

reply-timeout *time*

The default data authentication timeout time on the gRPC Server is **10** seconds.

1.5 Enabling Stream-based RPC for Events

1.5.1 Overview

Configure events to push gRPC packets through stream-based RPC.

1.5.2 Restrictions and Guidelines

- Events push gRPC packets through JsonSend RPC of ruijie-json.proto by default. After stream-based RPC is enabled, events push gRPC packets through JsonStreamSend RPC of ruijie-json.proto.
- After the RPC mode is changed, the current gRPC channel will be disconnected and a new session will be set up.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Enable stream-based RPC for events.

event stream enable

Stream-based RPC is disabled by default.

1.6 Configuring the DSCP Value of gRPC Packets

1.6.1 Overview

Configure the DSCP value of communication packets for the gRPC server and the gRPC client.

1.6.2 Restrictions and Guidelines

- After the DSCP value of gRPC communication packets is modified, the gRPC server and gRPC client services on the device are restarted. To ensure successful session establishment, you are advised to reconnect the gRPC server after it is restarted. The gRPC client automatically reconnects after it is restarted. The DSCP value takes effect after the gRPC server and gRPC client are reconnected and a session is established between them.
- If the **dscp** command does not carry the **ipv6** keyword, the DSCP value of IPv4 gRPC communication packets is configured. If the command carries the **ipv6** keyword, the DSCP value of IPv6 gRPC communication packets is configured.
- The range of the DSCP value is from 0 to 63.

1.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure the DSCP value of gRPC communication packets.

dscp [**ipv6**] *dscp-value*

The default DSCP value of gRPC communication packets is **0**.

1.7 Configuring Event Types Supported by gRPC

1.7.1 Overview

- The device is configured with all realtime events of concern which are subscribed to.
- The device is configured with all periodic events of concern which are subscribed to.
- The timer time of the periodic events is modified.

1.7.2 Restrictions and Guidelines

- The event which the server subscribes to cannot effect unless a specified event which is subscribed to is configured on the device.
- If the device subscribes to a realtime event, a realtime packet is generated and sent to the subscription server when the threshold of the realtime events is triggered. If the threshold of the realtime events is triggered frequently, many same packets are generated within a short time and sent to the subscription server. As a result, the cost of the server and the device is added. To prevent such threshold from being triggered frequently, Run the **subscr-realtime-interval** command to suppress realtime events to generate new packets within a period.
- When any periodic events of concern which are subscribed to is configured and the NETCONF manager sends periodic events which are subscribed to, the corresponding periodic timer starts and gRPC data is sent to the subscription server.
  - If the NETCONF manager does not subscribe to any periodic time, the timer period is determined by the **subscr-sample-interval** command.
  - If the NETCONF manager subscribes to both the periodic events and period time, the period time is subject to the subscription time of the NETCONF manager.
- After any events of concern which are subscribed to are configured by the gRPC process, the gRPC process waits for the NETCONF manager to send the subscription events:
  - If the received subscription message matches the local subscription event, the event triggering gRPC reporting server function is enabled.
  - When the NETCONF manager sends an unsubscription event and the received unsubscription message matches the local subscription event, the event triggering gRPC reporting server function is disabled.
- If no events of concerns which are subscribed to are configured by the gRPC process, the gRPC process ignores the subscription events and unsubscription events sent by the NETCONF manager.

1.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure an event of concern which is subscribed to on the device.

subscr [**realtime** | **sample**] **enable**

No events of concern which are subscribed to are configured by default.

(5) (Optional) Configure the timer time of the periodic events.

subscr-**sample**-**interval** *subscr-sample-interval-second* [**millisecond** *subscr-sample-interval-millisecond*]

The default periodic event timer time is **10** seconds.

(6) (Optional) Configure the realtime event suppression time.

subscr-realtime-interval { **all** *subscr-realtime-time* | **realtime** *json-event subscr-realtime-time* }

The realtime event suppression function is not enabled by default.

1.8 Configuring gRPC Server Attributes

1.8.1 Configuration Tasks

gRPC server attribute configuration includes the following tasks:

- Loading an SSL/TLS Certificate
- Loading a CA Certificate
- Configuring an ACL Filtering Rule
- Configuring GZIP Format for Compressed Data Transmission

1.8.2 Loading an SSL/TLS Certificate

1. Overview

After an SSL/TLS certificate is loaded, TLS encryption is enabled.

2. Restrictions and Guidelines

Before loading an SSL/TLS certificate, copy the certificate to the flash partition of the device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Load the SSL/TLS certificate of the gRPC server.

server certificate load { pem *cert-filename* **private-key** *key-filename* **} | { pfx** *cert-filename* **} [password** *password-text* **]**

No SSL/TLS certificate is loaded by default.

1.8.3 Loading a CA Certificate

1. Overview

After a CA certificate is loaded, certificate authentication is performed for the client.

2. Restrictions and Guidelines

Before loading a CA certificate, copy the certificate to the flash partition of the device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Load the CA certificate of the gRPC server.

server trusted-ca load *cert-filename*

No gRPC server CA certificate is loaded by default.

1.8.4 Configuring an ACL Filtering Rule

1. Overview

Packet filtering will be performed on the connected clients based on the configured ACL.

2. Restrictions and Guidelines

Create an ACL before specifying it as a filtering rule.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configuring the ACL filtering rule.

server access-class { **ipv4** *acl-name* | **ipv4** *acl-number* | **ipv6** *acl-name* }

No ACL filtering rule is configured by default.

1.8.5 Configuring GZIP Format for Compressed Data Transmission

1. Overview

Configure the gRPC server to transmit data in GZIP format.

2. Restrictions and Guidelines

After being configured with GZIP format for data transmission, the gRPC server will be restarted. You are advised to reconnect the client after the gRPC server is restarted. The gRPC server will transmit data in GZIP format upon session creation.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure the gRPC server to transmit data in GZIP format.

server compression gzip

No compression format is configured for the gRPC server by default.

1.9 Configuring gRPC to Support Pre-configuration Functions

1.9.1 Configuration Tasks

gRPC configuration includes the following tasks:

- Configuring gRPC to Pre-configure a Subscription Server and Event
- Configuring gRPC to Support User Login Information of the Functional Server

1.9.2 Configuring gRPC to Pre-configure a Subscription Server and Event

1. Overview

- After a server and event which are subscribed to is pre-configured, gRPC packets of corresponding events can be reported to the server.
- After a subscribing standby server is pre-configured, its services can be switched over to the standby server and packets are automatically sent to the standby server when the active server becomes faulty.
- After the GZIP format is configured for event data reporting, the gRPC clients associated with the subscription server will set up new sessions and report event data in GZIP format.
- If the user server is specified with a source IP address or a VRF, the device does not use the IP address or the VRF of the source interface for gRPC server connection.

2. Restrictions and Guidelines

- The address and port number of the subscription server must be legitimate.
- The events and parameters which are subscribed to must be correct and valid.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure gRPC to support a subscription server.

user-server { *ipv4-address* | **ipv6** *ipv6-address* } *port-number* **[source ip** <*source-ipv4-address*> **| ipv6 <** *source-ipv6-address*>] [**vrf** <*vrf_name*>]

No subscription server is pre-configured by default.

(5) Configure gRPC to support a subscription event.

type *json-event* **value** *value-interval* [**value-extend** *value-extend-interval*]

The event subscription function for a specified server is not enabled by default.

(6) (Optional) Configure gRPC to support the heartbeat detection function.

heartbeat inform [**interval** *interval*] [**keepalive-times** *keepalive-times*]

The heartbeat sending function is not enabled by default.

(7) (Optional) Configure gRPC to pre-configure a standby server for the subscription server.

standby-server *ipv4-address port-number*

No standby server is pre-configured for the subscription server by default.

Before configuring a standby server, enable the heartbeat detection function. When an active server becomes faulty, data can be automatically sent to the standby server.

(8) (Optional) Configure the gRPC protocol for the destination server.

protocol grpc { **tls** | **no-tls** }

The default protocol is the cleartext gRPC.

(9) (Optional) Configure the subscription server to report event data in GZIP format.

compression gzip

No format is specified for event data reporting by default.

(10) Exit the gRPC subscription server mode.

exit-user-server

1.9.3 Configuring gRPC to Support User Login Information of the Functional Server

1. Overview

After user login information is pre-configured for the server, the specified server can obtain correct information after initiating an obtainment event request.

2. Restrictions and Guidelines

- Running a configured command, the device can pre-create user login information of the server that passes authentication, simulate the server to initiate a login request, pass the authentication, and save the user login information.
- Running a configured command, the device can delete user login information of the server that passes authentication, simulate the server to initiate a logout request, and delete the user login information.
- The address of the server must be legitimate.
- The sequence number of the login user must be unique.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure gRPC to pre-configure user login information of the server.

user-client *user-id username* { *ipv4-address* | **ipv6** *ipv6-address* }

No user login information is pre-configured for the server by default.

1.10 Configuring a YANG Model Path

1.10.1 Overview

After the YANG model path and data source are collected, data can be sent to the corresponding server based on the path in the subscription association relationship.

1.10.2 Restrictions and Guidelines

- The YANG model path must be legitimate and corresponding data is available on the path.
- In the YANG model path group, a specific YANG model path is configured. For the detailed compilation method of the path, see the displayed information of the **display grpc yang tree** and **display grpcnode-path** commands in **gRPC Command Reference**.
- After a gRPC YANG model is configured, the system switches to the gRPC YANG model configuration group mode.
- Multiple YANG model paths can be configured for this group in this mode.

1.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure a YANG model path group.

sensor-group *sensor-group-id*

No YANG model path group is configured by default.

(5) Configure a gRPC YANG model path and its group.

sensor-**path** *sensor-path*

No YANG model path is configured by default.

(6) Exit the gRPC YANG model configuration group mode.

exit-sensor-group

1.11 Configuring the gRPC Subscription Function

1.11.1 Overview

After a subscription association relationship is configured, data of the associated YANG model path group is collected and sent to the destination server of the associated destination group.

1.11.2 Restrictions and Guidelines

A YANG model path group must be configured in advance.

1.11.3 Configuration Tasks

gRPC subscription configuration includes the following tasks:

(1) (Optional) Configuring gRPC to Specifying a Source Interface to Send Data

(2) Configuring gRPC Destination Server

(3) Configuring gRPC Subscription

1.11.4 Configuring gRPC to Specifying a Source Interface to Send Data

1. Overview

gRPC is configured to send packets based on the IP address of a specified interface.

2. Restrictions and Guidelines

- The bound interface must be up and be configured with an IP address.
- If a server receives only packets from a specified source address, this command can be used to configure the specified source interface.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) (Optional) Configure a source interface of realtime data and periodic data.

subscr-source-interface *interface-type interface-number*

The source interface of realtime data and periodic data is a random value by default.

1.11.5 Configuring gRPC Destination Server

1. Overview

The address of a destination server is configured for the server to subscribe to a YANG model path in the subscription association relationship.

2. Restrictions and Guidelines

- The address and port number of the server must be legitimate.
- After a gRPC destination group is configured, the system switches to the gRPC destination group configuration mode.
- Multiple destination addresses can be configured for this group in this mode.
- After the GZIP format is configured for gNMI data reporting, the gRPC clients in the destination configuration

group mode will set up new sessions and report gNMI data in GZIP format.

- If the gRPC destination address is specified with a source IP address or a VRF, the device does not use the IP address or the VRF of the source interface for gRPC server connection.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure a gRPC destination group.

destination-group *destination-group-id*

No destination group is configured for the subscription association relationship by default.

(5) Configure a gRPC destination address.

- Configure an IPv4 destination address.

  ip *ipv4-address* **port** *port-number* ***[source ip*** *<source-ipv4-address>* ***] [vrf*** *<vrf_name>* ***]***

  No IPv4 address or port is configured for the destination server by default.

- Configure an IPv6 destination address.

  ipv6 *ipv6-address* **port** *port-number* ***[source ipv6*** *<source-ipv6-address>* ***] [vrf*** *<vrf_name>* ***]***

  No IPv6 address or port is configured for the destination server by default.

(6) Configure a protocol for the destination server.

protocol grpc { **tls** | **no-tls** }

The default protocol is the gRPC protocol that supports plain text transmission.

(7) Configure the destination group to report gNMI data in GZIP format.

compression gzip

No format is specified for gNMI data reporting by default.

(8) Exit the gRPC destination group configuration mode.

exit-destination-group

1.11.6 Configuring gRPC Subscription

1. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure a subscription association relationship group.

subscription *subscriptionid*

No subscription association relationship is configured between the YANG model path group and destination group by default.

(5) Configure the associated YANG model path group and its period interval.

sensor-group *sensor-group-id* [**sample-interval** *sample-interval*]

(6) Configure a destination group of the subscription association relationship.

destination-**group** *destination-group-id*

No destination group is configured for the subscription association relationship by default.

(7) Exit the gRPC subscription association relationship configuration mode.

exit-grpc-subscription

1.12 Configuring gRPC to Support the Login and Logout Function

1.12.1 Overview

The authentication mode of the gRPC login and logout function is configured as authentication-exemption, AAA local authentication, or AAA server authentication (timeout time and retry times are configurable).

1.12.2 Restrictions and Guidelines

- If the authentication mode is configured as AAA server authentication, the new-model must be enabled.
- The authentication-exemption mode is not recommended.
- You are not advised to modify the timeout time and retry times. During server login in the AAA server authentication mode, if the retry times exceed a limit or the timeout time is too long, AAA may be locked for 15 minutes. The lock time can be canceled by running the AAA command **cleanup aaa local user lockout all**.
- Upon receiving a login request from the server, the device selects one of the three modes based on the configured authentication mode. After the server passes the authentication, the user login information, namely, IP address and username of the server, is recorded. The device returns responses to the obtainment event requests of the server that passes the authentication.

1.12.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure an authentication mode of the gRPC login and logout function.

authen login { **local** | **authentication** *mlist* }

The local authentication mode, namely, **authen login local** is used by default.

(5) (Optional) Configure the AAA server authentication attribute of the gRPC login and logout function.

authen aaa-config { **retry** *retry-times* | **timeout** *timeout-time* }

The retry times and timeout time of AAA server authentication are 1 and 4 seconds, respectively, by default.

This authentication mode can be configured only after AAA configuration (aaa new-model) is completed.

1.13 Configuring a Third-party Defined RPC Method

1.13.1 Overview

After a third-party defined RPC method is configured, the RPC method can be used to collect device information and data or perform RPC-related operation.

1.13.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure a third-party defined RPC method.

rpc { **gnmi** | **openr** } **enable**

No RPC defined by gNMI or openR is configured by default.

1.14 Configuring a Dial-out RPC Method

1.14.1 Overview

An RPC method is configured for dial-out so that the RPC method can be selected to receive and send messages.

1.14.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure a dial-out RPC method.

dial-out rpc { **sonic** | **custom** }

The RPC method defined by the open source of Sonic is used by default.

1.15 gRPC Packet Retransmission Timeout

1.15.1 Overview

After the gRPC packet retransmission timeout is configured, gRPC can identify connection exception upon timeout.

1.15.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the gRPC configuration mode.

grpc

(4) Configure the gRPC packet retransmission timeout.

tcp_user_timeout *value*

The default gRPC packet retransmission timeout is 30 seconds.

1.16 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** commands to clear information.

Caution

Running the **cleanup** commands may lose vital information and thus interrupt services.

Table 1-1 gRPC Monitoring

Command	Purpose
cleanup grpc channel [*ipv4-address port-number*] [**counter**]	Clears statistics about the server which gRPC subscribes to.
cleanup grpc subscr [**sample**] [**counter**]	Clears data about periodic events and realtime events of gRPC.
display grpc channel [**counter**]	Displays server address, port number, subscription events, and different statistics.
display grpc client	Displays gRPC server login information.
display grpc event subscription	Displays subscription delivered through events.
display grpc server information	Displays information related to the gRPC server.
display grpc status	Displays gRPC running status information.

Command	Purpose
display grpc subscr [**realtime** | **sample**]	Displays status and statistics of events which are subscribed to.
display grpc openconfig subscription	Displays information about static subscription association relationship (openconfig-telemetry.proto definition) of the device.
display grpc gnmi subscription	Displays information about subscription association relationship generated with the RPC method defined by gnmi.proto.
display grpc openr support-rpc	Displays openR RPC methods supported by the device.
display grpc yang tree [*module-name*]	Displays the YANG model tree that is loaded by gRPC.
display grpc yang node-path [*module-name*]	Displays paths corresponding to YANG nodes.

1.17 Configuration Examples

1.17.1 Configuring Basic Functions of gRPC

1. Requirements

A tested device is connected to a server and reports gRPC data.

2. Topology

Figure 1-3 Configuring Basic Function of gRPC



3. Notes

- Ensure that the device and the PC are mutually reachable via the Layer 3 route.
- Enable the gRPC function on the device.
- On the device, configure events in which the server must be reported.

4. Procedure

(1) Configure that the device and the PC are mutually reachable via the Layer 3 route.

```
Device> privilege
Device#config-system terminal
Device(cfg)#interface gigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# ip address 192.168.1.100 255.255.255.0
Device(cfg-if-GigabitEthernet 0/1)# exit
```

(2) Configure the gRPC function.

Enable the gRPC function.

```
Device(cfg)# grpc
```

Configure gRPC to bind a specified interface to send packets.

```
Device(cfg-grpc)# subscr-source-interface GigabitEthernet 0/1
```

(3) Configure events in which the server must be reported.

Configure the aaa new-model and the appropriate mlist test.

```
Device(cfg-grpc)# authen login authentication test
```

Configure the realtime event of concern subscribed to by gRPC.

```
Device(cfg-grpc)# subscr realtime enable
```

Configure the gRPC data response timeout time.

```
Device(cfg-grpc)# reply-timeout 30
```

Configure stream-based RPC for events.

```
Device(cfg-grpc)# event stream enable
```

Configure the DSCP value of IPv4 gPRC communication packets.

```
Device(cfg-grpc)# dscp 45
```

Enable the heartbeat mechanism, configure a heartbeat period and an interval, and specify a channel and subscription type, and specify a data compression format.

```
Device(cfg-grpc)# heartbeat inform interval 1 keepalive-times 1
Device(cfg-grpc)# user-server 192.168.1.1 12345
Device(cfg-grpc-us)# type 0x10000000 value 10
Device(cfg-grpc-us)# compression gzip
Device(cfg-grpc-us)# exit
```

Load the SSL/TLS certificate of the gRPC server.

```
Device(cfg-grpc)# server certificate load pem server.pem private-key server.key
```

Load the CA certificate of the gRPC server.

```
Device(cfg-grpc)# server trusted-ca load ca.pem
```

Configure IPv4 and IPv6 ACL filtering rules.

```
Device(cfg-grpc)# server access-class ipv4 ip4_acl
Device(cfg-grpc)# server access-class ipv6 ip6_acl
```

Configure the gRPC server to transmit data in GZIP format.

```
Device(cfg-grpc)# server compression gzip
```

Configure user login information of the specified server.

```
Device(cfg-grpc)# user-client 12345 test-user 192.168.1.100
```

Configure a YANG model path.

```
Device(cfg-grpc)# sensor-group 1
Device(cfg-grpc-us)# sensor-path /bridge:bridge/vlan/entry[id='1']/swport-list
Device(cfg-grpc-us)# exit-sensor-group
```

Configure a destination group.

```
Device(cfg-grpc)# destination-group 1
```

Configure an IPv4 address and port number of the destination group.

```
Device(cfg-grpc-destination)# ip 192.168.0.1 port 12345
```

Configure the destination group to report data in GZIP format.

```
Device(cfg-grpc-destination)# compression gzip
```

Enable the destination group to use the gRPC protocol that supports TLS encryption.

```
Device(cfg-grpc-destination)# protocol grpc tls
Device(cfg-grpc-destination)# exit-destination-group
```

Configure the subscription association relationship group, associated YANG model group and period interval, associated destination group, and data encoding format json-ietf.

```
Device(cfg-grpc)# subscription 1
Device(cfg-grpc-subscription)# sensor-group 1 sample-interval 1000
Device(cfg-grpc-subscription)# destination-group 1
Device(cfg-grpc-subscription)# exit-grpc-subscription
```

Configure an RPC method defined by gnmi.proto.

```
Device(cfg-grpc)# rpc gnmi enable
```

Configure an RPC method defined by openR.

```
Device(cfg-grpc)# rpc openr enable
```

Configure a private RPC method to receive and send messages.

```
Device(cfg-grpc)# dial-out rpc gnmi-ext
```

5. Verification

Run the **ping** command to check whether the device and the PC are reachable via the L3 route.

```
Device# ping oob 192.168.1.1
Sending 5, 100-byte ICMP Echoes to 192.168.1.1, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 2/3/5 ms.
```

Run the **display grpc status** command to check whether the gRPC function is enabled.

```
Device# display grpc status
GRPC:
  Uptime: 00h:28m:21s
```

Run the **display grpc client** command to display the user login information of the gRPC server.

```
Device# display grpc client
Client information all:
  1      : test-user@192.168.1.100, ID 12345
```

Run the **display grpc channel counter** command to display the statistics about the subscription server.

```
Device# display grpc channel counter
Server subscr all:
  server ip 192.168.1.1 port 50051
    state TRANSIENT_FAILURE
```

```
    heartbeat state FAILURE
  json list count: 0
  json send count: 1
  json drop count: 0
  call list count: 0
  call send count: 1
  call drop count: 0
```

Run the **display grpc channel counter** command to display the address, port number, and subscribing events of the subscription server.

```
Device# display grpc channel
Server subscr all:
  server ip 192.168.1.1 port 50051
    state TRANSIENT_FAILURE
    heartbeat state FAILURE
  subscr event:
    none
```

Run the **display grpc server information** command to display the gRPC server status, port number, TLS status, configured ACL, and certificate.

```
Device# display grpc server information
server information:
  status                  : enable
  port                    : 50051
  tls status              : enable
  ipv4 access-apply : ip4_acl
  ipv6 access-apply : ip6_acl
  certificate cn      : grpc
  ca certificate      : /O=RJ/emailAddress=1@123.com/C=CN/ST=FJ/L=FZ/OU=Ruijie/CN=ROOT
```

Run the **display grpc event subscription** command to display subscription delivered through events.

```
Device# display grpc event subscription
dynamic subscriptions:
  send count              : 405
  drop count              : 8
  call list count      : 0
  client(172.20.92.195:49340):
    subriptions:
      subription:
        json_event 0x10020001
        value 0
        value-extend 100
```

Run the **display grpc openconfig subscription** command to display the subscription relationship implemented according to openconfig-telemetry.yang.

```
Device# display grpc openconfig subscription
all subscriptions:
  subscription 1:
    sensor group 1:
      sample interval:1000
      sample timer count:858
      sensor path: /bridge:bridge/vlan/entry[id='1']/swport-list
        state: NORMAL
        recv count: 858
    destination group 1:
      destination: 192.168.1.100:12345
        state: TRANSIENT_FAILURE
        task list count: 0
        task drop count: 0
        send request count: 0
        send update count: 0
        recv response count: 0
      destination: [1::1]:12345
        state: TRANSIENT_FAILURE
        task list count: 0
        task drop count: 0
        send request count: 0
        send update count: 0
        recv response count: 0
```

6. Configuration Files

- Device configuration file

```
hostname Device
!
interface gigabitEthernet 0/1
 unset switch-port
 ip address 192.168.1.100 255.255.255.0
!
grpc
 dial-out rpc custom
 rpc gnmi enable
 rpc openr enable
 subscr enable
 reply-timeout 30
dscp 45
event stream enable
 subscr-source-interface Mgmt 0
 user-client 12345 test-user 192.168.1.100
 heartbeat inform interval 1 keepalive-times 100
```

```
server certificate load pem server.pem private-key server.key
server compression gzip
server trusted-ca load ca.pem
server access-apply ipv4 ip4_acl
server access-apply ipv6 ip6_acl
 sensor-group 1
  sensor-path /bridge:bridge/vlan/entry[id='1']/swport-list
  exit-sensor-group
 destination-group 1
  protocol grpc tls
  compression gzip
  ip 192.168.1.100 port 12345
  exit-destination-group
 subscription 1
  sensor-group 1 sample-interval 1000
  destination-group 1
  exit-grpc-subscription
 user-server 192.168.1.1 50051
  compression gzip
  exit-user-server
!
end
```

Contents

1 Configuring OpenFlow 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.1.3 Protocols and Standards 4
1.2 Restrictions and Guidelines 4
1.3 Configuration Task Summary 5
1.4 Configuring OpenFlow 5
1.4.1 Overview 5
1.4.2 Configuring OpenFlow 5
1.4.3 Configuring the OpenFlow Multi-controller Mode 5
1.4.4 Configuring the VLAN Tag for Packets Sent to the Controller 6
1.4.5 Configuring the Table-Lookup Mode 6
1.4.6 Configuring the OpenFlow Source IP Address 7
1.5 Monitoring 7
1.6 Configuration Examples 8
1.6.1 Configuring Central Control 8

1 Configuring OpenFlow

1.1 Introduction

1.1.1 Overview

OpenFlow is a network transmission protocol defined between a controller and the forwarding layer in the software defined network (SDN) architecture. OpenFlow allows the controller to directly access and operate network devices on the forwarding plane. These devices may be physical or virtualized.

OpenFlow separates the control plane from the forwarding plane of the network devices through standard communication protocols. Thus, the network devices focus on forwarding, and the control of an entire network focuses one controller. An OpenFlow connection is established between a controller and a device. The controller generates and delivers forwarding rules to the network device in a flow table through the OpenFlow connection, thereby centrally manage the control plane and reducing maintenance and management costs.

1.1.2 Principles

1. Basic Concepts

- Flow table

  A flow table is a core data structure for a device to control forwarding policies. Based on the flow table, the device determines an action to be taken for network traffic that enters the device.

  A device can have one or multiple flow tables. At present, a device supports only one flow table. A flow table is composed of multiple flow entries.

  According to the OpenFlow protocol, a flow entry consists of three parts: **header**, **counter**, and **action**.

  - **header**: It defines the index of a flow entry and consists of various packet fields such as source Media Access Control (MAC) address, destination MAC address, Ethernet protocol type, source IP address, destination IP address, IP protocol type, source port, and destination port. If a packet matches the header field of a flow entry, the packet matches this flow entry.
  - **counter**: It counts the number of packets and bytes that match this flow entry.
  - **action**: It defines an action taken for the packet that matches the flow entry, for example, discarding, broadcasting, or forwarding the packet.

- Message

  The OpenFlow protocol supports three categories of messages: **controller-to-switch**, **asynchronous**, and **symmetric**. Each category of messages include several subtypes of messages. The three categories of messages are described as follows:

  - **controller-to-switch**: This category of messages are initiated by the controller to manage and obtain network device statuses. For the subtypes of messages, see Table 1-1.
  - **asynchronous**: This category of messages are initiated by a network device to update network events or network device status changes (most commonly link up/down of a network port) to the controller. For the subtypes of messages, see Table 1-2.
  - **symmetric**: This category of messages are initiated either by a device or a controller to achieve an

initial handshake of the protocol and detect connection statuses. For the subtypes of messages, see Table 1-3.

Table 1-1 Subtypes of controller-to-switch Messages

Subtype	Description
Features	Used by a controller to send Features requests to understand device capabilities and responses.
Configuration	Used by a controller to query device configuration and make responses.
Modify-State	Used to manage device statuses, for example, flow entry and port statuses. This type of messages are used to add, delete, and modify flow entries, group entries, and port attributes of a device.
Multipart	Used by a controller to request device information, for example, current configuration and statistics.
Packet-Out	Used by a controller to send packets to a device.
Barrier	Used to check whether an action taken before a barrier message is successful. A controller sends a barrier request message. If a device confirms that the preceding action is successful, the device returns a barrier response message.
Role-Request	Used by a network device to request the role of a controller.
Asynchronous-Configuration	Used by a controller to set an asynchronous message filter and filter out unwanted asynchronous messages or query this filter.

Table 1-2 Subtypes of Asynchronous Messages

Subtype	Description
Packet-in	Used by a device to send packets to the controller.
Flow-Removed	Used to notify a controller of removing a flow entry from the flow table.
Port-Status	Used to notify a controller of the change in the device port status or configuration.
Error	Used to notify a controller of a device fault or error.

Table 1-3 Subtypes of Symmetric Messages

Subtype	Description
Hello	A message for interaction sent by between a controller and a device when an OpenFlow connection is established.

Subtype	Description
Echo	Keep-alive (echo request/reply) messages sent by a controller and a device.
Experimenter	A message reserved for new features.

2. Roles of the Controller

Table 1-4 Roles of the Controller

Role	Permission
Master	The controller in this role has all the permissions to control devices, deliver a flow table, query statistics, and receive status messages reported by the devices. In the cluster deployment mode, only one controller plays the master role.
Equal	The controller in this role has all the permissions of the controller. Compared with the master role, the equal role can be played by multiple controllers.
Standby	The controller in this role has partial permissions, and it is not permitted to deliver flow entries, group entries, or meter entries. The controller is not permitted to modify device configuration or perform packet out operation.

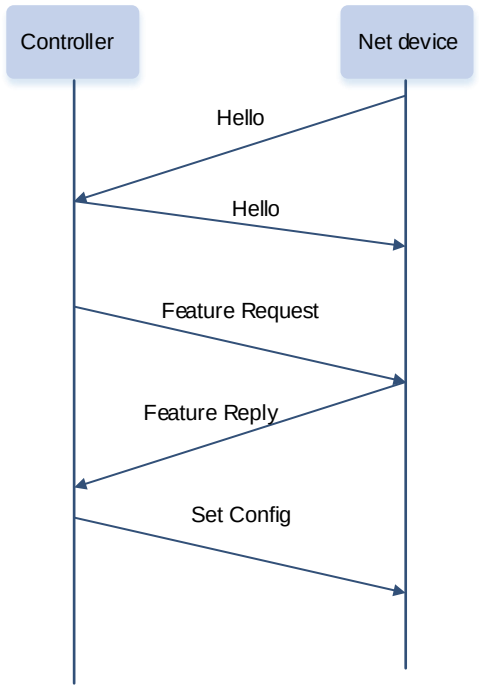
3. Separating Control from Forwarding

OpenFlow enables centralized management of the network control plane, so that the entire network can be centrally managed at ease (as compared with the status quo of the network), reducing maintenance and management costs.

The controller sends flow table information to the network devices, so as to control the method for forwarding network data packets and configuration parameters. Each network device will send a notification message to the controller when its link is interrupted or when the network device receives a data packet in which no forwarding action has been specified. In this way, the interaction between the controller and the network devices is implemented to eventually control the transmission of the entire network.

The mutual discovery between a controller and a network device must be completed before the controller and the network device interact with each other. Figure 1-1 shows the specific actions involved in this process.

Figure 1-1 Principle of the Controller and Network Device Interaction



Hello packets are sent between a controller and a network device to achieve a handshake. After the handshake is completed, the controller requests specific information about the device, including the number of ports on the network device and the capability of each port (such as the Feature Request/Reply shown in Figure 1-1). Then the controller delivers configurations (such as Set Config shown in Figure 1-1) to the specific network device. After a connection is established, the controller defines various flows and corresponding actions for the flows, and delivers them in a flow table to the network device. When a data packet enters the network device, the network device matches the data packet with the flow table according to present flow table rules and performs a corresponding action (including forwarding, discarding, and modifying the packet). At the same time, a corresponding counter is updated. If no match is found in the flow table, the network device forwards the data packet to the controller.

The network device locally maintains the flow table delivered from the controller. If the data packet to be forwarded is already defined in the flow table, the network device directly forwards the data packet. Otherwise, the data packet is sent to the controller to confirm the transmission path (which can be understood as control plane parsing to generate the flow table) and then forwarded based on the flow table delivered from the controller.

1.1.3 Protocols and Standards

- OpenFlow Switch Specification Version 1.0.0
- OpenFlow Switch Specification Version 1.3.0

1.2 Restrictions and Guidelines

- Only the Table 0 flow table is supported.
- A maximum of three controllers can be connected.

1.3 Configuration Task Summary

OpenFlow configuration includes the following tasks:

(1) Configuring OpenFlow

(2) (Optional) Configuring the OpenFlow Multi-controller Mode

(3) (Optional) Configuring the VLAN Tag for Packets Sent to the Controller

(4) (Optional) Configuring the Table-Lookup Mode

(5) (Optional) Configuring the OpenFlow Source IP Address

1.4 Configuring OpenFlow

1.4.1 Overview

As a network transmission protocol, OpenFlow runs over Transport Layer Security (TLS) or unprotected TCP connections, and defines the interaction between a controller and a network device.

1.4.2 Configuring OpenFlow

1. Overview

After the OpenFlow function is enabled on a device, packet forwarding control policies are generated and delivered by a controller.

2. Restrictions and Guidelines

- Before switching the address of a controller, disable and then enable the OpenFlow function again.
- When the OpenFlow function is enabled, the LLDP function on the device must be disabled. Otherwise, the LLDP packets delivered by a controller cannot be sent out.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the OpenFlow function.

of controller-ip *ipv4-address* [**port** *port-number*] [**aux**] **interface** *interface-type interface-number*

The OpenFlow function is disabled by default.

1.4.3 Configuring the OpenFlow Multi-controller Mode

1. Overview

After the multi-controller mode is configured, multiple controllers can be connected once.

2. Restrictions and Guidelines

Before the multi-controller mode is configured, the OpenFlow function must be disabled first. The multi-controller mode can take effect after the OpenFlow function is enabled again.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Connect an OpenFlow device to the controller.

of mode { **single** | **multiple** }

An OpenFlow device works in multi-controller mode by default.

(4) (Optional) Configure the interval at which an OpenFlow device reconnects to a controller.

of connect-interval *interval*

The default interval at which an OpenFlow device reconnects to a controller is 6 seconds.

(5) (Optional) Configure the echo request sending interval of an OpenFlow device.

of echo-interval *interval*

The default echo request sending interval of an OpenFlow device is 5 seconds.

1.4.4 Configuring the VLAN Tag for Packets Sent to the Controller

1. Overview

This function is used to configure a VLAN tag for packets sent to a controller by a device. Packets sent to a controller by a device contain the VLAN tag by default.

2. Restrictions and Guidelines

The configuration takes effect immediately.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a VLAN tag for packets sent by an OpenFlow device to a controller.

of packet vlantag

Packets sent to a controller by an OpenFlow device contain the VLAN tag by default.

1.4.5 Configuring the Table-Lookup Mode

1. Overview

This function is used to enable the table-lookup mode for received packets. A device looks up the flow table for the received packets by default, and processes the packets based on the matched entry. If no entry is matched, the packets are discarded.

2. Restrictions and Guidelines

The configuration takes effect immediately.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Connect an OpenFlow device to a controller.

of packet table-lookup { **enable** | **disable** }

The table-lookup mode is enabled for a network device by default.

1.4.6 Configuring the OpenFlow Source IP Address

1. Overview

The IP address of a device is configured as a source IP address for connecting to a controller. The default source IP address is the IP address of the connection port.

2. Restrictions and Guidelines

The configuration takes effect immediately and the network device is disconnected from the controller. Then, the source IP address is used for reconnection.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a source IP address for connecting an OpenFlow device to a controller.

of source-ip *ipv4-address*

The IP address of the connection port is the source IP address for connecting to a controller by default.

1.5 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Table 1-5 Monitoring

Command	Purpose
display of	Displays the connection between the OpenFlow device and the controller.
display of port	Displays the port status of the current OpenFlow device.
display of flowtable	Displays the flow table of the current OpenFlow device.
display of group (only available in OpenFlow 1.3)	Displays the group table of the current OpenFlow device.

Command	Purpose
display of meter (only available in OpenFlow 1.3)	Displays the meter table of the current OpenFlow device.
display of mergedflow(only available in OpenFlow 1.3)	Displays the merged flow table of the current OpenFlow device.

1.6 Configuration Examples

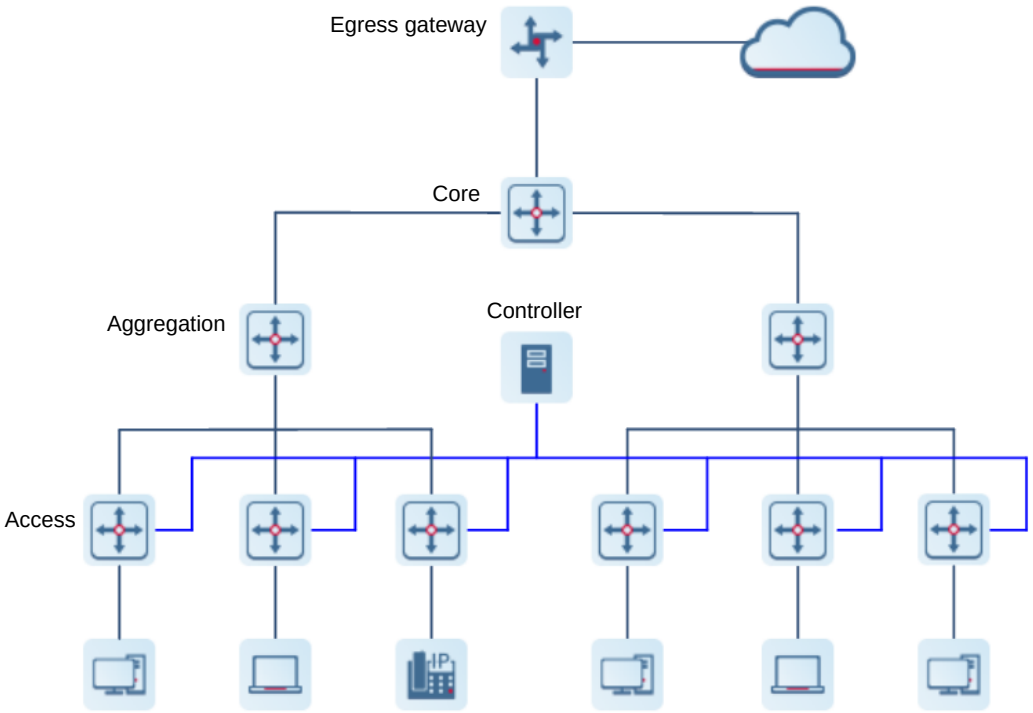
1.6.1 Configuring Central Control

1. Requirements

OpenFlow can be used to perform centralized management of access device authentication. As displayn in Figure 1-2, a controller is deployed on an access device to authenticate and control the access device, so that the authentication function (on the control plane) is shifted from the access device to the controller.

2. Topology

Figure 1-2 Topology of Central Control



3. Notes

- Enable the OpenFlow function on a device and establish an OpenFlow management channel with a controller.

- Run OpenFlow Client on the access device to interconnect with the controller.
- The controller asks the access device to send an authentication packet to itself over the OpenFlow protocol.
- The controller completes the authentication process, and sends authentication results to the access device over the OpenFlow protocol to complete admission control on end users.

4. Procedure

Enable the OpenFlow function on a device, set the IP address of a controller to 172.18.2.36/24 and the port number to 65535, and disable the LLDP function on the device.

```
Device> privilege
Device# config-system terminal
Device(cfg)# unset lldp enable
Device(cfg)# interface GigabitEthernet 0/1
Device(cfg-if-GigabitEthernet 0/1)# unset switch-port
Device(cfg-if)# ip address 172.18.2.36 255.255.255.0
Device(cfg-if)# exit
Device(cfg)# of controller-ip 172.18.2.35 port 65535 interface mgmt 0
```

Configure the single-controller mode on the device.

```
Device(cfg)# of mode single
```

Remove the VLAN tag from packets sent to the controller.

```
Device(cfg)# no of packet vlantag
```

Configure the table-lookup mode on the device.

```
Device(cfg)# ofpacket table-lookup enable
```

Set the OpenFlow source IP address for the device to 192.168.197.25.

```
Device(cfg)# of source-ip 192.168.197.25
```

5. Verification

Display the connection status of OpenFlow 1.3.

```
Device# display of
[0] Controller ID=0 Info=tcp:172.18.2.35 port=6653 interface Mgmt 0, Main is Connected, Aux is Disabled
```

Display the port status of OpenFlow 1.3.

```
Device# display of port
ID   IFX  INTERFACE            SPEED    LINK  DUPLEX   TX_PKT  RX_PKT  CONFIG
1    1    GigabitEthernet 0/1  Unknown  DOWN  Unknown  0       0       NA
2    2    GigabitEthernet 0/2  Unknown  DOWN  Unknown  0       0       NA
3    3    GigabitEthernet 0/3  Unknown  DOWN  Unknown  0       0       NA
4    4    GigabitEthernet 0/4  Unknown  DOWN  Unknown  0       0       NA
5    5    GigabitEthernet 0/5  Unknown  DOWN  Unknown  0       0       NA
6    6    GigabitEthernet 0/6  Unknown  DOWN  Unknown  0       0       NA
7    7    GigabitEthernet 0/7  Unknown  DOWN  Unknown  0       0       NA
8    8    GigabitEthernet 0/8  Unknown  DOWN  Unknown  0       0       NA
9    9    GigabitEthernet 0/9  Unknown  DOWN  Unknown  0       0       NA
10   10   GigabitEthernet 0/10 Unknown  DOWN  Unknown  0       0       NA
```

11	11	GigabitEthernet 0/11 Unknown	DOWN	Unknown	0	0	NA
12	12	GigabitEthernet 0/12 Unknown	DOWN	Unknown	0	0	NA
13	13	GigabitEthernet 0/13 Unknown	DOWN	Unknown	0	0	NA
14	14	GigabitEthernet 0/14 Unknown	DOWN	Unknown	0	0	NA
15	15	GigabitEthernet 0/15 Unknown	DOWN	Unknown	0	0	NA
16	16	GigabitEthernet 0/16 Unknown	DOWN	Unknown	0	0	NA

Display the flow table status of OpenFlow 1.3.

```
Device# display of flowtable
/****************** openflow flow table[0]---flow number[2] ******************/

{table="0", duration_sec="5222", priority="0", flags ="0x1",idle_timeout="0", hard_timeout="0", cookie="0x0",
packet_count="0", byte_count="0". match=oxm{all match} instructions=[apply{acts=[output{port="controller",
max_len="3333"}] }] }
xid=472, sync_flag=0x0, ss_index=0x7b6

{table="0", duration_sec="5222", priority="66", flags ="0x1",idle_timeout="0", hard_timeout="0", cookie="0x0",
packet_count="0", byte_count="0". match=oxm{eth_dst="00:25:64:c4:54:e4", eth_type="0x800",
ipv4_dst="1.1.0.0", ipv4_dst_mask="255.255.240.0"} instructions=[apply{acts=[output{port="controller",
max_len="65535"}] }] }
xid=473, sync_flag=0x0, ss_index=0x7b1

/****************************** openflow flow table end ***************************************/
flow total number = 2
```

Configure the single-controller mode on the device. The device can be connected to only one controller. If more controllers are connected, a notification will be displayed.

```
Device(cfg)# unset of controller-ip
Device(cfg)# of mode single
Device(cfg)# of controller-ip 172.18.122.24 interface gigabitEthernet 0/1
Device(cfg)# of controller-ip 172.18.122.25 interface gigabitEthernet 0/1
Controller Mode is Single, can't connected
```

Run the **display of** command to check whether the current mode is the table-lookup mode and whether the source IP address is 192.168.197.25.

```
Device# display of
version:openflow1.3, controller[0]:tcp:172.18.105.11 port 6653 interface GigabitEthernet 1/0/7, main is
connected, aux is disable, role is master.
Current controller mode : multiple.
Current packet process mode : Lookup all flow.
Datapath id = 897516188948
Source IP = 192.168.197.25
```

6. Configuration Files

```
hostname Device
unset lldp enable
!
of source-ip 192.168.197.25
```

```
of controller-ip 172.18.2.35 port 65535 interface Mgmt 0
unset of packet vlantag
!
interface GigabitEthernet 0/1
unset switch-port
 ip address 172.18.2.36 255.255.255.0
```

7. Common Errors

- The controller IP address is incorrectly configured.
- The TCP port of the controller is incorrectly configured.
- The IP address of the local management channel is not configured.

Contents

1 Configuring SEM 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Basic Concepts 1

1.1.3 Principle 3

1.2 Configuration Task Summary 5

1.3 Enabling the SEM Function 5

1.3.1 Overview 5

1.3.2 Restrictions and Guidelines 6

1.3.3 Configuring SEM Policies 6

1.3.4 Configuring Parameters of SEM Policies 8

1.4 Monitoring 10

1.5 Configuration Examples 11

1.5.1 Configuring an Interface Monitoring Event 11

1.5.2 Configuring a Counter Monitoring Event 13

1.5.3 Configuring a Timer Monitoring Event 17

1.5.4 Configuring a Syslog Monitoring Event 19

1.5.5 Configuring a Track Monitoring Event 21

1.5.6 Configuring a Route Monitoring Event 24

1.5.7 Configuring a CLI Monitoring Event 25

1 Configuring SEM

1.1 Introduction

1.1.1 Overview

The smart embedded manager (SEM) is a network management tool. The tool can be separately deployed through commands, independent from external network management tools, which makes the tool easy to deploy.

Traditional external network management tools must be connected to devices over a network to manage the devices. If a network fails, affecting network connection, the external network management tools lose effect. SEM can directly manage devices under any circumstance and troubleshoot different types of network or device faults or capture key information.

SEM detects configuration events in real time. When an event occurs, SEM takes preset actions. The entire procedure is highly customizable. SEM provides fault detection and processing and automatic management functions, which improves the availability of the device and network.

SEM supports a variety of events, including key events of a device such as key syslog, trap and time point, or input operations such as Command Line Interface (CLI) commands and Simple Network Management Protocol (SNMP) operations, or threshold excess such as interface statistics, SNMP object values, and system resource statistics. SEM supports many actions, including user command running, log sending, and active device reset.

1.1.2 Basic Concepts

- Event

  An event indicates an event that a user cares about. It is configured by users and attached to a policy. One or multiple events can be configured. Each type of events are detected by a specific detector. When conditions are met, corresponding event is triggered by the event detector. For example, to detect the inclusion of the **shutdown** command in the privileged EXEC mode, run the **event tag** *example* **cli pattern** *shutdown* **mode** *exec* command to configure this event to the corresponding command line event detector.

- Action

  An action indicates a measure taken after an event occurs. It is configured by users and attached to a policy. One or multiple actions can be configured. Each type of actions correspond to a type of action configurations. When an event occurs, the policy corresponding to this event runs configured actions in order. For example, to restart a device after an event occurs, run the **act** *example* **reload** command.

- Policy

  A policy organizes relationships between events and between events and actions. A policy must be submitted after it is configured. When an event corresponding to this policy meets a preset rule, this policy is triggered and runs actions in order.

- Event detector

  An event detector is embedded in specific business and monitors the business based on user configuration. When the business complies with the user configuration, corresponding event is triggered. The event detector sends an event notification to the smart management server.

- Smart management server

  A smart management server receives the event notification from the event detector and determines whether to trigger a policy based on the event. After the policy is triggered, the smart management server runs the policy in policy management.

- Policy manager

  A policy manager manages SEM policies and runs actions in the policies that are triggered by the smart management server.

- SEM environmental variable

  It indicates a variable that is used by an action of a policy during SEM running. There are three types of SEM environmental variables:

  - Global variable
  - Local system variable
  - Local user variable

  The words and signs located between the United States Dollar sign ($) and the first non-letter, non-digit, and non-underline character are used as the name of a variable to be replaced. Global variables can be used in all policies. Local system variables and local user variables are local variables and can be used in specific policies only. The local system variables are read-only. The local user variables are defined by the actions in the policies during policy running. Therefore, the local system variables start with the underline (_) to avoid confusion with the local user variables.

- Specific application event of SEM

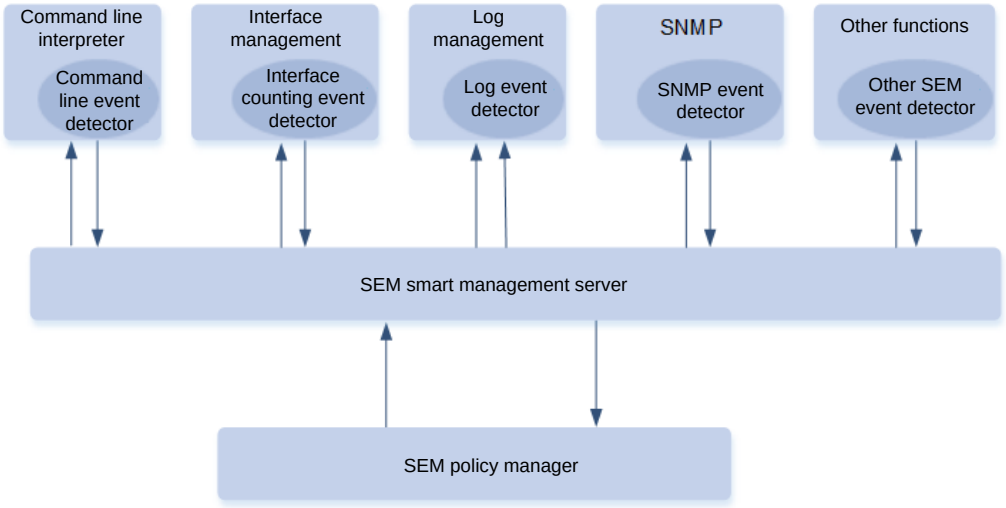
  SEM supports specific application events in the SEM system. These events are used to trigger a policy during running of another policy. Subsets and types are used to distinguish specific application events. Specific application events of SEM are detected by the specific application event detector based on subsets and types, and are released by their actions in the running policy. When an action in the running policy releases a specific application event of a subset and type, the detected event of the subset and type is triggered.

- SEM naming counter

  SEM supports a naming counter in the SEM system. Change in the SEM naming counter is detected by the counting event detector of SEM. The value of the SEM naming counter is operated by the counter action in the running policy. The SEM naming counter can be used for policy trigger counting and numerical statistics.

1.1.3 Principle

Figure 1-1 SEM Structure



As shown in Figure 1-1, SEM provides different types of event detectors. These events detectors are embedded in different business to monitor the business in real time. The detectors try to match event triggering conditions in user configurations with business running events. If a match is hit, the event detector notifies the smart management server of the event.

On the basis of the event configuration in the policy, the smart management server determines whether to trigger a policy to run. If the event triggering conditions are met, the smart management server runs the policy in the policy manager.

The policy in the policy manager runs preset actions in order based on user configuration.

1. Event Detectors Supported by SEM

SEM supports a variety of event detectors. These event detectors are embedded in different business. They detect the running business, and decide whether any event occurs. SEM supports the following detector types:

- Counting event detector

  The counting event detector detects the naming counter in SEM. When the naming counter exceeds the specified threshold, a counter event is triggered. After the counter event is triggered, the counting event detector stops working temporarily, and resumes until the counter reaches the threshold. The value of the naming counter is changed by the action of the running counter and can thus be accumulated by other policies. When the counter reaches a threshold, a counter event is triggered, which triggers the policy to run. During the running of this policy, the value of the counter is reset to realize circulation.

- Interface counting event detector

  The interface counting event detector detects interface statistics of a device. By periodic collection and statistics, the detector determines whether the statistics of an interface exceeds a threshold. If yes, an interface counting event is triggered. After the interface counting event is triggered, the interface counting event detector stops working temporarily and resumes until the interface counter reaches the recovery threshold or the suspension time exceeds the recovery cycle.

- Syslog event detector

  The syslog event detector detects log information of a device and uses regular expressions to match the log information. If the log information is matched using the regular expressions, a syslog event is triggered.

- Timer event detector

  The timer event detector detects time-related events. There are four types of timer events as follows:

  - Absolute time timer event: The timer time is set to some time point in the future. When this time point is up, the timer event is triggered.
  - Countdown timer event: The timer time is the number of seconds before which a policy takes effect. When the effective time reaches the number of such seconds, the timer event is triggered.
  - Watchdog timer event: The timer time is the number of seconds set before which a policy takes effect. When the effective time reaches the number of such seconds, the timer event is triggered.
  - Cron timer event: Cron (derived from Greek word chronos, meaning time) is used widely in a Unix-like operating system and used to set a periodically run instruction. SEM supports time detection in cron mode. When the time point described by the cron character string of the cron timer event is up, the timer event is triggered.

- **CLI event detector**

  The CLI event detector detects the CLI input. When the input command matches the detector configuration, the CLI event is triggered. Only the **trigger** command event in the privileged EXCE mode can be detected currently.

- Track event detector

  The track event detector detects tracks. Track refers to status change of a reliable network service (RNS). The track event detector detects change in the status of a tracked entity. When the status of a tracked entity changes, the track event is triggered.

- **Route event detector**

  The route event detector detects the route-unreachable events. When a route is detected unreachable, the route event is triggered.

2. Policy Actions Supported by SEM

SEM supports different actions. An action runs after a policy is triggered by an event. The actions include information collection and device/network troubleshooting. SEM supports the following actions:

- Run CLI commands: Run commands set by users.
- Send syslog: Send specified log messages.
- Operate naming counter: Operate the naming counter in SEM.
- Reset device: Reset devices.
- Wait during policy running: Wait for some time during policy running.
- Configure environmental variables: Configure local policy variables.

3. Environmental Variables Supported by SEM

SEM can use environmental variables in policies. SEM supports the following environmental variables:

- Global variable

Global variables can be used in all policies after they are defined by using the **smart manager env** command. The values of global variables can be read-only in a policy.

- Local variable

  Local variables generated during policy running are divided into local system variables and local user variables. The local system variables are created by detectors and used to describe events. Therefore, they are read-only. Local user variables are created by actions in policies, and can be modified during policy running.

  The priority of local variables in a policy are higher than that of global variables. Reading an inexistent local variable or global variable or modifying a read-only local variable may generate an error. Configuring an inexistent local variable will cause the local variable to automatically generate.

4. Management of the SEM Smart Management Server

The SEM smart management server provides management interfaces to allow users to view SEM running information and perform management. Users can view different types of SEM information, including:

- Supported detector types
- Versions of the smart management server and event detectors
- User configuration and submitted policies
- Policies being run
- Policies to be run
- Event history
- Currently defined counter
- Currently defined SEM global variables

During running of the SEM smart management server, users can manage the running policies and perform the following operations:

- Suspend and restore entire SEM policy scheduling.
- Pause and restore specific policies or specified classes of policies.
- Adjust priority of specified types of policies.
- Forcibly terminate specific policies, specified classes of policies, or all policies.

1.2 Configuration Task Summary

SEM configuration includes the following tasks:

(1) Configuring SEM Policies

(2) (Optional) Configuring Parameters of SEM Policies

1.3 Enabling the SEM Function

1.3.1 Overview

SEM is separately deployed in a device independent from external network management tools. SEM allows customers to customize different types of events and policies so that customers can freely obtain or monitor their desired event information and actions taken for the events.

1.3.2 Restrictions and Guidelines

- The syslog function must be enabled by using the **logging** command before syslog actions are configured.
- The device must be equipped with a normal standby engine before actions are switched to the standby engine.
- Run the **act reload** command with caution. It may cause repeated restart of the system or repeated active/standby switchover.

1.3.3 Configuring SEM Policies

1. Overview

By configuring SEM policies, users can monitor events based on actual needs.

2. Restrictions and Guidelines

- When multiple events are configured for a policy, the events are automatically arranged in the alphabetical order of tags. You are advised to use the naming method based on a sequencing rule, for example, 01_cli, 02_timer, and 03_counter.
- The relationship between events refers to the relationship of the current event with the combination of all other events. Only when the first event is triggered, SEM checks whether to trigger a policy. If users try to configure juxtaposition for other events except the first event, the default juxtaposition is an AND relation. In this case, the juxtaposition of the first event is ignored.
- In either of the following scenarios, the **rollback** command can be used to roll back the policy configuration.
  - A new configured policy has not been submitted.
  - A configured policy has been submitted and registered, but changes to the policy have not been submitted.
- If no action is configured for a submitted policy, the policy can be registered. However, no action is taken when the policy is triggered.
- Generally, multiple actions are configured for a policy. When the policy is triggered, the actions are run in the alphabetical order of the *label* parameter.
- To save the I/O of CLI commands, run the **policy-info record** command.
  The **smart manager policy-info record** command does not save the command I/O as well.
- If the policy is being edited but is submitted, run the **list-config** command to display the configuration not submitted in the currently edited policy. The **commit** command is not included in the displayed content. If the policy is submitted, the configuration of the submitted policy is displayed. The **commit** command is included in the displayed content.
- The submittable policies are divided into new policies and edited and registered policies. When a registered but unedited policy is not submitted, a notification will be displayed. If users want to quit the changes to the policy configuration before submitting the changes, run the **rollback** command to roll back the policy configuration.
- During submission, the validity of the policy configuration is checked. If the checking fails, the policy configuration fails to be submitted. In this case, the policy is not registered and remains in the editing status. For example, if no event is configured for the policy, the policy submission checking fails.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an SEM policy and enter the SEM configuration mode.

smart manager app *app-name* [**class** *class-options*]

No policy is configured by default.

(4) Configure events. Configure at least one of the following tasks.

- Configure a counter monitoring event.

  event tag *event-name* [**correlate** { **andnot** | **and** | **or** }] **counter name** *counter-name* **entry-operation** *operator* **entry-value** *entry-value* **exit-operation** *operator* **exit-value** *exit-value*

  No counter monitoring event is configured by default.

- Configure an interface monitoring event.

  event tag *event-name* [**correlate** { **andnot** | **and** | **or** }] **interface name** *interface-type interface-number* **parameter** { **link_up** | **link_down** }

  No interface monitoring event is configured by default.

- Configure a log monitoring event.

  event tag *event-name* [**correlate** { **andnot** | **and** | **or** }] **syslog pattern** *regular-expression* [**priority-level** *priority-level*] [**occurs** *num-occurrences*] [**period** *period-value*] [**skip** { **yes** | **unset** }]

  No log monitoring event is configured by default.

- Configure a timer monitoring event.

  event tag *event-name* [**correlate** { **andnot** | **and** | **or** }] **timer** { **absolute** { **unix** *time-value* | **date** date-value } | **countdown time** *countdown-timer* | **cron cron-entry** *cron-entry* | **watchdog time** *watchdog-timer* } [**name** *timer-name*]

  No timer monitoring event is configured by default.

- Configure a track monitoring event.

  event tag *event-name* [**correlate** { **andnot** | **and** | **or** }] **track** [**state** { **up** | **down** }] [*track-id*]

  No track monitoring event is configured by default.

- Configure a route monitoring event.

  event tag *event-name* [**correlate** { **andnot** | **and** | **or** }] **route** { **unreachable** | **reachable** } { **ipv4** *ipv4-address* | **ipv6** *ipv6-address* } [**vrf** *vrf-name*]

A complete policy is composed of events and actions. If no event is configured for the policy, an error occurs during policy submission, resulting in submission failure.

(5) Configure actions. Configure one of the following tasks.

- Run CLI commands of a device.

  act *action-label* **cli command** *cli-string* [**pattern** *pattern-string*]

- Operate SEM counters.

act *action-label* **counter name** *counter-name* **op** { **dec** | **inc** | **nop** | **set** } **value** *counter-value*

- Terminate a policy script and set the exiting status.

  act *action-label* **exit** [*result*]

- Restart the device.

  act *action-label* **reload**

- Set local variables.

  act *action- label* **set** *variable-name variable-value*

- Carry out logging.

  act *action-label* **syslog** [**priority-level** *priority-level*] **msg** *syslog-message* [**facility** *mnemonics*]

- Pause a policy script.

  act *action-label* **wait** *wait-time*

- Configure a CLI monitoring event.

  event tag *event-name* [**correlate** { **andnot** | **and** | **or** }] **cli pattern** *regular-expression* [**occurs** *number-occurrences*] [**period** *period-value*]

No action is configured by default.

(6) Enable the function of recording CLI action output.

policy-info record

The recording function is not enabled for CLI command action output by default.

(7) (Optional) Configure description.

description *string*

No description is configured for an SEM policy by default.

(8) (Optional) Display current policy configuration.

list-config

(9) (Optional) Roll back configuration.

rollback

(10) Submit configuration.

commit

1.3.4 Configuring Parameters of SEM Policies

1. Restrictions and Guidelines

- Modification to policy description takes effect immediately without submission.
- A CLI policy can be triggered only after the **trigger** command is configured and the CLI detection event is submitted.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Clear CLI records generated during SEM policy running.

smart manager policy record clean { **all** | **dir** *record-directory* | **no-registed** | **policy-info** *registed-policy-name* }

CLI records generated during SEM policy running are not cleared by default.

(3) Trigger a CLI detection event.

trigger *cmd*

(4) Enter the global configuration mode.

config-system terminal

(5) Configure a thread pool and thread pool size of an SEM policy class.

smart manager schdulr thread-pool class *class-options* **number** *thread-number*

The thread pool size of the default SEM policy is **32** by default, and no thread pool is specified for other classes of SEM policies.

(6) Configure upper limits of SEM parameters. Configure at least one of the following tasks.

- Configure the upper limits of SEM detector parameters.

  smart manager finder { **counter** | **interface** | **timer** } **event-number** *detector-number*

  By default, the maximum number of events for the counting detector, the maximum number of events for the interface detector, and the maximum number of events for the timer detector are **256**, respectively.

- Configure the maximum number of SEM global variables.

  smart manager global-variant number *global-variant-number*

  By default, the maximum number of global variables is **512**.

- Configure the upper limits of SEM policy parameters.

  smart manager policy-info { **act-number** *policy-number* | **config-number** *policy-number* | **event-number** *policy-number* | **register-number** *policy-number* | **trigger-delay-number** *policy-number* }

  By default, the maximum number of actions in a policy is **64**; the maximum numbers of policy detectors and registered policies are **128**, respectively; and the maximum numbers of configured policies and policy delayed triggers are **256**, respectively.

- Configure the upper limits of SEM configuration instance parameters.

  smart manager record { **size-of-instance** *record-number* | **size-of-policy** *record-number* }

  By default, the maximum number of policy instances is **50** and the maximum size of a policy file is **1024** KB.

- Configure the upper limits of SEM scheduler parameters.

  smart manager schdulr { **pending-number** *schdulr-number* | **running-number** *schdulr-number* }

  By default, the maximum number of wait policies of the scheduler and the maximum number of policies run by the scheduler are **128**, respectively.

(7) Configure the maximum number of saved SEM history messages.

smart manager hist-info size events *size*

By default, the maximum number of saved SEM history messages is **50**.

(8) Configure the application delay time of an SEM policy.

smart manager policy-info startup-delay *delay time*

The application delay time of an SEM policy is **60** seconds.

(9) Enter the SEM configuration mode.

smart manager app *app-name* [**class** *class-options*]

(10) Configure the trigger attributes of a policy.

trigger { **correlate-period** *correlate-period-value* | **correlate-start** *period-start-value* | **delay** *delay-value* | **maxrun** *maxruntime-number* | **occurs-period** *occurs-period-value* | **occurs-value** *occurs-value* }*

No trigger attribute is configured for a policy by default.

1.4 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Table 1-1 SEM Monitoring

Command	Purpose
display smart manager finder [**all** | *detector-name*] [**detailed** | **statistics**]	Displays SEM detector information.
display smart manager environment [**all** | *variable-name*]	Displays global variables.
display smart manager hist-info events	Displays SEM history information.
display smart manager policy-info all	Displays all policies and policy submission information.
display smart manager policy-info active [**class** *class-options*] [**detailed**]	Displays active policy instances.
display smart manager policy-info pending [**class** *class-options*] [**detailed**]	Displays policies to be run.
display smart manager policy-info registered [**policy-info** *policy-name*] [**incident-type** *event-name*] [**class** *class-options*] [**time-ordered** | **name-ordered**]	Displays information of registered policies.
display smart manager schdulr thread-pool [**detailed**]	Displays the running status of a SEM scheduler by thread pool.
display smart manager version	Display SEM versions.

1.5 Configuration Examples

1.5.1 Configuring an Interface Monitoring Event

1. Requirements

Device A is directly connected to Device B through GigabitEthernet 0/1. Due to the problem of the line or Device B, Device A often receives many consecutive error frames, which affects communication. After GigabitEthernet 0/1 of Device A is shut down and restarted, the problem is solved.

2. Topology

Figure 1-2 Configuring an Interface Monitoring Event



3. Notes

(1) Create a policy.

(2) Configure an interface monitoring event.

(3) Configure CLI command actions.

(4) Submit the policy configuration.

4. Procedure

(1) Create a policy with the name policy_interface on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA (cfg)# smart manager app policy_interface
```

(2) Configure an event with the name event_1 in the policy_interface policy and set the type to interface. The specific parameter configuration is as follows:

Configure the interface counting type as link_down and set the interface status to down.

Configure the interface name as GigabitEthernet 0/1 to specify the interface to be detected.

```
DeviceA (sem-app)# event tag event_1 interface name GigabitEthernet 0/1 parameter link_down
```

(3) Configure actions of the policy_interface policy.

Configure action_1 to run the **enable** command to enter the privileged EXEC mode.

```
DeviceA (sem-app)#act action_1 cli command “privilege”
```

Configure action_2 to run the **config-system terminal** command to enter the global configuration mode.

```
DeviceA (sem-app)#act action_2 cli command "config-system terminal"
```

Configure action_3 to run the **interface GigabitEthernet 0/1** command to enter the interface configuration mode.

```
DeviceA (sem-app)#act action_3 cli command "interface GigabitEthernet 0/1"
```

Configure action_4 to run the **unset shutdown** command to set the interface status to up.

```
DeviceA (sem-app)#act action_4 cli command "unset shutdown"
```

(4) Submit the policy.

```
DeviceA(sem-app)# commit
```

(5) Exit the policy editing status.

```
DeviceA(sem-app)# exit
```

5. Verification

Run the **display smart manager policy-info registered** command to display a registered SEM policy.

```
DeviceA# display smart manager policy-info registered
 No.   Name                                                    Type      Event Type
   1   policy_interface                                        app   interface
  event_1: interface: parameter link_down name GigabitEthernet 0/1
  act action_1 cli command “privilege”
  act action_2 cli command "config-system terminal"
  act action_3 cli command "interface GigabitEthernet 0/1"
  act action_4 cli command "unset shutdown"
```

Run the **display smart manager policy-info active** command to display an active policy.

```
DeviceA# display smart manager policy-info active
Key: P - Priority              :L - Low, H - High, N - Normal
       S - Scheduling node :A - Active, P - Pending

No.   Job id    P S Status       Time Of Event          Event Type      Policy Name
1     2          N A suspend     2021-03-19 18:05:59    interface       policy_interface
```

Set GigabitEthernet 0/1 to the down status to verify whether the policy takes effect.

```
DeviceA# config-system terminal
DeviceA (cfg)# interfce GigabitEthernet 0/1
DeviceA (cfg-if-GigabitEthernet 0/1)#shutdown
*Mar 19 18:06:19: %SEM-6-PLCY_TRG: Policy policy_1 id 2 is triggered by event 2, and the policy-info instance id is 1.
```

Run the **display smart manager hist-info events detailed** command to display SEM event history.

```
DeviceA# display smart manager hist-info events detailed
No.         Job id    Event Type          Time                        Policy name
1           2          interface          2021-03-22 17:19:38         policy_interface
            Class: default, Policy Type: app
```

6. Configuration Files

Device configuration file

```
hostname DeviceA
!
smart manager app policy_interface
 event tag event_1 interface name GigabitEthernet 0/1 parameter link_down
 act action_1 cli command “privilege”
```

```
 act action_2 cli command "config-system terminal"
 act action_3 cli command "interface GigabitEthernet 0/1"
 act action_4 cli command "unset shutdown"
!
end
```

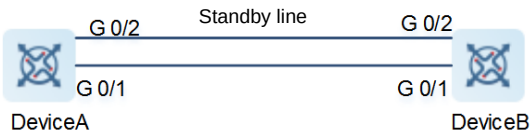
1.5.2 Configuring a Counter Monitoring Event

1. Requirements

Device A is directly connected to Device B through GigabitEthernet 0/1. A low-bandwidth standby line is created between Device A and Device B through GigabitEthernet 0/2. Due to the problem of the line or Device B, Device A often receives many consecutive error frames on GigabitEthernet 0/1, which affects communication. After GigabitEthernet 0/1 of Device A is shut down and restarted, the problem is solved. If the counter monitoring event is triggered more than 50 times, GigabitEthernet 0/1 is shut down permanently and the standby line is used.

2. Topology

Figure 1-3 Configuring a Counter Monitoring Event



3. Notes

(1) Configure an interface monitoring event, including interface counting event, CLI actions, and counter actions.

(2) Configure a counter monitoring event for detecting interface monitoring events.

4. Procedure

(1) Create a policy with the name policy_interface on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# smart manager app policy_interface
```

(2) Configure an event with the name event_1 in the policy_interface policy and set the type to interface. The specific parameter configuration is as follows:

Configure the interface counting type as link_down and set the interface status to down.

Configure the interface name as GigabitEthernet 0/1 to specify the interface to be detected.

```
DeviceA (sem-app)# event tag event_1 interface name GigabitEthernet 0/1 parameter link_down
```

(3) Configure actions of the policy_interface policy.

Configure action_1 to run the **enable** command to enter the privileged EXEC mode.

```
DeviceA (sem-app)#act action_1 cli command “privilege”
```

Configure action_2 to run the **config-system terminal** command to enter the global configuration mode.

```
DeviceA (sem-app)#act action_2 cli command "config-system terminal"
```

Configure action_3 to run the **interface GigabitEthernet 0/1** command to enter the interface configuration mode.

```
DeviceA (sem-app)#act action_3 cli command "interface GigabitEthernet 0/1"
```

Configure action_4 to run the **unset shutdown** command to set the interface status to up.

```
DeviceA (sem-app)#act action_4 cli command "unset shutdown"
```

Configure action_5, and increase the value of the naming counter for the SEM policy name of policy_interface by 1.

```
DeviceA(sem-app)# act action_5 counter name policy_interface op inc value 1
```

(4) Submit the policy.

```
DeviceA(sem-app)# commit
```

(5) Exit the policy editing status.

```
DeviceA(sem-app)# exit
```

(6) Create a policy with the name policy_counter.

```
DeviceA(cfg)# smart manager app policy_counter
```

(7) Configure an event with the name event_counter in the policy_counter policy and set the event type to SEM counter. The specific parameter configuration is as follows:

Set the counter name to policy_interface, which is the policy name in step (1) .

Set the counter trigger parameter **entry-operation ge entry-value** to **6**, indicating that the event is triggered when the value of the SEM counter is equal to or greater than 6.

Set the counter restoration parameter **exit-operation gt exit-value** to **5**, indicating that event monitoring is restored immediately when the value of the SEM counter is greater than 5.

```
DeviceA(sem-app)# event tag event_counter counter name policy_interface entry-operation ge entry-value 6 exit-operation gt exit-value 5
```

(8) Configure actions of the policy_counter policy.

Configure action_counter_1 to set the value of the counter policy_interface to 0.

```
DeviceA(sem-app)# act action_counter_1 counter name policy_interface op set value 0
```

Configure action_counter_2 to run the **enable** command to enter the privileged EXEC mode.

```
DeviceA(sem-app)# act action_counter_2 cli command “privilege”
```

Configure action_counter_3 to run the **config-system terminal** command to enter the global configuration mode.

```
DeviceA(sem-app)# act action_counter_3 cli command "config-system terminal"
```

Configure action_counter_4 to run the **interface GigabitEthernet 0/1** command to enter the interface configuration mode.

```
DeviceA(sem-app)# act action_counter_4 cli command "interface GigabitEthernet 0/1"
```

Configure action_counter_5 to run the **shutdown** command to set the interface status to down permanently.

```
DeviceA(sem-app)# act action_counter_5 cli command "shutdown"
```

(9) Submit the policy.

```
DeviceA(sem-app)# commit
```

(10) Exit the policy editing status.

```
DeviceA(sem-app)# exit
```

5. Verification

Run the **display smart manager policy-info registered** command to display a registered SEM policy.

```
DeviceA# display smart manager policy-info registered
No.    Name                                                                    Type       Event Type
  1    policy_interface                                                        app    interface
 event_1: interface: parameter link_down name GigabitEthernet 0/1
 act action_1 cli command “privilege”
 act action_2 cli command "config-system terminal"
 act action_3 cli command "interface GigabitEthernet 0/1"
 act action_4 cli command "unset shutdown"
   policy-info record per-instance 50 per-policy 1000

  2    policy_counter                                                          app    counter
 event_counter: counter: name policy_interface entry-operation ge entry-value 6 exit-operation gt exit-value 5
 act 1 counter name policy_interface op set value 0
 act action_counter_1 cli command “privilege”
 act action_counter_2 cli command "config-system terminal"
 act action_counter_3 cli command "interface GigabitEthernet 0/1"
 act action_counter_4 cli command "shutdown"
 act action_counter_5 cli command "end"
```

Set GigabitEthernet 0/1 to the down status five times. Run the **display smart manager finder** command to display the monitoring event information.

```
DeviceA# display smart manager finder counter statistics
finder counter events number: 1
    event id: 3, event name: event_counter
      counter name: policy_interface
      counter value: 5
      status: enable
      trigger times: 0
      trigger policy-info times: 0
```

Set GigabitEthernet 0/1 to down status again to trigger the policy_counter policy, and run the **display smart manager finder** command to display the monitoring event information.

```
DeviceA# display smart manager finder counter statistics
finder counter events number: 1
    event id: 3, event name: event_counter
      counter name: policy_interface
      counter value: 1
      status: enable
      trigger times: 1
      trigger policy-info times: 1
```

Run the **display smart manager hist-info events detailed** command to display SEM event history.

```
DeviceA# display smart manager hist-info events detailed
No.       Job id    Event Type        Time                        Policy name
1         65        interface         2021-03-24 11:27:30         policy_interface
          Class: reset, Policy Type: app
2         54        counter           2021-03-24 11:27:29         policy_counter
          Class: reset, Policy Type: app
3         64        interface         2021-03-24 11:27:29         policy_interface
          Class: reset, Policy Type: app
4         63        interface         2021-03-24 11:24:43         policy_interface
          Class: reset, Policy Type: app
5         62        interface         2021-03-24 11:24:41         policy_interface
          Class: reset, Policy Type: app
6         61        interface         2021-03-24 11:24:29         policy_interface
          Class: reset, Policy Type: app
7         60        interface         2021-03-24 11:24:25         policy_interface
          Class: reset, Policy Type: app
8         59        interface         2021-03-24 11:24:07         policy_interface
          Class: reset, Policy Type: app
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
smart manager app policy_interface
 event tag event_1 interface name GigabitEthernet 0/1 parameter link_down
 act action_1 cli command “privilege”
 act action_2 cli command "config-system terminal"
 act action_3 cli command "interface GigabitEthernet 0/1"
 act action_4 cli command "unset shutdown"
 act action_5 counter name policy_interface op inc value 1
!
smart manager app policy_counter
 event tag event_counter counter name policy_interface entry-operation ge entry-value 6 exit-operation gt
exit-value 5
 act 1 counter name policy_interface op set value 0
 act action_counter_1 cli command “privilege”
 act action_counter_2 cli command "config-system terminal"
 act action_counter_3 cli command "interface GigabitEthernet 0/1"
 act action_counter_4 cli command "shutdown"
 act action_counter_5 cli command "end"
 commit
!
end
```

1.5.3 Configuring a Timer Monitoring Event

1. Requirements

Device A is connected to a Trivial File Transfer Protocol (TFTP) server, and Device A archives log files to the TFTP server at 00:00:00 each day and deletes the original log files.

2. Topology

Figure 1-4 Configuring a Timer Monitoring Event



3. Notes

(1) Configure that Device A and the TFTP server are mutually reachable via routes.

(2) Create a policy.

(3) Configure a timer monitoring event.

(4) Configure actions.

(5) Submit the policy configuration.

4. Procedure

(1) Configure that Device A and the TFTP server are mutually reachable via routes.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1) ip address 172.16.0.1 255.255.255.0
DeviceA(cfg-if-GigabitEthernet 0/1) exit
```

(2) Create a policy with the name policy_timer on Device A.

```
DeviceA# config-system terminal
DeviceA(cfg)# smart manager app policy_timer
```

(3) Configure an event with the name event_timer in the policy_timer policy, set the type to cron timer, and specify the logging time to 00:00:00 each day.

```
DeviceA(sem-app)# event tag event_timer timer cron cron-entry "0 0 * * *"
```

(4) Configure actions of the policy_timer policy.

Configure action_timer1 to run the **enable** command to enter the privileged EXEC mode.

```
DeviceA(sem-app)# act action_timer1 cli command “privilege”
```

Configure action_timer2 to run the **copy** command to archive the log file named syslog_rfc5424_flag.txt to the server with the IP address 172.16.0.2 through TFTP.

```
DeviceA(sem-app)# act action_timer2 cli command "copy flash:syslog_rfc5424_flag.txt
tftp://172.16.0.2/syslog_rfc5424_flag.txt"
```

(5) Submit the policy.

```
DeviceA(sem-app)# commit
```

(6) Exit the policy editing status.

```
DeviceA(sem-app)# exit
```

5. Verification

Run the **ping** command to verify that Device A and the TFTP server are mutually reachable via routes.

```
DeviceA# ping 172.16.0.2
Sending 5, 100-byte ICMP Echoes to 172.16.0.2, timeout is 2 seconds:
  < press Ctrl+C to break >
!!!!!
Success rate is 100 percent (5/5), round-trip min/avg/max = 1/1/3 ms.
```

Run the **display smart manager policy-info registered** command to display a registered SEM policy.

```
DeviceA# display smart manager policy-info registered
 No.   Name                                                    Type      Event Type
   2   policy_timer                                          app   timer
  event_timer: timer: cron cron-entry 0 0 * * *
  act aciton_timer1 cli command "enble"
  act action_timer2 cli command "copy flash:syslog_rfc5424_flag.txt tftp://172.26.1.33/syslog_rfc5424_flag.txt"
```

Run the **display smart manager policy-info active** command to display an active policy.

```
DeviceA# display smart manager policy-info active
Key: P - Priority              :L - Low, H - High, N - Normal
       S - Scheduling node :A - Active, P - Pending

No.   Job id    P S Status       Time Of Event          Event Type      Policy Name
1     1          N A suspend     2021-03-19 18:05:59   timer           policy_timer
```

Run the **display smart manager hist-info events detailed** command to display SEM event history.

```
DeviceA# display smart manager hist-info events detailed
No.         Job id    Event Type          Time                          Policy name
1           1          timer              2021-03-22 0:0:10          policy_timer
            Class: reset, Policy Type: app
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
smart manager app policy_2
!
smart manager app policy_timer
 event tag event_timer timer cron cron-entry "0 0 * * *"
 act aciton_timer1 cli command "enble"
 act action_timer2 cli command "copy flash:syslog_rfc5424_flag.txt tftp://172.26.1.33/syslog_rfc5424_flag.txt"
 commit
```

```
!
end
```

1.5.4 Configuring a Syslog Monitoring Event

1. Requirements

Though the interface status is down, Device A sends logs. When the logs are detected, the interface status must be restored to up.

2. Topology

Figure 1-5 Configuring a Syslog Monitoring Event



3. Notes

(1) Create a policy.

(2) Configure a syslog monitoring event.

(3) Configure actions.

(4) Configure policy trigger parameters.

(5) Submit the policy configuration.

4. Procedure

(1) Create a policy with the name policy_syslog on Device A.

```
Device> privilege
DeviceA# config-system terminal
DeviceA(cfg)# smart manager app policy_syslog
```

(2) Configure an event with the name event_syslog1 in the policy_syslog policy, set the type to syslog, and put logs with the content "LINEPROTO-5-UPDOWN" within the detection scope.

```
DeviceA(sem-app)# event tag event_1 syslog pattern "LINEPROTO-5-UPDOWN" priority-level critical
```

(3) Configure an event with the name event_syslog2 in the policy_syslog policy, set the type to syslog, and put logs with the content "GigabitEthernet 0/1" within the detection scope.

```
DeviceA(sem-app)# event tag event_syslog2 syslog pattern "GigabitEthernet 0/1"
```

(4) Configure an event with the name event_syslog3 in the policy_syslog policy, set the type to syslog, and put logs with the content "changed state to down" within the detection scope. The conditional relationship between event_syslog3 and event_syslog1/event_syslog2 is an AND relation.

```
DeviceA(sem-app)# event correlate and tag event_syslog3 syslog pattern "changed state to down"
```

(5) Configure actions of the policy_syslog policy.

Configure action_syslog_1 to run the **enable** command to enter the privileged EXEC mode.

```
DeviceA(sem-app)# act action_syslog_1 cli command “privilege”
```

Configure action_syslog_2 to run the **config-system terminal** command to enter the global configuration mode.

```
DeviceA(sem-app)# act action_syslog_2 cli command "config-system terminal"
```

Configure action_syslog_3 to run the **interface GigabitEthernet 0/1** command to enter the interface configuration mode.

```
DeviceA(sem-app)# act action_syslog_3 cli command "interface GigabitEthernet 0/1"
```

Configure action_syslog_4 to wait for five seconds.

```
DeviceA(sem-app)# act action_syslog_4 wait 5
```

Configure action_syslog_5 to run the **unset shutdown** command to set the interface status to up.

```
DeviceA(sem-app)# act action_syslog_5 cli command "unset shutdown"
```

Configure action_syslog_6 to run the **end** command to exit the privileged EXEC mode.

```
DeviceA(sem-app)# act action_syslog_6 cli command "end"
```

Configure action_syslog_7 to run the **display interfaces status | include UP** command to display the interface status.

```
DeviceA(sem-app)# act action_syslog_7 cli command "display interfaces status | include UP"
```

(6) Submit the policy.

```
DeviceA(sem-app)# commit
```

(7) Exit the policy editing status.

```
DeviceA(sem-app)# exit
```

5. Verification

Run the **display smart manager policy-info registered** command to display a registered SEM policy.

```
DeviceA# display smart manager policy-info registered
 No.   Name                                                    Type      Event Type
   1   policy_syslog                                           app   syslog
  event_syslog1: syslog: pattern "LINEPROTO-5-UPDOWN"
  event_syslog2: syslog: pattern "GigabitEthernet 0/1"
  event_syslog3: syslog: pattern "changed state to down"
  act action_syslog_1 cli command “privilege”
  act action_syslog_2 cli command "config-system terminal"
  act action_syslog_3 cli command "interface GigabitEthernet 0/1"
  act action_syslog_4 wait 5
  act action_syslog_5 cli command "unset shutdown"
  act action_syslog_6 cli command "end"
  act action_syslog_7 cli command "display interfaces status | include UP"
```

Run the **display smart manager policy-info active** command to display an active policy.

```
DeviceA# display smart manager policy-info active
Key: P - Priority             :L - Low, H - High, N - Normal
       S - Scheduling node :A - Active, P - Pending

No.   Job id     P S Status       Time Of Event          Event Type      Policy Name
```

```
1      10          N A suspend      2021-03-23 11:01:49   syslog            policy_syslog
```

Set GigabitEthernet 0/1 to the down status to verify whether the policy takes effect.

```
DeviceA# config-system terminal
DeviceA(cfg)# interfce GigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)#shutdown
*Mar 23 11:02:03: %LINK-5-CHANGED: Interface GigabitEthernet 0/1, changed state to administratively down.
*Mar 23 11:02:03: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet 0/1, changed state to down.
*Mar 23 11:02:03: %SEM-6-PLCY_TRG: Policy policy_syslog id 7 is triggered by event 10, and the policy-info instance id is 12.
*Mar 23 11:02:11: %LINK-3-UPDOWN: Interface GigabitEthernet 0/1, changed state to up.
*Mar 23 11:02:11: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet 0/1, changed state to up.
```

Run the **display smart manager hist-info events detailed** command to display SEM event history.

```
DeviceA# display smart manager hist-info events detailed
No.        Job id     Event Type          Time                          Policy name
1          10          syslog              2021-03-23 11:02:03          policy_syslog
           Class: reset, Policy Type: app
```

6. Configuration Files

```
hostname DeviceA
!
smart manager app policy_syslog
 event tag event_syslog1 syslog pattern "LINEPROTO-5-UPDOWN"
 event tag event_syslog2 syslog pattern "GigabitEthernet 0/1"
 event correlate and tag event_syslog3 syslog pattern "changed state to down"
 act action_syslog_1 cli command “privilege”
 act action_syslog_2 cli command "config-system terminal"
 act action_syslog_3 cli command "interface GigabitEthernet 0/1"
 act action_syslog_4 wait 5
 act action_syslog_5 cli command "unset shutdown"
 act action_syslog_6 cli command "end"
 act action_syslog_7 cli command "display interfaces status | include UP"
 commit
!
end
```

1.5.5 Configuring a Track Monitoring Event

1. Requirements

To facilitate management, a log is printed when the status of a specified or random tracked entity of Device A changes.

2. Topology

Figure 1-6 Topology of a Track Monitoring Event



3. Notes

(1) Create a policy.

(2) Configure a track monitoring event.

(3) Configure actions.

(4) Submit the policy configuration.

4. Procedure

(1) Create a policy with the name policy_track on Device A.

```
DeviceA> privilege
DeviceA# config-system terminal
DeviceA(cfg)# smart manager app policy_track
```

(2) Configure an event with the name event_track in the policy_track policy, set the type to track, and detect all tracked objects in the down status.

```
DeviceA(sem-app)# event tag event_track track state down
```

(3) Configure actions of the policy_track policy.

Configure action_track1 to print "track obj down".

```
DeviceA(sem-app)# act action_tarck1 syslog msg “track obj down.”
```

Configure action_track2 to run the **enable** command to enter the privileged EXEC mode.

```
DeviceA(sem-app)# act action_track2 cli command “privilege”
```

Configure action_track3 to run the **config-system terminal** command to enter the global configuration mode.

```
DeviceA(sem-app)# act action_track3 cli command "config-system terminal"
```

Configure action_track4 to run the **interface GigabitEthernet 0/1** command to enter the interface configuration mode.

```
DeviceA(sem-app)# act action_track4 cli command "interface GigabitEthernet 0/1"
```

Configure action_track5 to run the **unset shutdown** command to set the interface status to up.

```
DeviceA(sem-app)# act action_track5 cli command "unset shutdown"
```

(4) Submit the policy.

```
DeviceA(sem-app)# commit
```

(5) Exit the policy editing status.

```
DeviceA(sem-app)# exit
```

(6) Configure a track monitoring event.

```
DeviceA(cfg)# track 1 Interface GigabitEthernet 0/1 line-protocol
DeviceA(cfg-track)# delay down 10
```

5. Verification

Run the **display smart manager policy-info registered** command to display a registered SEM policy.

```
DeviceA# display smart manager policy-info registered
No.   Name                                                    Type      Event Type
  1   policy_track                                            app   track
  event_track: track: all state down
  act action_track1 syslog msg "track obj down"
  act action_track2 cli command “privilege”
  act action_track3 cli command "config-system terminal"
  act action_track4 cli command "interface GigabitEthernet 0/1"
  act action_track5 cli command "unset shutdown"
```

Run the **display smart manager policy-info active** command to display an active policy.

```
DeviceA# display smart manager policy-info active
Key: P - Priority          :L - Low, H - High, N - Normal
     S - Scheduling node :A - Active, P - Pending

No.   Job id    P S Status      Time Of Event          Event Type     Policy Name
1     4          N A suspend    2021-03-22 17:21:49   track          policy_track
```

Set GigabitEthernet 0/1 to the down status to verify whether the policy takes effect.

```
DeviceA# config-system terminal
DeviceA(cfg)# interface gigabitEthernet 0/1
DeviceA(cfg-if-GigabitEthernet 0/1)# shutdown
DeviceA(cfg-if-GigabitEthernet 0/1)# end
*Mar 22 17:21:59: %SEM-6-PLCY_TRG: Policy track id 4 is triggered by event 4, and the policy-info instance id is 8.
*Mar 22 17:21:59: %SEM-6-COMM: track: track obj down
*Mar 22 17:22:01: %LINK-3-UPDOWN: Interface GigabitEthernet 0/1, changed state to up.
*Mar 22 17:22:01: %LINEPROTO-5-UPDOWN: Line protocol on Interface GigabitEthernet 0/1, changed state to up.
```

Run the **display smart manager hist-info events detailed** command to display SEM event history.

```
DeviceA# display smart manager hist-info events detailed
No.         Job id    Event Type        Time                      Policy name
1           4          track            2021-03-22 17:21:59       track
            Class: reset, Policy Type: app
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
smart manager app policy_track
```

```
 event tag event_track track state down
act action_track1 syslog msg "track obj down"
 act action_track2 cli command “privilege”
 act action_track3 cli command "config-system terminal"
 act action_track4 cli command "interface GigabitEthernet 0/1"
 act action_track5 cli command "unset shutdown"
commit
!
track 1 Interface GigabitEthernet 0/1 line-protocol
 delay down 10
!
end
```

1.5.6 Configuring a Route Monitoring Event

1. Requirements

A log is printed when a specified route of device A is unreachable.

2. Topology

Figure 1-7 Configuring a Route Monitoring Event



3. Notes

(1) Create a policy.

(2) Configure a route event.

(3) Configure actions.

(4) Configure policy trigger parameters.

(5) Submit the policy configuration.

4. Procedure

(1) Create a policy with the name policy_a on Device A.

(2) Configure an event with the name event_1 for the policy_a policy, set the type to route, and detect the unreachable route destined for 192.168.1.5/32.

(3) Configure log “route 192.168.1.5/32 unreachable.” for policy_a.

(4) Submit the policy.

(5) Exit the policy configuration.

```
Device# config-system terminal
Device(cfg)# smart manager applet policy_a
Device(sem-app)# event tag event_1 route unreachable ipv4 192.168.1.5/32
Device(sem-app)# action action_1 syslog msg “route 192.168.1.5/32 unreachable.”
Device(sem-app)# commit
```

```
Device(sem-app)# exit
Device(cfg)#
```

5. Verification

```
Device# display smart manager policy registered
No.    Name                Type       Event Type
    1    policy_a          applet    route
   event_1: route: unreachable ipv4 192.168.1.5/32
   action action_1 syslog msg "route 192.168.1.5/32 unreachable"
```

6. Configuration Files

Device A configuration file

```
hostname DeviceA
!
smart manager applet policy_a
 event tag event_1 route unreachable ipv4 192.168.1.5/32
 action action_1 syslog msg "route 192.168.1.5/32 unreachable."
 Commit
!
End
```

1.5.7 Configuring a CLI Monitoring Event

1. Requirements

A log is printed when the **trigger test** command is input in the privileged EXEC mode on device A.

2. Topology

Figure 1-8 Configuring a CLI Monitoring Event



3. Notes

(1) Create a policy.

(2) Configure a CLI event.

(3) Configure actions.

(4) Commit the policy configuration.

4. Procedure

(1) Create a policy with the name policy_a on Device A.

(2) Configure an event with the name event_1 for the policy_a policy, set the type to CLI, and detect command **trigger test**.

(3) Configure log “trigger test is running.” for policy_a.

(4) Commit the policy.

(5) Exit the policy configuration.

```
Device# config-system terminal
Device(cfg)# smart manager applet policy_a
Device(sem-app)# event tag event_1 cli pattern "trigger test"
Device(sem-app)# action action_1 syslog msg “trigger test is running.”
Device(sem-app)# commit
Device(sem-app)# exit
Device(cfg)#
```

5. Verification

```
Device# display smart manager policy registered
No.   Name        type      Event Type
1      policy_a applet    cli
  event_1: cli: pattern “trigger test”
  action action_1 syslog msg "track obj down."
```

6. Configuration Files

Device A configuration file

```
hostname Device A
!
smart manager applet policy_a
  event tag event_1 cli pattern "trigger test"
  action action_1 syslog msg “trigger test is running.”
  commit
!
End
```

Contents

1 Configuring IDS 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.2 Configuring IDS 1

1.2.1 Overview 1

1.2.2 Restrictions and Guidelines 1

1.2.3 Procedure 1

1.3 Monitoring 2

1 Configuring IDS

1.1 Introduction

1.1.1 Overview

Intelligent Diagnosis System (IDS) is used for device diagnosis in fault scenarios.

A device may fail to work for various internal and external reasons. When a failure occurs, the O&M personnel needs to identify the fault by collecting related information and generate a rectification plan immediately. To reduce the time for identifying a fault and minimize the impact of faults on services, you can use IDS as the assistant for fault identification.

1.1.2 Principles

IDS extracts typical fault scenarios for diagnosis based on software and hardware issues. The running of IDS includes the following phases:

(1) Failure occurrence: If a fault is detected by a software component associated with a specific fault scenario, the fault is reported to IDS.

(2) Troubleshooting: IDS receives the fault report and asks the software component associated with the fault scenario to start diagnosis.

(3) Information collection: IDS receives a message when the software component finishes diagnosis. If all software components associated with the fault scenario finish diagnosis, the log framework is instructed to collect and package the relevant operational and diagnostic logs. The O&M personnel can locate the fault based on the packaged logs.

1.2 Configuring IDS

1.2.1 Overview

You can specify a specific fault scenario or all fault scenarios that can trigger the IDS function.

1.2.2 Restrictions and Guidelines

- **all**: specifies all fault scenarios. *scenario-name*: specifies a specific fault scenario.
- **enable**: enables the IDS function. **disable**: disables the IDS function.
- Run the **no** { *scenario-name* | **all** } command to restore the default configuration.

1.2.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the IDS configuration mode.

ids

(4) Configure the IDS function.

{ *scenario-name* | **all** } { **enable** | **disable** }

The IDS function is enabled by default.

1.3 Monitoring

Run the **display** command to display the configurations.

Run the **debugging** command to output debugging information.

Caution

The output debugging information of the **debugging** command occupies system resources. Therefore, disable debugging immediately after use.

Table 1-1 Monitoring

Command	Purpose
display ids scenario { *scenario-name* | **all** }	Displays the state of the IDS function.
display base-on-scene debugging { *scenario-name* | **all** }	Displays the state of the IDS debugging function.

Contents

1 Configuring Intelligent Monitoring 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Configuration Task Summary 4
1.3 Notes 4
1.4 Configuring the Energy-saving Mode of Fans 5
1.4.1 Overview 5
1.4.2 Restrictions and Guidelines 5
1.4.3 Procedure 5
1.5 Configuring Temperature Thresholds 5
1.5.1 Overview 5
1.5.2 Restrictions and Guidelines 5
1.5.3 Procedure 5
1.6 Monitoring 6

1 Configuring Intelligent Monitoring

1.1 Introduction

1.1.1 Overview

Intelligent monitoring is an intelligent hardware management function of devices, including intelligent power management, intelligent fan speed adjustment, and intelligent temperature monitoring. The intelligent monitoring function monitors the power supply of the chassis, processes card power-on/off and monitors its power, automatically adjusts fan speed based on the change of the ambient temperature, and monitors temperature change of the cards in real time in order to give notifications to users. When the card temperature exceeds the maximum temperature of the device, the intelligent monitoring function automatically powers off the cards to protect the device.

By default, the intelligent monitoring function is enabled after the device is powered on. Users can use this function without need of any configuration. Apart from the default configuration, the system provides a variety of configuration methods so that you can change the default configuration of the system as necessary.

1.1.2 Principles

1. Basic Concepts

- Operating mode of power supply

  The power supply can operate in one of the following two modes: non-redundancy mode and N+M redundancy mode.

  - Non-redundancy mode: It is also called the superposition mode. The power supply operates in non-redundancy mode by default. In this mode, the total power of the system is the sum of power of all power modules.
  - N+M redundancy mode: The system has N+M power modules in total, where M indicates the number of configured redundant power modules and N indicates the number of available power modules of the system. The total power of the system is the sum of output power of the N power modules.

- Power supply priority

  The power supply priorities of cards refer to the power-on/off priorities of the cards. A card with a higher power supply priority is powered on earlier, and powered off later. When the system power is insufficient and thus some line cards need to be powered off, a line card with the lowest priority is powered off first. The default power supply priorities of the cards are as follows: supervisor module > FE card > VSL card > line card. If the cards are of the same type, a card with a smaller slot ID has a higher priority. You can change the power supply priorities of the line cards only. The power supply priority of a line card ranges from 1 to 16, where 1 indicates the lowest priority and 16 the highest priority.

Note

When the system power is insufficient and thus some cards need to be automatically powered off, only the line cards will be powered off but the supervisor modules and FE cards will not be automatically powered off.

- Operating modes of fans

  A fan tray on a chassis device normally operates in one of the following three modes:

  - Normal mode: It is the default operating mode. The fans initially run at intermediate speed. As the ambient temperature changes, the rotating speed of the fans automatically decreases or increases to achieve the best heat dissipation effect.
  - Quiet mode: The fans initially run at relatively low speed. As the ambient temperature changes, the rotating speed of the fans automatically decreases or increases. In this mode, the rotating speed of the fans is lower than that of a fan in normal mode. Therefore, in quiet mode, fans' noise is further reduced.
  - User-defined mode: In this mode, users define the rotating speed of the fans and the rotating speed of all the fan trays is the same (user-defined value). The rotating speed of the fans can be set to seven levels.

Note

In Virtual Switching Unit (VSU) mode, the fans in two chassis operate in the same working mode.

- Temperature threshold

  During temperature monitoring, if the temperature reaches a certain threshold, the system takes a corresponding action. For example, if the temperature reaches an alarm threshold, the system generates a Syslog alarm. If the temperature reaches a hazard threshold, the corresponding card is powered off.

2. Power Monitoring

A device supports multiple power modules. However, different types of power modules cannot be mixed. When a power supply is inserted, the system identifies this power supply and checks whether it conflicts with the original power supply type. If yes, a power supply with a lower power is selected. The alarm indicator turns yellow and a log alarm is generated. After the power supply is removed, the abnormality is cleared and the system restores to normal.

3. Configuration of Power Input Mode

A device supports DC or AC power supplies. The two types of power supplies have different input voltage calculation methods. Therefore, the power input mode of the device must be configured manually based on the DC/AC mode of the power supply to ensure that the input voltage value displayed by the **display power** command is correct.

4. Power Redundancy

You can implement redundancy of power supplies in the chassis by configuring power redundancy. As the standby power supply, the redundant power supplies are not involved in system power allocation to avoid in-service power supply failures. When setting the power supply to the redundancy mode, you need to specify the number of redundant power supplies.

After the N+M redundancy mode is configured, the system automatically monitors the redundant power supplies. As the system continuously operates or a card or power supply is removed or inserted, the remaining system power changes. The redundancy mode configured by you may fail to meet the normal power supply needs of the system. In redundancy mode, when the remaining system power is insufficient, the system alarm indicator turns yellow. In this case, you need to modify power redundancy configuration. If you have configured the function of powering off some cards in the case of insufficient power, the specified line cards will be powered off to meet system power allocation needs.

Caution

- When the power modules are mixed, they cannot work in redundancy mode.
- When the specified number of redundant power modules is equal to or greater than the number of available power modules, the redundancy configuration fails.
- The maximum number of redundant power modules is the remaining system power divided by the power of a single power supply. If the system determines that redundancy will result in insufficient remaining system power (exceeding the maximum allowed value), the redundancy configuration fails.

5. Automatic Power-off of Line Cards

After this function is enabled, some line cards are powered off automatically based on the current power supply priorities to meet system power allocation needs in the case of power deficiency.

Meanwhile, the hardware FPGA automatically notifies the CPU of the power deficiency, enabling the software to further process this problem.

The automatic power-off of line cards is disabled by default and can be enabled by running the **power priority card auto-down** command. After this function is enabled, the system automatically monitors the system power and inspects power allocation status of the current system every two minutes. If power deficiency is detected, the system automatically selects a line card with the lowest priority from the line cards in power-on status and powers off this line card.

If this function is not enabled, the active supervisor module prints a Syslog alarm when receiving the power deficiency notification of the field programmable gate array (FPGA) and the alarm indicator turns red.

The power supply priority of a line card can be modified by running the **power priority** command.

6. Manual Power-on/off of Cards

You can manually power off/on a specified card by running a power command.

- When a card is in power-on status, you can run the **power off** command to power off the card.
- When a card is in power-off status, you can run the **power on** command to power on the card.

Caution

- The device hardware boasts the power-on protection function. When the temperature of a card exceeds 50°C, the card cannot be powered on and a notification indicates card power-on failure.
- The active supervisor module of the chassis cannot be manually powered off. If you attempt to specify a card as the active supervisor module of the chassis, a notification is displayed. In VSU mode, the local active supervisor modules in two chassis cannot be manually powered off.
- When the supervisor module of a chassis is reset, the cards that are powered off by running the **power off** command can be powered on again.

7. Intelligent Speed Adjustment of Fans

When the fans work in normal, quiet or energy-saving mode, the rotating speed is automatically adjusted as the temperature changes.

The system automatically specifies a default start rotating speed for the fans based on the current operating mode of the fans. As the ambient temperature rises or drops, the fans automatically raise or reduce their rotating speed to dissipate heat and ensure that the noise is low.

The system provides a user-defined mode to ensure that all the fans operate at a user-specified rotating speed. In user-defined mode, the rotating speed of fans is fixed and not adjusted as the ambient temperature changes. Therefore, the normal mode or quiet mode is recommended so that the rotating speed of fans will automatically change as the temperature changes to best protect the device from a fault due to over-temperature.

8. Intelligent Temperature Monitoring

The active supervisor module monitors the temperature of all cards and notifies users of temperature change when the CA system runs normally.

The system defines three types of temperature, including supervisor module temperature, CPU temperature and media access controller (MAC) temperature, and two temperature thresholds, namely, alarm threshold and hazard threshold. Only hazard thresholds are defined for the CPU temperature and MAC temperature. In intelligent monitoring mode, the system monitors the temperature of all cards every two minutes. When the system discovers that the temperature of a card reaches or exceeds the alarm threshold, the active supervisor module generates a Syslog message and the alarm indicator on the panel turns yellow. If the temperature of a card reaches or exceeds the hazard threshold, the active supervisor module powers off the card and generates a Syslog message. The alarm indicator on the panel turns red.

Warning

The active supervisor module of the chassis or the global active supervisor module in VSU mode generates a Syslog message when discovering that its temperature exceeds the hazard threshold. At the same time, the alarm indicator on the panel turns red. However, the active supervisor module or global active supervisor module will not be automatically powered off. In this case, you need to take effective actions to cool the active supervisor module or manually remove it for cooling.

1.2 Configuration Task Summary

The intelligent monitoring configuration includes the following tasks: All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring the Energy-saving Mode of Fans
- Configuring Temperature Thresholds

1.3 Notes

This section describes notes for configuring intelligent monitoring functions.

Table 1-1 describes intelligent monitoring functions supported by different products.

Table 1-1 Intelligent Monitoring Functions Supported by Different Products

Function	Product Supporting the Function
Configuring the Energy-saving Mode of Fans	S6980-64QC
Configuring Temperature Thresholds	S6980-64QC, S6580-48CQ8QC, S6580-24DC8QC

1.4 Configuring the Energy-saving Mode of Fans

Note

Only S6980-64QC supports configuration of the energy-saving mode of fans.

1.4.1 Overview

This function can be used to enable or disable the energy-saving mode of fans based on needs.

1.4.2 Restrictions and Guidelines

When a 400G ZR module is inserted into the device, the system automatically disables the energy-saving mode of fans to facilitate heat dissipation of the device.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the operating mode of fans.

fan energy-saving enable

The energy-saving mode of fans is disabled by default.

1.5 Configuring Temperature Thresholds

Note

Only S6980-64QC, S6580-48CQ8QC, and S6580-24DC8QC support configuration of temperature thresholds.

1.5.1 Overview

This function can be used to modify the default alarm thresholds and hazard thresholds of temperature.

1.5.2 Restrictions and Guidelines

- Temperature thresholds are set based on chassis configuration. All line cards of a chassis share the same set of thresholds.
- When the set temperature thresholds exceed the allowed values, the threshold settings are invalid.
- When the set alarm thresholds are excessively low, the system frequently generates alarm logs.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure temperature thresholds of cards.

threshold set temperature [**switch** *device-id*] { **board** *temperature* | **board shutdown** *temperature* } | **cpu shutdown** *temperature* | **mac shutdown** *temperature* }

By default, the alarm threshold of the supervisor module is 56°C and the hazard threshold is 80°C. No alarm thresholds are defined for the CPU and MAC, but the hazard thresholds are 100°C.

1.6 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Figure 1-1 Monitoring

Command	Purpose
display fan [[[*device-id*] *fan-id*] **detail** | **version | speed**]	Displays fan information.
display temperature	Displays line card information such as current temperature and temperature thresholds.
display power [**priority** | **version**]	Displays power information about line cards and respective power supply priorities.

Contents

1 Configuring HDF 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Restrictions and Guidelines 2
1.3 Configuration Task Summary 2
1.4 Configuring Basic Features 2
1.4.1 Overview 2
1.4.2 Procedure 2
1.5 Configuring Single-time Real-time Detection for a Board 3
1.5.1 Overview 3
1.5.2 Restrictions and Guidelines 3
1.5.3 Procedure 3
1.6 Configuring Alerting for HDF Detection Items 3
1.6.1 Overview 3
1.6.2 Procedure 3
1.7 Configuring the Detection Period for HDF Detection Nodes 4
1.7.1 Overview 4
1.7.2 Procedure 4
1.8 Monitoring 4

1 Configuring HDF

1.1 Introduction

1.1.1 Overview

Network devices are made up of various semiconductor chips and components. After the devices run for a long period, components may experience functional failures due to factors such as the environment, external forces, and aging, which affects normal services. You can configure Hardware Diagnostics Function (HDF), which allows you to query the status of hardware components on the device in real time and automatically reports faults by using alarm notifications, Trap messages, and Syslog records.

1.1.2 Principles

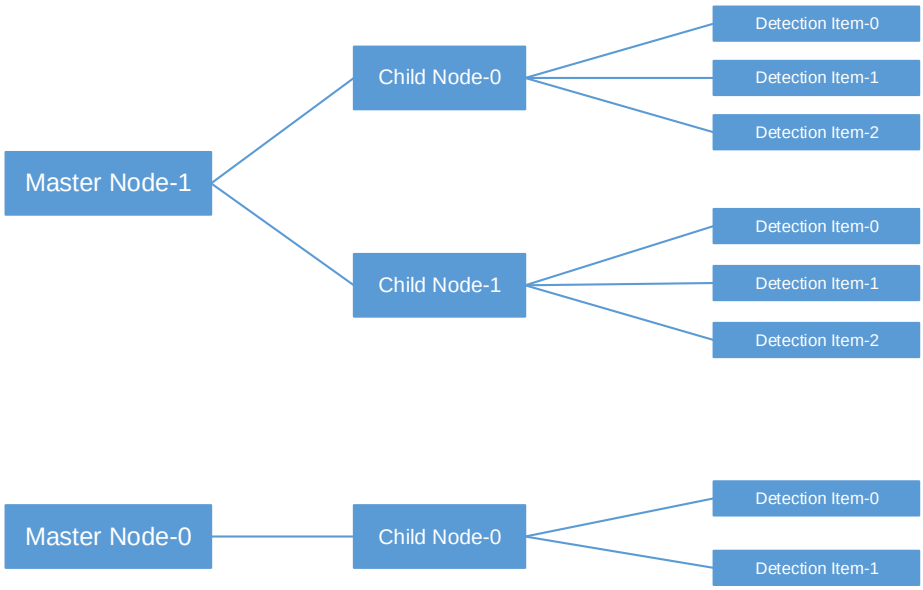
Based on the idea of logical detection nodes, HDF adopts unified logical detection nodes for anomaly detection on hardware components, implementing various functions such as node management, detection mechanism, diagnostic analysis, self-healing actions, and log alarms.

The master nodes classified based on hardware component types include the CPU, storage, electromechanical module, logic module, Double Data Rate (DDR) Synchronous Dynamic Random Access Memory (SDRAM), interface optical module, and MAC switching chip. As shown in Figure 1-1, HDF supports splitting detection items by using child nodes and provides management capabilities for detection objects in three levels.

HDF manages hardware device detection and shows the status of hardware components of the device. The supported functions are as follows:

- Enabling detection/configuring the detection period
- Displaying the status of detection objects
- Single execution of all detection nodes and items

Figure 1-1 HDF Node Management



1.2 Restrictions and Guidelines

- The running environment of HDF includes a management engine board, a line card, and a switch fabric module. When you configure HDF, ensure that the related components run properly.
- A detection node dynamically registers for a board in the system. In this case, the information of some detection nodes may fail to be queried instantly.

1.3 Configuration Task Summary

HDF configuration includes the following tasks:

(1) Configuring Basic Features

(2) Configuring Single-time Real-time Detection for a Board

(3) (Optional) Configuring Alerting for HDF Detection Items

(4) (Optional) Configuring Alerting for HDF Detection Items

1.4 Configuring Basic Features

1.4.1 Overview

You can manage logical detection nodes using HDF by configuring the detection function on logical detection nodes.

1.4.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the HDF detection function. All the configuration steps below are optional. Perform the configuration steps as needed.

hdf enable { **all** | **slot** *slot-id* [**main-node** *name* [**child-node** *name* [**item** *item-id*]]] }

The detection function is enabled for a board by default.

1.5 Configuring Single-time Real-time Detection for a Board

1.5.1 Overview

The detection periods of different components of a board vary. In most cases, a round-robin HDF detection is performed for each item periodically. You can configure single-time real-time detection for a board in particular scenarios. For example, you require that the status of a device and all detection items are checked in real-time for troubleshooting the device during initial setup.

1.5.2 Restrictions and Guidelines

After you configure the single-time real-time detection function, real-time detection is performed for all enabled detection nodes and detection items of the board. You must enable detection nodes and detection items by using commands before you perform real-time detection.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Configure single-time real-time detection for a board.

hdf once-run slot *slot-id*

After you configure this function, real-time detection is performed for all enabled detection nodes and detection items of the board. You can run the **display hdf** command to display the detection result.

1.6 Configuring Alerting for HDF Detection Items

1.6.1 Overview

You can configure the fault alerting function for specific detection items of a board in the CLI. You can enable or disable this function as needed.

1.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the fault alerting function for an HDF detection item of a board.

hdf alarm enable slot *slot-id* **main-node** *name* **child-node** *name* **item** *item-id*

By default, the fault alerting function for HDF detection items is automatically configured.

1.7 Configuring the Detection Period for HDF Detection Nodes

1.7.1 Overview

By configuring the detection period of the HDF detection nodes of a hardware component, you can dynamically adjust the detection period of detection child nodes, which helps meet your service requirements.

1.7.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the detection period for detection nodes.

hdf set cycle slot *slot-id* **main-node** *name* **child-node** *name* **time** *cycle-num*

If you do not configure the detection period, the default detection period is used.

1.8 Monitoring

Run the **display** command to check the configuration.

Table 1-1 Monitoring

Command	Purpose
display hdf all	Displays the status of the hardware components of the device.
display hdf fault	Displays the status of the faulty hardware components of the device.
display hdf slot *slot-id* [**fault**]	Displays the status of the hardware components of a specific slot.
display hdf detail-info slot *slot-id* **main-node name** *child-node name* [**item** *item-id*]	Displays the status information of the child detection nodes or detection items of a slot.

Contents

1 Configuring Hotkeys 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.2 Restrictions and Guidelines 2

1 Configuring Hotkeys

1.1 Introduction

1.1.1 Overview

Hotkeys, also known as magic keys, help you to perform maintenance and debugging on devices. You can view the memory information, interrupt information, lock usage, and process information of a device as the network administrator by using hotkeys when the device is faulty or cannot be controlled properly. You can also perform other actions, including one-click collection and enabling or disabling of the logging function of the console, by using hotkeys. You can use the functions provided by hotkeys to obtain key debugging information of the device, locate the cause of exceptions, and restore the device.

1.1.2 Principles

1. Hotkeys

Hotkeys, also known as magic keys, are input to a device to obtain debugging information of the device in the event of exceptions. You can also restore the services on the device by using hotkeys. Supported hotkeys are 4Esc key combinations (press the **Esc** key on the keyboard four consecutive times) and 4$ key combinations (press and hold the **Shift** key and then press the **$** key on the keyboard four consecutive times).

Note

Hotkeys can be input only on the serial console of the device and are not available on other types of terminals.

2. 4Esc Key Combination

The 4Esc key combinations are supported by the device and provide the following functions:

- Enable and disable the output function of the console.
- Display the information about usage of all locks.
- Perform one-click collection of debugging information.
- Display the help Information about hotkeys.
- Display the CLI debugging information.
- Display the information about interrupts.
- Stop a process by specifying the process identifier (PID).
- Display the information about the running processes.
- Display the memory statistics of the device.
- Enable the timer for periodic output of the information about interrupts.
- Disable the timer for periodic output of the information about interrupts.
- Disable logging of the console.
- Display the information about the context switch and running time for all threads.

- Display the information about the stacks of all CPU cores.
- Display the status information about all threads.
- Display the kernel stack information of all threads.
- Start the task to output periods (.) on the console.
- Stop the task to output periods (.) on the console.

Note

- 4Esc key combinations: Press the **Esc** key on the keyboard four consecutive times and then press a letter key. For more information about letter keys of specific functions, see *Hotkey Commands*.
- You can use the 4@ key combinations (press and hold the **Shift** key, press the @ key on the keyboard four consecutive times, and then press the same letter key as that of the 4Esc key combinations) to replace the 4Esc key combinations. For more information about letter keys of specific functions, see *Hotkey Commands*.

3. 4$ Key Combinations

The 4$ key combinations are supported by the device and provide the following functions:

- Set the kernel log priority level.
- Display the help information about system hotkeys.
- Display the stack information of all active CPUs.
- Display the memory information of the device.
- Display the register information of the device.
- Display the timer information of the device.
- Display the kernel stack information of all threads.
- Display the information about threads that are blocked of the device.

Note

4$ key combinations: Press and hold the **Shift** key, press the **$** key on the keyboard four consecutive times, and then press a letter key. For more information about letter keys of specific functions, see *Hotkey Commands*.

1.2 Restrictions and Guidelines

The hotkeys can be used only on serial consoles rather than other types of remote terminals, such as terminals connected through Telnet and SSH.

Contents

1 Configuring Network Packet Loss Detection 1

1.1 Introduction 1

1.1.1 Overview 1

1.1.2 Principles 1

1.2 Configuring Network Packet Loss Detection 2

1.2.1 Overview 2

1.2.2 Restrictions and Guidelines 2

1.2.3 Configuration Tasks 2

1.2.4 Configuring a Standard Parameter-based Probe Packet Profile 2

1.2.5 Configuring a Customized Probe Packet Profile 3

1.2.6 Sending Probe Packets Based on the Profile 4

1.3 Monitoring 4

1.4 Configuration Examples 5

1.4.1 Configuring Packet Forwarding on the Local Device 5

1 Configuring Network Packet Loss Detection

Note

Network packet loss detection is not supported by the S6980-64QC, S6920-4C, S6580-24DC8QC, or S6510-48VS8CQ.

1.1 Introduction

1.1.1 Overview

Data transmitted over the network passes through many devices and network links. Packets may be lost at any link. The causes of packet loss between devices are complex and diverse. Even if the device causing packet loss is identified, it is difficult to reproduce the fault in an environment with complex network configurations. The network packet loss detection function can enable the construction of packets containing simulated packet loss issues, facilitating troubleshooting by identifying the causes of packet loss through alternative methods.

1.1.2 Principles

When a device simulates packet loss, the device constructs a probe packet based on the configured profile and simulates the forwarding process of the discarded packet on the device by specifying the source interface of the probe packet.

Users should configure a probe packet profile based on the characteristics of lost packets. The device supports two methods to configure the probe packet profile to simulate lost packets:

- Directly specify the packet content and configure a customized probe packet profile.
- Specify key fields of the packet data link layer, network layer, transport layer, and application layer based on the packet format, and configure a standard parameter-based probe packet profile.

  This method supports IPv4 TCP, IPv4 UDP, IPv6 TCP, and IPv6 UDP packets. For different types of probe packet profiles, the following key fields can be configured:

  - Key fields in IPv4 TCP probe packets: source MAC address, destination MAC address, source IPv4 address, destination IPv4 address, VLAN ID, source port number, destination port number, 802.1p priority, DSCP value, TTL value, transport layer and network layer checksum, and TCP flags.
  - Key fields in IPv4 UDP probe packets: source MAC address, destination MAC address, source IPv4 address, destination IPv4 address, VLAN ID, source port number, destination port number, 802.1p priority, DSCP value, TTL value, transport layer and network layer checksum, and UDP packet length.
  - Key fields in IPv6 TCP probe packets: source MAC address, destination MAC address, source IPv6 address, destination IPv6 address, VLAN ID, source port number, destination port number, 802.1p priority, flow label, hop limit, Traffic Class field, transport layer and network layer checksum, and TCP Flags.
  - Key fields in IPv6 UDP probe packets: source MAC address, destination MAC address, source IPv6 address, destination IPv6 address, VLAN ID, source port number, destination port number, 802.1p priority, flow tag, hop limit, traffic class, transport layer and network layer checksum, and UDP packet length.

1.2 Configuring Network Packet Loss Detection

1.2.1 Overview

The network packet loss detection function is used to locate the causes of packet loss in forwarding. Configure probe packet profiles on network devices. After the devices successfully send probe packets, you can use other methods to locate the causes of packet loss.

1.2.2 Restrictions and Guidelines

- Probe packets are not sent to the network but are only forwarded to the local device.

1.2.3 Configuration Tasks

The probe packet profile can be configured either as a standard parameter-based profile or a customized profile. Select one of the configuration modes.

- Configuring a Standard Parameter-based Probe Packet Profile
- Configuring a Customized Probe Packet Profile
- Sending Probe Packets Based on the Profile

1.2.4 Configuring a Standard Parameter-based Probe Packet Profile

1. Overview

Configure key fields in the packet header and application layer data to create a standard parameter-based probe packet profile to simulate lost packets.

2. Restrictions and Guidelines

- For a standard parameter-based probe packet profile, basic network parameters must be configured. Otherwise, the profile cannot be used to construct a probe packet.
- It is not allowed to create a standard parameter-based probe packet profile with the same name as that of a configured customized probe packet profile. If you need to modify the type of the probe packet profile, you should delete the current profile before you configure it. For a standard parameter-based probe packet profile, if different probe packet profiles use the same profile name, the later configured profile will overwrite the earlier configured profile. For some protocol-specific key fields, such as the DSCP field of IPv4 packets and the flow label field of IPv6 packets, if the value of the field is not changed in the subsequent configuration, the value is still saved in the device and takes effect when the probe packet profile is configured for the corresponding protocol.
- A maximum of 20 probe packet profiles can be configured on a device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure network parameters for the probe packet profile. Configure one of the following tasks.

- Configure network parameters for the IPv4 TCP probe packet profile.

detector packet-trace profile *profile-name* **source-mac** *src-mac-address* **destination-mac** *dest-mac-address* [**vlan** *vlan-id* [**8021p** *8021p-value*]] **ip source-ip** *src-ipv4-address* **destination-ip** *dest-ipv4-address* **tcp** [**csum** *csum-value*] [**dscp** *dscp-value*] [**network-length** *network-length-value*] [**ttl** *ttl-value*] **source-port** *src-port-number* **destination-port** *dest-port-number* [**csum** *transport-csum*] [**ack** | **cwr** | **ecn** | **fin** | **nonce** | **push** | **reset** | **syn** | **urgent**]

- Configure network parameters of the IPv4 UDP probe packet profile.
  detector packet-trace profile *profile-name* **source-mac** *src-mac-address* **destination-mac** *dest-mac-address* [**vlan** *vlan-id* [**8021p** *8021p-value*]] **ip source-ip** *src-ipv4-address* **destination-ip** *dest-ipv4-address* **udp** [**csum** *csum-value*] [**dscp** *dscp-value*] [**network-length** *network-length-value*] [**ttl** *ttl-value*] **source-port** *src-port-number* **destination-port** *dest-port-number* [**csum** *transport-csum*] [**udp-length** *udp-length-value*]
- Configure network parameters for the IPv6 TCP probe packet profile.

  detector packet-trace profile *profile-name* **source-mac** *src-mac-address* **destination-mac** *dest-mac-address* [**vlan** *vlan-id* [**8021p** *8021p-value*]] **ipv6 source-ip** *src-ipv6-address* **destination-ip** *dest-ipv6-address* **tcp** [**flowlabel** *flowlabel-value*] [**hoplimit** *hoplimit-value*] [**network-length** *network-length-value*] [**tc** *traffic-class*] **source-port** *src-port-number* **destination-port** *dest-port-number* [**csum** *transport-csum*] [**ack** | **cwr** | **ecn** | **fin** | **nonce** | **push** | **reset** | **syn** | **urgent**]
- Configure network parameters for the IPv6 UDP probe packet profile.

  detector packet-trace profile *profile-name* **source-mac** *src-mac-address* **destination-mac** *dest-mac-address* [**vlan** *vlan-id* [**8021p** *8021p-value*]] **ipv6 source-ip** *src-ipv6-address* **destination-ip** *dest-ipv6-address* **udp** [**flowlabel** *flowlabel-value*] [**hoplimit** *hoplimit-value*] [**network-length** *network-length-value*] [**tc** *traffic-class*] **source-port** *src-port-number* **destination-port** *dest-port-number* [**csum** *transport-csum*] [**udp-length** *udp-length-value*]

  No probe packet profile is configured by default.

(4) (Optional) Configure application layer data for the probe packet profile.
detector packet-trace profile *profile-name* { **payload-base** *payload-data* | **payload-extern-1** *payload-data* | **payload-extern-2** *payload-data* | **payload-extern-3** *payload-data* | **payload-extern-4** *payload-data* }

By default, the application layer data of the probe packet profile is not configured. Instead, it is filled with zeros to its minimum length.

1.2.5 Configuring a Customized Probe Packet Profile

1. Overview

You can create a customized probe packet profile by directly configuring the packet content.

2. Restrictions and Guidelines

- It is not allowed to create a customized probe packet profile with the same name as that of a configured standard parameter-based probe packet profile. If you need to modify the type of the probe packet profile, you should delete the current profile before you configure it.
- Different probe packet profiles have different names. If different probe packet profiles use the same name, the later configured profile will overwrite the earlier configured profile.
- A maximum of 20 probe packet profiles can be configured on a device.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a customized probe packet profile.

detector packet-trace profile *profile-name* { **packet-base** *packet-data* | **packet-extern-1** *packet-data* | **packet-extern-2** *packet-data* | **packet-extern-3** *packet-data* | **packet-extern-4** *packet-data* }

No customized probe packet profile is configured by default.

1.2.6 Sending Probe Packets Based on the Profile

1. Overview

After the probe packet profile is configured, the profile is used to send probe packets to simulate the forwarding process of discarded packets within the device.

2. Restrictions and Guidelines

- The source interface of the probe packet must be a physical port.
- The device forwards and processes packets without differentiating whether they are probe packets. Packet statistics are collected for probe packets.
- The number of sent packets is limited to 1 to 1000.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Send probe packets based on a specified profile to locate the cause of packet loss.

send-detector packet-trace profile *profile-name packet-count src-interface*

No probe packet based on a specified profile is sent by default.

1.3 Monitoring

Run the **display** command to display the profile configuration.

Table 1-1 Displaying Profile Configuration

Command	Purpose
display detector packet-trace profile [*profile-name*]	Displays the configuration of a probe packet profile.
display detector packet-trace detail	Displays the sending statistics of the packet profile.

Run the **debugging** command to check the running status of the function configuration.

Table 1-2 Displaying Debugging Information

Caution

The output of debugging information consumes system resources. Therefore, disable the debugging function immediately after use.

Command	Purpose
debugging packet-trace all	Debugging
debugging packet-trace error	Debugs errors.
debugging packet-trace event	Debugs events.
debugging packet-trace packet	Debugs packets.

1.4 Configuration Examples

1.4.1 Configuring Packet Forwarding on the Local Device

1. Networking Requirements

Packets sent from device A to device C are discarded on device B. The cause of packet discard needs to be determined.

2. Networking Diagram

Figure 1-1 Topology of Network Packet Loss Detection



3. Configuration Roadmap

- Based on the packets sent by device A, configure device B to send packets for probing.

4. Procedure

Configure a probe packet profile on device B and send the probe packet.

```
DEVICE-B(cfg)# detector packet-trace profile test payload-base
01234567890123456789012345678901234567890123456789012345678901234567890123456789
DEVICE-B(cfg)# detector packet-trace profile test payload-extern-1 0123
DEVICE-B(cfg)# detector packet-trace profile test payload-extern-2 0123
DEVICE-B(cfg)# detector packet-trace profile test payload-extern-3 0123
DEVICE-B(cfg)# detector packet-trace profile test payload-extern-4 0123
DEVICE-B(cfg)# detector packet-trace profile test source-mac 1111.2222.3333 destination-mac
1111.4444.5555 vlan 1 8021p 5 ip source-ip 1.1.1.3 destination-ip 1.3.4.5 udp dscp 3 ttl 3 network-length 3
source-port 5223 destination-port 6553
```

```
DEVICE-B(cfg)# end
DEVICE-B# send-detector packet-trace profile test GigabitEthernet 1/0/1 100
```

5. Verification

Run the **display detector packet-trace profile test** command to check whether the test profile is configured.

```
DEVICE-B# display detector packet-trace profile test
---------------------------------------
profile name            : test
source-mac              : 1111.2222.3333
destination-mac         : 1111.4444.5555
vlan                    : 1
8021p                   : 5
source-ip               : 1.1.1.3
destination-ip          : 1.3.4.5
proto                   : udp
dscp                    : 3
ttl                     : 3
network-length          : 3
source-port             : 5223
destination-port        : 6553
payload-base            :
01234567890123456789012345678901234567890123456789012345678901234567890123456789
payload-extern-1        : 0123
payload-extern-2        : 0123
payload-extern-3        : 0123
payload-extern-4        : 0123
---------------------------------------
```

Contents

1 Configuring IFA 1
1.1 Introduction 1
1.1.1 Overview 1
1.1.2 Principles 1
1.2 Restrictions and Guidelines 2
1.3 Configuration Task Summary 2
1.4 Configuring Basic Functions 2
1.4.1 Overview 2
1.4.2 Restrictions and Guidelines 2
1.4.3 Procedure 3
1.5 Configuring Optional Features 4
1.5.1 Overview 4
1.5.2 Restrictions and Guidelines 4
1.5.3 Procedure 4
1.6 Monitoring 4
1.7 Configuration Examples 5
1.7.1 Configuring IFA sampling 5

1 Configuring IFA

Note

IFA is only supported by the S6980-64QC, S6580-48CQ8QC, or S6580-24DC8QC.

1.1 Introduction

1.1.1 Overview

In-band Flow Analyzer (IFA) can provide sampling for traffic visualization. IFA can be used to accurately determine the traffic forwarding path and latency, encapsulate the data into UDP packets, and send them to the customer's server for analysis.

1.1.2 Principles

1. Basic Concepts

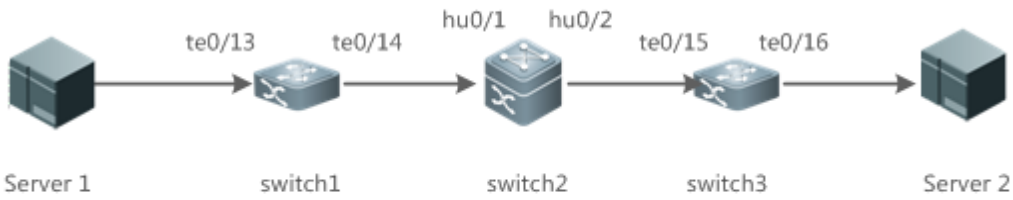
Traffic visualization utilizes network devices as nodes and device interfaces as basic sampling units. It obtains traffic forwarding path through hardware chip-based sampling and sends path information the server for traffic monitoring. This method can monitor traffic conditions in real-time, detect traffic anomalies in a timely manner, and ensure the normal and stable operation of the network. IFA sampling is to encapsulate traffic forwarding path and timestamp into UDP packets and send them to the server for analysis.

The first and last hops of IFA sampling can be selected. There can be more than one intermediate device. On the first -hop device, when a packet reaches the IFA interface and matches the IFA ACL, the device processes the packet according to the IFA configuration on the interface, including adding an INT probe header (containing probe marks, payload length, serial number, and so on) and an MD bit (containing device number, ingress and egress interfaces, and ingress and egress timestamps). By using the loopback interface, IFA enables the packets added with the INT probe header to be forwarded through the default original egress interfaces. The packets passing through the intermediate device and last-hop device are also added with the MD bit. The last-hop device encapsulates the IFA sampling information into UDP packets and send them to the IFA server for traffic analysis. Traffic forwarding path and latency can be identified based on the MD bit in IFA packets.

2. Basic Framework

As shown in Figure 1-1, traffic visualization solution can be used to monitor communication between server 1 and server 2. The devices between them are omitted and only three switches are reserved. The configuration of the omitted intermediate devices is the same as that of switch 2. Enable IFA on switch 1, the IFA first-hop device. The IFA packet is added with the MD bit on switch 1, switch 2, and switch 3 respectively, and sent by the last-hop device to the IFA server for traffic analysis.

Figure 1-1 Traffic Visualization Scheme Monitoring



- Configure an IFA ACL and sampling rate, and enable the interface to add the MD bit to the packet on switch 1.
- Configure the interface to add the MD bit on switch 2.
- Configure the server information and enable the interface to add the MD bit on switch 3.
- Configure MOD sampling on any switch to output packet drop statistics on the sampling interface.

1.2 Restrictions and Guidelines

- The IFA last-hop device can communicate with the server.
- Process the packets passing through the interface according to the sampling rate and send them to the server.

1.3 Configuration Task Summary

IFA configuration includes the following tasks:

(1) Configuring Basic Functions

(2) (Optional) Configuring Optional Features

1.4 Configuring Basic Functions

1.4.1 Overview

Configure the basic functions of IFA.

1.4.2 Restrictions and Guidelines

- The IFA last-hop device can communicate with the server.
- The IFA ACL and sampling rate can be configured on a physical or an aggregate interface. An interface can be configured with up to four sampling policies, each including an ACL and sampling rate.
- Only one sampling rate that is not equal to 1 is supported. For example, if an interface is configured with a sampling rate of 100, the sampling rate for other interfaces can only be 100 if not being 1.
- Run the **ifa set-header md** command in interface configuration command, and the specific interface will add the MD bit to IFA packet header. This command can be configured on a physical or an aggregate interface. After an interface is configured with this command, all interfaces cannot be configured with the **ifa passive md** command.
- Run the **ifa set-header md** command in global configuration command, and all interfaces will add the MD bit to IFA packet header.

- To disable this function on a specific interface, run the **ifa passive md** command on the interface.
- The MD bits added in interface configuration mode and global configuration mode are mutually exclusive. That is, the **ifa set-header md** command be configured in both interface configuration mode and global configuration mode.
- The **ifa set-header md** command must be configured on all non-last-hop devices that the IFA packets pass through included in a Layer 3 route. Otherwise, the IFA packets may not be re-encapsulated and sent to the server. Even if these packets are sent to the server, the packets may have incomplete information. Configure an expert-level ACL before running the **ifa set-header md** [**acl** *acl-id* | **acl** *acl-name*] command. The packets matching the ACL will be added with the MD bit in the header.
- The IFA version includes IFA 1.0 and IFA 2.0. IFA 1.0 are IFA 2.0 are not interchangeable. Running the **show ifa version** command can display IFA 2.0, but not IFA 1.0.
- The IFA 2.0 first-hop device does not need to be configured with the **ifa set-header md** and **ifa passive md** commands.
- Before using IFA, run the **switch-mode ifa** { **global** | **slot** *num* } command to switch the device to the IFA mode.
- The priority of the IFA ACL on each interface is unique.
- The **ifa set-header md** command takes effect on only incoming packets.
- The **ifa set-header md acl** command with the **acl** parameter is supported by only some products (chip-specific). Whether this command is supported is subject to the device.
- Applying both security ACL and IFA ACL on an interface may cause IFA to sample the packets dropped by the security ACL. A packet is dropped if it matches the security ACL, but it is sampled if it matches the IFA ACL. The security ACL and IFA ACL belong to different FP groups. So both of them can match the packet and non-conflicting actions can be implemented simultaneously. In this case, the packet dropped by the security ACL can still be sampled by IFA ACL. This packet is included in the IFA statistics.
- Configure the IFA ACL to allow the packet matching the deny rule to pass through and sample the packet matching the permit rule.
- The packets forwarded within the IFA domain are added with IFA information including IFA header, MD header, or MD bit according to IFA configuration. When the size of a packet exceeds the interface MTU, the packet is dropped. The first-hop device adds IFA information of up to 40 bytes to packets while the intermediate devices add the IFA information of up to 32 bytes to packets. The size of the packet with added IFA information should not exceed the interface MTU.

1.4.3 Procedure

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

configure terminal

(3) Enter the interface configuration mode.

interface *interface-type interface-number*

(4) Configure the IFA ACL and priority.

ifa acl { *acl-id* | *acl-name* } **rate** *rate* [**priority** *priority*]

(5) Configure the interface to add the MD bit to the IFA packet header.

ifa set-header md [**acl** *acl-id* | **acl** *acl-name*]

1.5 Configuring Optional Features

1.5.1 Overview

- Configure IFA packet filtering.
- Disable the interface to add the MD bit to IFA packet header.
- Configure the IPv4 address of the IFA device.
- Configure IFA statistics.

1.5.2 Restrictions and Guidelines

- This command can be used with the **ifa set-header md** command in global configuration mode to disable this function on a specific interface. This command is mutually exclusive with the **ifa set-header md** command in interface configuration mode.
- This command can be configured on a physical or an aggregate interface.
- The IP address of the IFA device must be authorized.

1.5.3 Procedure

(1) Enter the privileged EXEC mode.

enable

(2) Enter the global configuration mode.

configure terminal

(3) Enable IFA statistics.

ifa statistics enable

(4) Enter the interface configuration mode..

interface *interface-type interface-number*

(5) Configure the IP address or ID of the IFA device.

ifa set-device { *ipv4-address* | *id* }

(6) Configure IFA packet filtering.

ifa filter enable

(7) Disable the interface to add the MD bit to the IFA packet header.

ifa passive md

1.6 Monitoring

Run the **show** commands to check the running status of a configured function to verify the configuration effect.

Run the **clear** command to clear information.

Caution

Running the **clear** command may lose vital information and thus interrupt services.

Table 1-1 Monitoring

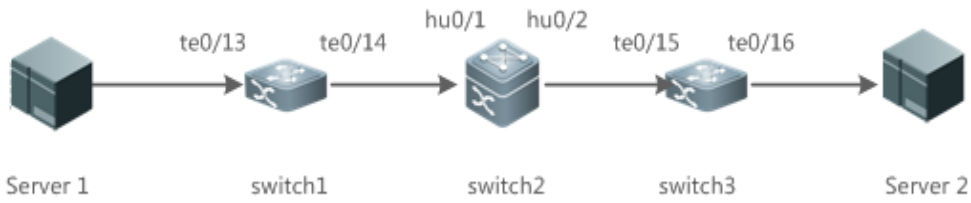
Command	Purpose
clear ifa statistics	clear ifa statistics
show ifa statistics	show ifa statistics.
show ifa version	show ifa version.

1.7 Configuration Examples

1.7.1 Configuring IFA sampling

1. Topology

Figure 1-2 Topology



2. Note

- On switch 1, enable IFA on the interface included in the route, and add the MD bit to the IFA packet header.
- On switch 2, configure the interface included in the route to add the MD bit to the IFA packet header.
- On switch 3, configure the server information and enable the interface included in the route to add the MD bit to the IFA packet header.

3. Procedure

The following example enables IFA on switch 1.

```
Switch1# configure terminal
Switch1(config)# access-list 100 permit ip any any
Switch1(config)# interface TenGigabitEthernet 0/13
Switch1(config-if-TenGigabitEthernet 0/13)# ifa acl 100 rate 100 priority 10
Switch1(config-if-TenGigabitEthernet 0/13)# ifa set-header md
Switch1(config-if-TenGigabitEthernet 0/13)# end
```

The following example enables IFA on switch 2.

```
Switch2# configure terminal
Switch2(config)# interface HundredGigabitEthernet0/1
Switch2(config-if-HundredGigabitEthernet0/1)# ifa set-header md
Switch2(config-if-HundredGigabitEthernet0/1)# end
```

The following example enables IFA on switch 3.

```
 Switch3# configure terminal
Switch3(config)# ifa statistics enable
Switch3(config)# interface TenGigabitEthernet 0/15
Switch3(config-if-TenGigabitEthernet 0/15)# ifa set-header md
Switch3(config-if-TenGigabitEthernet 0/15)# end
```

4. Verification

Run the **show run** command to check whether the command output is consistent with the configuration.

Steer the packet matching the ACL to a path from interface TenGigabitEthernet0/13 on switch 1 toward the proceeding interfaces. Check whether the server receives the IFA packets sent from switch 3 and whether sampling statistics are correct.

Run the **show ifa statistics** command to display statistics.

```
Ruijie# show ifa statistics
Type          Acl    Statistic(pkts)        Applied interface
------------------------------------------------------------------------------
--------------
sample         100    20426                Te0/13-14
sample         101    0                   TF0/2
insert-md       -      20586                Hu0/1-2, Te0/13-16
```

5. Configuration File

- Configuration on switch 1

```
access-list 100 permit ip any any

interface TenGigabitEthernet 0/13
ifa acl 100 rate 100 priority 10
ifa set-header md
```

- Configuration on switch 2

```
interface HundredGigabitEthernet0/1
ifa set-header md
```

- Configuration on switch 3

```
ifa statistics enable

interface TenGigabitEthernet 0/15
ifa set-header md
```

Data Center Configuration

1. Configuring VXLAN
2. Configuring Service Chain
3. Configuring PFC

Contents

1 Configuring VXLAN 1
1.1 Overview 1
1.2 Principles 2
1.2.1 Basic Concepts 2
1.2.2 Packet Format 4
1.2.3 Introduction to the VXLAN Operation Mechanism 6
1.2.4 Establishment of a VXLAN Tunnel 6
1.2.5 Identifying the VXLAN to Which a Packet Belongs 7
1.2.6 MAC Address Learning 7
1.2.7 Forwarding Model 8
1.2.8 Forwarding Process 11
1.2.9 VXLAN Routing Deployment Model 24
1.2.10 Anycast Gateway 26
1.2.11 ARP/ND Flood Suppression 26
1.3 Protocols and Standards 27
2 Configuring VXLAN Basic Features 27
2.1 Overview 27
2.1.1 Centralized VXLAN Networking Using Static Tunnels 28
2.1.2 Centralized VXLAN Networking Using Dynamic Tunnels 28
2.1.3 Distributed Asymmetric VXLAN Networking 28
2.1.4 Distributed Symmetric VXLAN Networking 29
2.2 Restrictions and Guidelines 29

2.2.1 Centralized VXLAN Networking Using Static Tunnels 29
2.2.2 Centralized VXLAN Networking Using Dynamic Tunnels 29
2.2.3 Distributed Asymmetric VXLAN Networking 29
2.2.4 Distributed Symmetric VXLAN Networking 29
2.3 Configuration Task Summary 30
2.3.1 Centralized VXLAN Networking Using Static Tunnels 30
2.3.2 Centralized VXLAN Networking Using Dynamic Tunnels 30
2.3.3 Distributed Asymmetric VXLAN Networking 30
2.3.4 Distributed Symmetric VXLAN Networking 30
2.4 Configuring a VXLAN Routed Interface 31
2.4.1 Overview 31
2.4.2 Restrictions and Guidelines 31
2.4.3 Procedure 31
2.5 Configuring a Static Tunnel 31
2.5.1 Overview 31
2.5.2 Restrictions and Guidelines 31
2.5.3 Prerequisites 32
2.5.4 Procedure 32
2.6 Configuring a VXLAN Instance 32
2.6.1 Overview 32
2.6.2 Restrictions and Guidelines 32
2.6.3 Procedure 32
2.7 Configuring a VTEP 33
2.7.1 Overview 33

2.7.2 Restrictions and Guidelines 33
2.7.3 Procedure 34
2.8 Configuring an EVPN 34
2.8.1 Overview 34
2.8.2 Restrictions and Guidelines 34
2.8.3 Configuration Tasks 35
2.8.4 Configuring a BGP EVPN 35
2.8.5 Configuring an EVI 35
2.9 Enabling ARP/ND Automatic Learning 36
2.9.1 Overview 36
2.9.2 Restrictions and Guidelines 36
2.9.3 Procedure 36
2.10 Configuring an Anycast Gateway 37
2.10.1 Overview 37
2.10.2 Restrictions and Guidelines 37
2.10.3 Procedure 37
2.11 Enabling the Anycast Gateway Synchronization Function 38
2.11.1 Restrictions and Guidelines 38
2.11.2 Prerequisites 38
2.11.3 Procedure 38
2.12 Configuring a Symmetric VXLAN 39
2.12.1 Overview 39
2.12.2 Restrictions and Guidelines 39
2.12.3 Prerequisites 39

2.12.4 Procedure....39
3 Configuring VXLAN Extended Functions....40
3.1 Overview....40
3.2 Restrictions and Guidelines....40
3.3 Configuration Task Summary....40
3.4 Enabling Storm Control....41
3.4.1 Overview....41
3.4.2 Restrictions and Guidelines....41
3.4.3 Procedure....41
3.5 Configuring the VXLAN UDP Destination Port....41
3.5.1 Overview....41
3.5.2 Restrictions and Guidelines....41
3.5.3 Procedure....41
3.6 Configuring a Static VXLAN MAC Address Table....42
3.6.1 Overview....42
3.6.2 Procedure....42
3.7 Configuring Multi-Tenant Layer 3 Packet Isolation....42
3.7.1 Overview....42
3.7.2 Restrictions and Guidelines....43
3.7.3 Procedure....43
3.8 Configuring the EVPN Synchronization Attribute....43
3.8.1 Overview....43
3.8.2 Restrictions and Guidelines....43
3.8.3 Procedure....45

3.9 Configuring Rate Limits for a Tunnel Interface....45
3.9.1 Overview....45
3.9.2 Restrictions and Guidelines....46
3.9.3 Procedure....46
3.10 Configuring ARP/ND Suppression....46
3.10.1 Overview....46
3.10.2 Procedure....46
3.11 Configuring ARP/ND Proxy....47
3.11.1 Overview....47
3.11.2 Restrictions and Guidelines....47
3.11.3 Prerequisites....47
3.11.4 Procedure....47
3.12 Creating VXLAN Instances in Batches....48
3.12.1 Overview....48
3.12.2 Restrictions and Guidelines....48
3.12.3 Procedure....48
3.13 Configuring a Data Center Interconnection Tunnel....48
3.13.1 Overview....48
3.13.2 Restrictions and Guidelines....49
3.13.3 Procedure....49
3.14 Configuring VNI Mapping....49
3.14.1 Overview....49
3.14.2 Restrictions and Guidelines....49
3.14.3 Prerequisites....50

3.14.4 Procedure....50
3.15 Configuring a Layer 2 Ethernet Sub-interface to Access a VXLAN....51
3.15.1 Overview....51
3.15.2 Restrictions and Guidelines....51
3.15.3 Procedure....51
3.16 Configuring a Delay in the Protocol Status Change of a VXLAN Tunnel....52
3.16.1 Overview....52
3.16.2 Procedure....52
3.17 Configuring the Optimal Minimum Route Length for VXLAN Tunnels....52
3.17.1 Overview....52
3.17.2 Procedure....52
4 Monitoring....53
5 Configuration Examples....55
5.1 Configuring a VXLAN SDN....55
5.1.1 Requirements....55
5.1.2 Topology....56
5.1.3 Notes....56
5.1.4 Procedure....57
5.1.5 Verification....60
5.1.6 Configuration Files....61
5.2 Configuring a Scenario of EVPN-based Multi-Tenant Centralized VXLAN Networking....64
5.2.1 Requirements....64
5.2.2 Topology....64
5.2.3 Notes....65

5.2.4 Procedure....65
5.2.5 Verification....68
5.2.6 Configuration Files....69
5.3 Configuring a Scenario of EVPN-based Multi-Tenant Centralized Anycast Gateway Deployment....73
5.3.1 Requirements....73
5.3.2 Topology....73
5.3.3 Notes....73
5.3.4 Procedure....74
5.3.5 Verification....79
5.3.6 Configuration Files....81
5.4 Configuring a Scenario of EVPN-based Multi-Tenant Distributed VXLAN Networking (Asymmetric Deployment)....87
5.4.1 Requirements....87
5.4.2 Topology....87
5.4.3 Notes....87
5.4.4 Procedure....88
5.4.5 Verification....91
5.4.6 Configuration Files....93
5.4.7 Common Errors....96
5.5 Configuring a Scenario of EVPN-based Multi-Tenant Distributed VXLAN Networking (Symmetric Deployment)....96
5.5.1 Requirements....96
5.5.2 Topology....97
5.5.3 Notes....97

5.5.4 Procedure....98
5.5.5 Verification....103
5.5.6 Configuration Files....104
5.5.7 Common Errors....109
5.6 Configuring a Scenario of EVPN-based Single-Tenant VXLAN Routing....109
5.6.1 Requirements....109
5.6.2 Topology....110
5.6.3 Notes....110
5.6.4 Procedure....111
5.6.5 Verification....117
5.6.6 Configuration Files....117
5.7 Configuring a Scenario of EVPN-based Multi-Tenant VXLAN Routing....124
5.7.1 Requirements....124
5.7.2 Topology....125
5.7.3 Notes....125
5.7.4 Procedure....126
5.7.5 Verification....134
5.7.6 Configuration Files....135
5.8 Configuring an EVPN Distributed Network to Support Non-EVPN VTEPs....145
5.8.1 Requirements....145
5.8.2 Topology....146
5.8.3 Notes....146
5.8.4 Procedure....147
5.8.5 Verification....149

5.8.6 Configuration Files 150
5.9 Configuring Non-EVPN VTEPs to Connect to an EVPN Distributed Network in Full-Mesh Pattern 153
5.9.1 Requirements 153
5.9.2 Topology 154
5.9.3 Notes 154
5.9.4 Procedure 155
5.9.5 Verification 158
5.9.6 Configuration Files 158
5.10 Configuring a Scenario of Data Center Interconnection Based on VNI Mappings 162
5.10.1 Requirements 162
5.10.2 Topology 162
5.10.3 Notes 162
5.10.4 Procedure 163
5.10.5 Verification 166
5.10.6 Configuration Files 167
5.11 Configuring a Layer 2 Ethernet Sub-interface to Access a VXLAN 171
5.11.1 Requirements 171
5.11.2 Topology 172
5.11.3 Notes 172
5.11.4 Procedure 172
5.11.5 Verification 173
5.11.6 Configuration Files 174
5.12 Configuring a VXLAN in an M-LAG Scenario 175
5.12.1 Requirements 175

5.12.2 Topology 175

5.12.3 Notes 175

5.12.4 Procedure 176

5.12.5 Verification 178

5.12.6 Configuration Files 179

5.12.7 Common Errors 182

1 Configuring VXLAN

Note

VXLAN is supported by only the S6510-32CQ, S6510-48VS8CQ, S6510-4C, and S6250-48XS8CQ.

1.1 Overview

Virtual extensible local area network (VXLAN) is a virtual Ethernet built over a physical IP network. It is an Layer 2 over User Datagram Protocol (UDP) technology.

The development of network and virtualization technologies brings widespread application of server virtualization in data centers. Deploying virtual machines (VMs) on servers in a data center improves server utilization and reduces hardware costs and operation costs in the data center.

However, the introduction of the virtualization technology poses new challenges to conventional data center networks.

- Layer 2 network boundary limit

  With application of the virtualization technology, many VMs may run on each physical server. The number of VM MAC addresses in an Layer 2 network may increase sharply. To properly forward data packets, an access switch needs to learn the MAC addresses of VMs in each Layer 2 network. Limited by hardware, the capacity of the MAC address table on an access switch is small and cannot meet requirements.

- Inadequate network isolation capability

  In a conventional data center network, the traffic of VMs is isolated using the virtual local area network (VLAN) technology. The VMs of different tenants are classified into different VLANs. The VLAN technology distinguishes VLANs by VLAN ID and the actual number of supported VLAN IDs is 4094, which is insufficient for some large-scale data centers. In addition, conventional VLANs cannot meet the requirement of dynamic network adjustment, and network expansion and adjustment are very difficult.

- VM migration limit

  VMs need to be migrated to other servers because of insufficient server hardware resources (for example, very high CPU usage, insufficient disk space, and insufficient memory) or the relocating of the equipment room. The MAC and IP addresses of the VMs must remain unchanged before and after migration, to ensure that services are not interrupted during VM migration. Therefore, VMs can be migrated only in the current Layer 2 network.

The VXLAN technology effectively solves the problems above.

- Breaking the Layer 2 network boundary limit to solve the MAC address capacity problem

  VXLAN uses UDP to encapsulate data packets sent by VMs. The outer IP or MAC headers of UDP packets use the IP or MAC address of the physical network. Therefore, except edge devices, devices in a VXLAN do not need to identify the MAC addresses of VMs, which reduces the MAC address learning pressure of the devices.

- Solving the VLAN shortage problem

VXLAN uses a 24-bit VXLAN network ID (VNI) akin to a VLAN ID for identification. It can implement isolation of up to 16 million VXLAN segments and accommodate mass tenants. With VXLAN, user isolation and identification are no longer limited.

- Removing the VM migration limit

  VXLAN uses MAC in UDP encapsulation to extend an Layer 2 network and decouple virtual networks from a physical network. VXLAN allows tenants to plan virtual networks without considering restrictions in IP addresses and broadcast domains of physical networks. It reduces network management difficulties. Building a large Layer 2 network through VXLAN can ensure that the IP addresses, MAC addresses, and other parameters of VMs keep unchanged during VM migration.

1.2 Principles

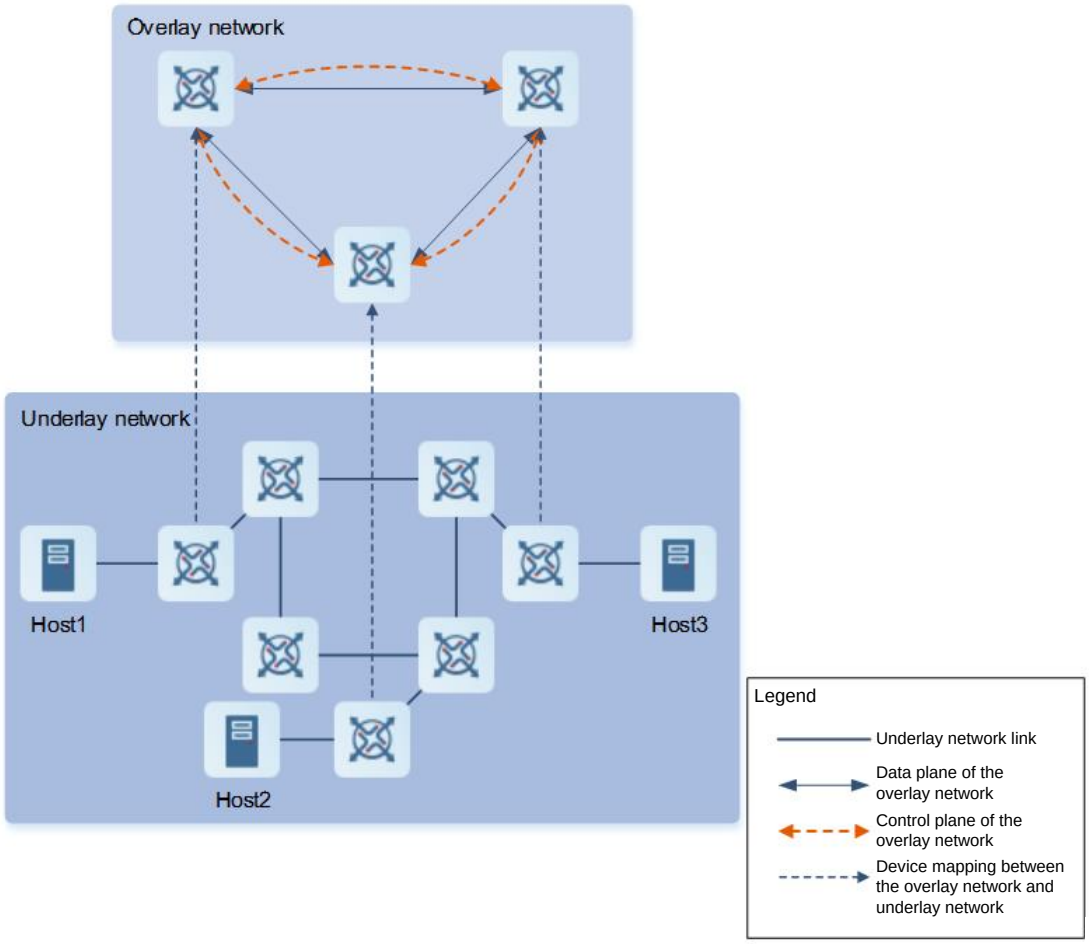
1.2.1 Basic Concepts

1. Overlay Network and Underlay Network

The VXLAN technology can be used to build an overlay logical network on the basis of a physical network (such as an IP network).

- The bearer network (underlay network) and virtual network (overlay network) are completely separated and do not perceive each other.
- The underlay network needs to be simple and easy to manage, and can provide stable and reliable forwarding capability for the overlay network.
- The overlay network needs to provide flexibility and scalability.

Figure 1-1 Overlay Network and Underlay Network



2. VTEP

A VXLAN tunnel end point (VTEP) is used to encapsulate and decapsulate VXLAN packets. At the start point of a VXLAN tunnel, a VTEP encapsulates data generated by VMs into UDP headers and sends out the packets. After encapsulation, the MAC addresses and VLAN information of the VMs are no longer used the basis for data forwarding. At the end point of the VXLAN tunnel, a VTEP decapsulates received VXLAN packets.

3. VNI

VNIs identify different VXLAN segments. Only VMs with the same VNI in the same VXLAN segment can communicate with each other through VXLAN Layer 2 forwarding. Otherwise, VMs communicate with each other through VXLAN Layer 3 forwarding. A VNI is represented by a 24-bit binary number.

4. VXLAN Tunnel

A VXLAN tunnel is a logical tunnel for transmitting encapsulated VXLAN packets between two VTEPs. When a service packet reaches a VXLAN tunnel, a VXLAN header, UDP header, and IP header are added to the front of the service packet. The encapsulated packet is then transparently transmitted to the remote VTEP by the Layer 3 forwarding function of the underlay network. After receiving the service packet, the remote VTEP decapsulates the packet.

5. VXLAN Gateway

A VXLAN gateway routes VXLAN packets based on IP addresses in the packets. It serves as the gateway of VMs in a VXLAN network. After configuring a VXLAN gateway, you can build an Layer 3 overlay network to implement communication between VXLAN subnets.

6. VXLAN Instance

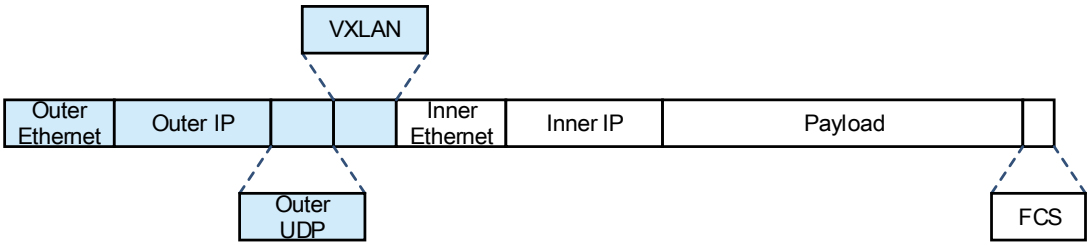
A VXLAN instance is a node in a VXLAN-based overlay network. This node can implement conventional Layer 2 switching functions such as MAC address learning and flooding in an overlay network. After being specified as a VXLAN gateway, this node can also perform Layer 3 forwarding. On the same device, VXLAN instances and VNIs are in one-to-one mappings.

1.2.2 Packet Format

After receiving a data frame from a VM, a VTEP encapsulates four elements to form a new frame header. The original destination/source MAC address, inner VLAN tag, and Ethernet type carried in the inner data frame remain unchanged. The four elements from inside out are the VXLAN header, outer UDP header, outer IPv4 header, and outer Ethernet frame header.

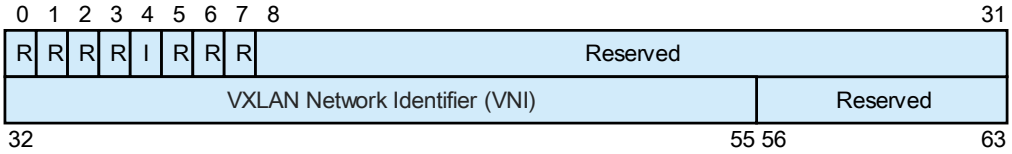
The figure below shows the frame format after VXLAN encapsulation.

Figure 1-2 Frame Format After VXLAN Encapsulation



- VXLAN header

Figure 1-3 VXLAN Header

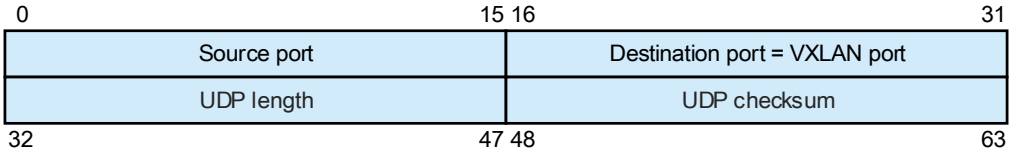


A VXLAN header has 64 bits. In the design of the current protocol version, the sole purpose of a VXLAN header is to carry the 24-bit VNI assigned by a VTEP.

- **Flag** (8 bits): The I bit must be set to 1 to indicate a valid VNI, and the R bits must be set to 0.
- **VXLAN Segment ID/VXLAN Network Identifier (VNI)**: It occupies 24 bits to identify a VXLAN. Only VMs belonging to the same VXLAN can communicate with each other.
- **Reserved**: There are two reserved fields, which occupy 24 bits and 8 bits separately and are filled with zeros.

- Outer UDP header

Figure 1-4 Outer UDP Header

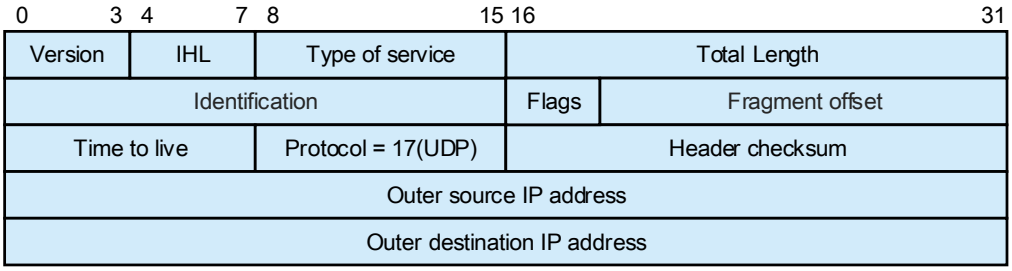


The fields in an outer UDP header are described as follows:

- **Source port**: Indicates the UDP source port ID. Assigned by a VTEP, it is the result of the hash operation performed on the Layer 2 header of a data frame. This hash result can serve as the basis for traffic load balancing.
- **Destination port = VXLAN port**: Indicates a destination port ID. The destination port ID assigned by the Internet Assigned Numbers Authority (IANA) is 4789.
- **UDP length**: Indicates the length of a UDP header.
- **UDP checksum**: Indicates the UDP checksum, which is set to **0** for transmission.

- Outer IP header

Figure 1-5 Outer IP Header



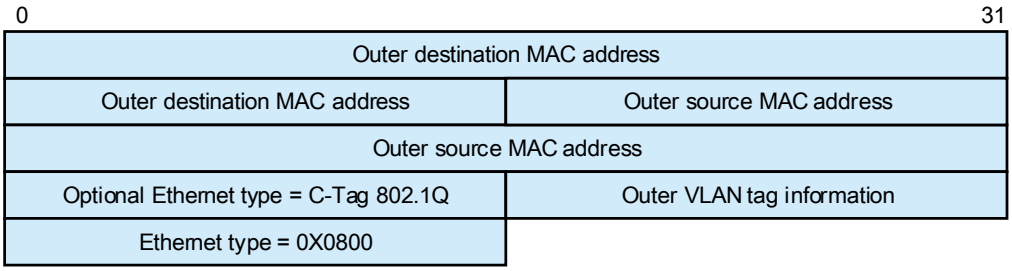
The fields in an outer IP header are described as follows:

- **Outer source IPv4 address**: Indicates the IP address of a VTEP that forwards packets of VMs.
- **Outer destination IPv4 address**: It can be a unicast IP address or multicast IP address. If it is a unicast IP address, the field indicates the IP address of a VTEP corresponding to a VM that needs to make communication.

The IP addresses in an outer IP header are no longer the addresses of VMs that make communication but the IP addresses of VTEPs at both ends of a tunnel. If the hypervisor on a server, on which a VM runs, directly undertakes the work of a VTEP, the IP address here is that of the network interface card (NIC) of the server that runs the VM software. If a VTEP is a switch, the IP address is that of an outbound interface or an Layer 3 switch virtual interface (SVI).

- Outer Ethernet header

Figure 1-6 Outer Ethernet Header



The fields in an outer Ethernet header are described as follows:

- **Outer destination MAC address**: Indicates the next-hop MAC address pointed to the IP address of the destination VTEP.
- **Outer source MAC address**: Indicates the MAC address of the local VTEP.
- **Outer VLAN tag information**: Optional.

1.2.3 Introduction to the VXLAN Operation Mechanism

The VXLAN operation consists of the following parts:

(1) Create a VXLAN instance on a device.

(2) Establish a tunnel between VTEPs and associate the tunnel with the VXLAN instance.

(3) VTEPs learn MAC addresses for data forwarding.

(4) A VTEP determines the VXLAN, to which a received MAC frame belongs, encapsulates the packet into a VXLAN packet, and selects a proper tunnel to forward the packet based on the MAC address learning result. The peer VTEP receives the packet from the tunnel, and decapsulates and forwards the packet.

(5) The VXLAN gateway routes and forwards packets across VXLAN subnets at Layer 3.

1.2.4 Establishment of a VXLAN Tunnel

VXLAN packets are transmitted between two VTEPs through a VXLAN tunnel. Service packets are encapsulated upon entering a tunnel. The Layer 3 forwarding module of the underlay network transmits the encapsulated packets to the remote VTEP, which then decapsulates the packets and sends them out of the tunnel. Therefore, a VXLAN tunnel needs to be established between VTEPs. The established VXLAN tunnel needs to associate with a VXLAN instance so that the tunnel transmits packets of the VXLAN instance. A VXLAN tunnel can be created manually or automatically:

- Manual creation

  Manually configure VXLAN tunnel interfaces and specify the source IP address and destination IP address of the tunnel. A manually created tunnel needs to be manually associated with a VXLAN instance.

- Automatic creation

  An Ethernet virtual private network (EVPN) is used as the control plane of a VXLAN. A VTEP needs to establish a Border Gateway Protocol (BGP) neighbor relationship with another VTEP or an Interior Border Gateway Protocol (IBGP) route reflector first. Then, the VTEP sends an EVPN type-3 route (containing the multicast Ethernet tag) to its neighbor that supports BGP-EVPN. VTEPs send EVPN type-3 routes to each

other to discover VTEPs and establish VXLAN tunnels. The local IP address of a tunnel is IP address of the local VTEP and the peer IP address is the peer VTEP IP address learned from a type-3 route.

VXLAN tunnels automatically established through EVPN are automatically associated with VXLAN instances.

1.2.5 Identifying the VXLAN to Which a Packet Belongs

1. Identifying the VXLAN to Which an Access Packet Belongs

The device can identify the VXLAN, to which an access packet belongs, in two ways:

- Configuring a VLAN to be associated with a VXLAN instance: When the device receives a MAC frame from a VLAN, which is associated with a VXLAN, the device encapsulates the received MAC frame into a VXLAN packet and forwards it through the corresponding VXLAN tunnel. On the same access device, different VXLAN instances cannot be associated with the same VLAN. After a VLAN is associated with a VXLAN, all packets of the VLAN are encapsulated into VXLAN packets. Therefore, no SVI can be created for the VLAN to serve as an IP gateway.
- Configuring a Layer 2 Ethernet sub-interface to access a specified VXLAN: After a Layer 2 Ethernet sub-interface receives a MAC frame, if the sub-interface has been configured to access a VXLAN network, the packet belongs to the VXLAN network. The device encapsulates the MAC frame into a VXLAN packet for forwarding. Layer 2 Ethernet sub-interfaces can access a VXLAN in untagged VLAN encapsulation mode, tagged VLAN encapsulation mode, and QinQ encapsulation mode. After a sub-interface is available, the encapsulation rule of the sub-interface has a higher priority than that of the VLAN associated with a VXLAN instance.

2. Identifying the VXLAN, to Which Packets Received from a VXLAN Tunnel Belongs

The header of a VXLAN packet carries VNI information. After receiving a packet from a VXLAN tunnel, a VTEP can determine the VXLAN network, to which the packet belongs based on the VNI carried in the packet.

1.2.6 MAC Address Learning

A VTEP needs to learn user MAC addresses to forward VXLAN packets. The MAC address learning of a VTEP includes local MAC address learning and remote MAC address learning. You can also manually configure MAC address entries and specify the VLANs, interfaces, and VNIs for the MAC address entries. Binding static MAC address entries and static ARP entries manually is more stable and secure than dynamic MAC address learning.

1. Local MAC Address Learning

Local MAC address learning is triggered by actions on the forwarding plane. After receiving a data frame from a local host, a VTEP determines the VXLAN instance, to which the data frame belongs.. If it is certain that the data frame belongs to a VXLAN instance, the VTEP adds the source MAC address of the data frame to the MAC address table of the VXLAN instance and records the VNI and the interface that receives the data frame.

2. Remote MAC Address Learning

Remote MAC address learning varies with the VXLAN tunnel establishment mode.

A VTEP can learn the MAC address of an Layer 2 broadcast domain where the peer VTEP resides in either of the following ways:

- Source MAC address learning: VXLAN packets are transmitted between VTEPs through a VXLAN tunnel.

After receiving a VXLAN packet from a VXLAN tunnel, a VTEP determines the VXLAN instance, to which the packet belongs, based on the VNI. Then, the VTEP decapsulates the packet, adds the source MAC address in the data frame to the MAC address table of the VXLAN instance, and records the VNI and the VXLAN tunnel interface that receives the packet.

- BGP EVPN type-2 routing packets: After EVPN is imported, a VTEP can transmit hosts' MAC addresses by advertising the type-2 MAC/IP advertisement routes of EVPN, so as to reduce traffic flooding in the VXLAN. After learning a host's MAC address, a VTEP can advertise a MAC-only type-2 route to advertise the MAC entry to other VTEP neighbors. If the VTEP learns a host's ARP entry, it advertises a MAC-IP type-2 route to advertise the host's MAC address to its VTEP neighbors.

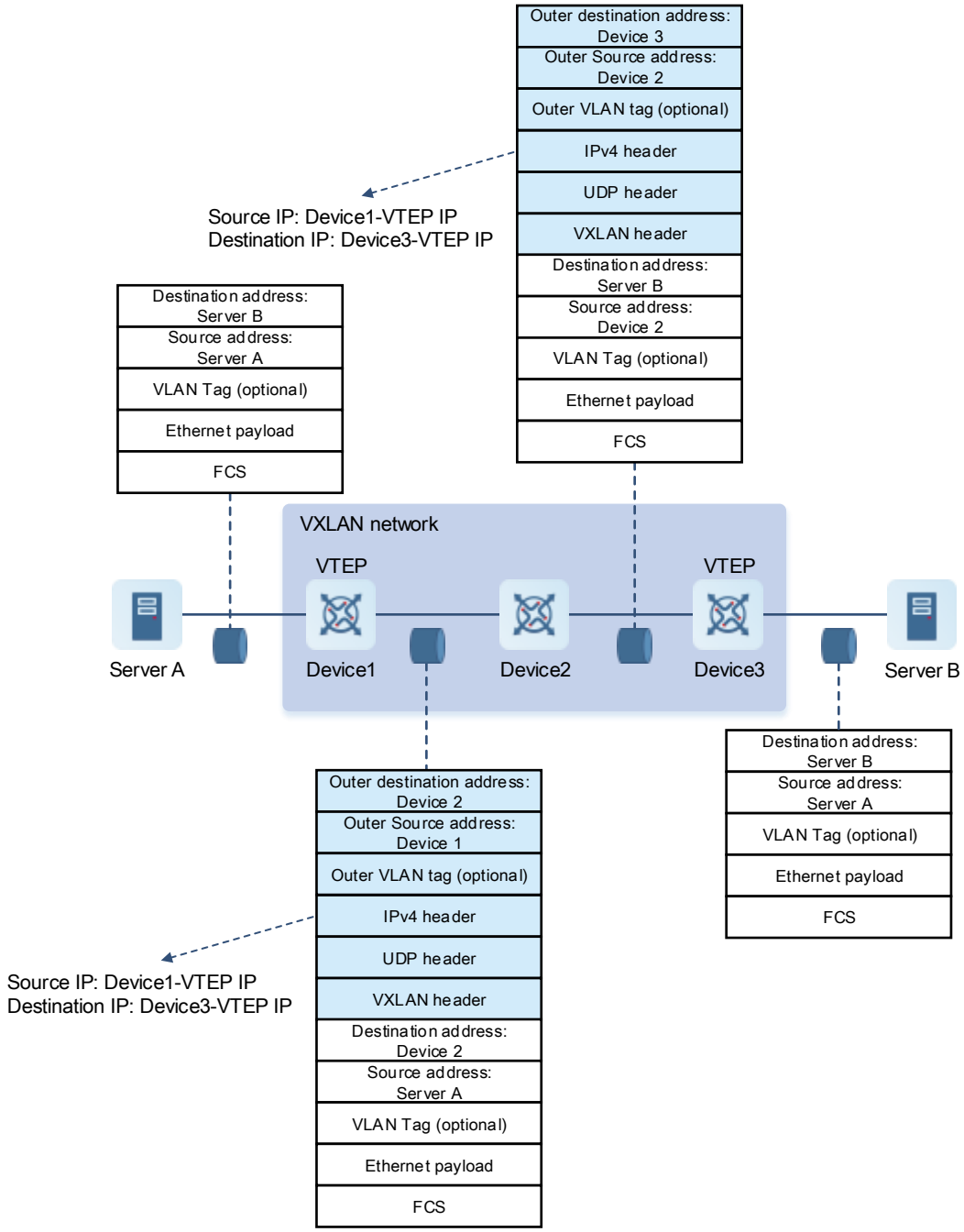
1.2.7 Forwarding Model

VXLAN forwarding supports bridging mode and routing mode. User communication in the same network segment is forwarded in VXLAN bridging mode, in which hosts using the same VNI communicate with each other through VXLAN Layer 2 forwarding. User communication in different network segments is forwarded in VXLAN routing mode, in which hosts using different VNIs communicate with each other through VXLAN routing and forwarding. A VXLAN gateway is needed for the communication between different VNIs.

1. VXLAN Bridging Principle

After determining the VXLAN, to which a MAC frame belongs, a VTEP encapsulates the MAC frame into a VXLAN packet. VXLAN packets are carried over UDP packets. After IP headers are added, the VXLAN packets are transmitted over the IP network. At the receive end, a VTEP decapsulates VXLAN packets to retrieve MAC frames and then forwards the MAC frames. See the following figures.

Figure 1-7 VXLAN Bridging



(1) Device 1 (switch) receives an Ethernet packet and encapsulates it into a VXLAN packet.

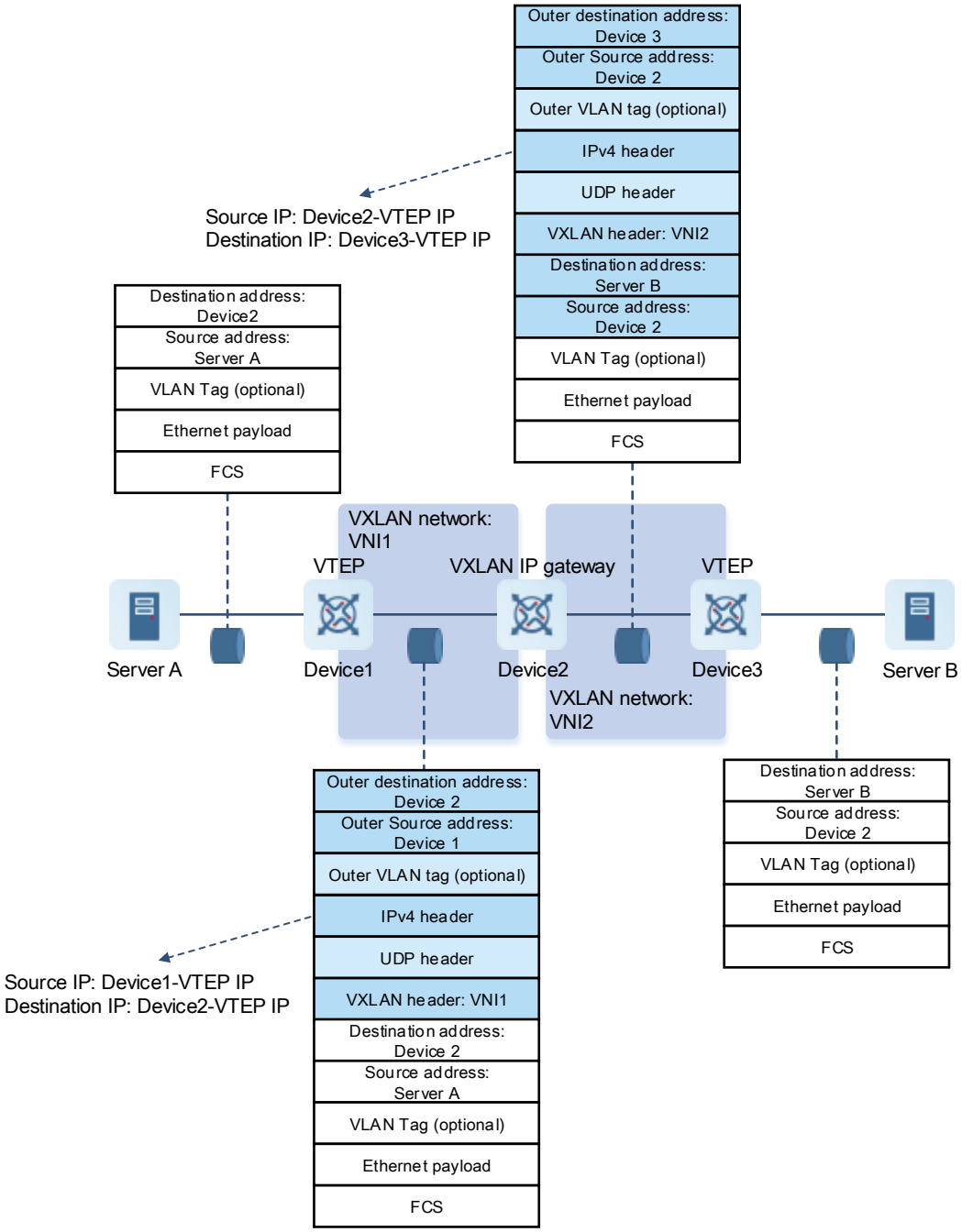
(2) The VXLAN packet is forwarded over the IP core network. As shown in the figure above, device 2 forwards the VXLAN packet.

(3) After receiving the VXLAN packet, device 3 decapsulates the packet, and forwards it at Layer 2 in the LAN.

2. VXLAN Routing Principle

Cross-VXLAN communication is implemented through a VXLAN gateway. See the following figures.

Figure 1-8 VXLAN Routing



(1) To make cross-VXLAN communication, server A first sends a packet to the IP gateway, which is deployed on device 2.

(2) The packet sent by server A is encapsulated by device 1 into a VXLAN packet and then sent to device 2.

(3) After receiving the VXLAN packet, device 2 finds that the inner destination MAC address is not its MAC address and then sends the packet to device 3 after VXLAN routing.

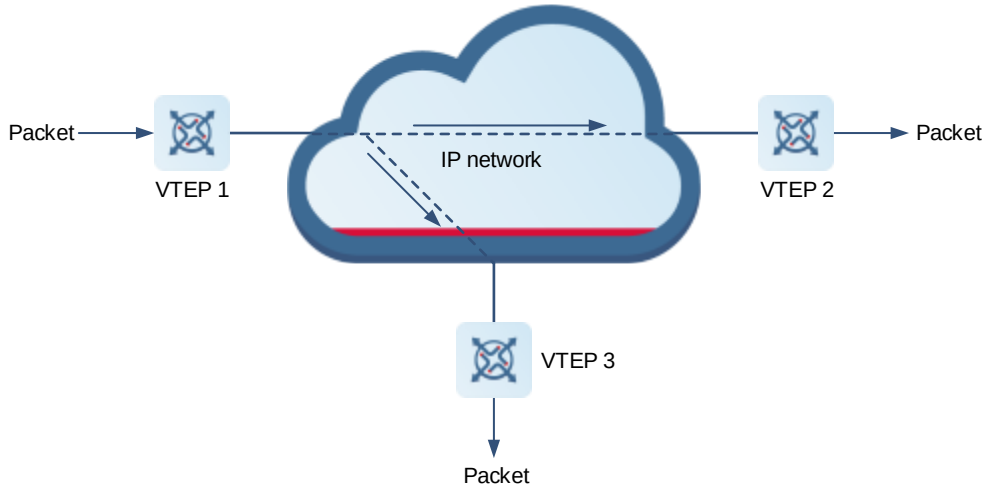
(4) After receiving the packet from device 2, device 3 decapsulates the packet and sends it to server B.

1.2.8 Forwarding Process

1. Packet Flooding Using VXLAN Tunnel Header Replication

A VXLAN uses the tunnel header replication mode to flood broadcast, multicast, and unknown unicast packets. After receiving an Address Resolution Protocol (ARP) request, VTEP 1 encapsulates the ARP request packet into two unicast packets, and forwards the packets to VTEP 2 and VTEP 3, as shown in Figure 1-9.

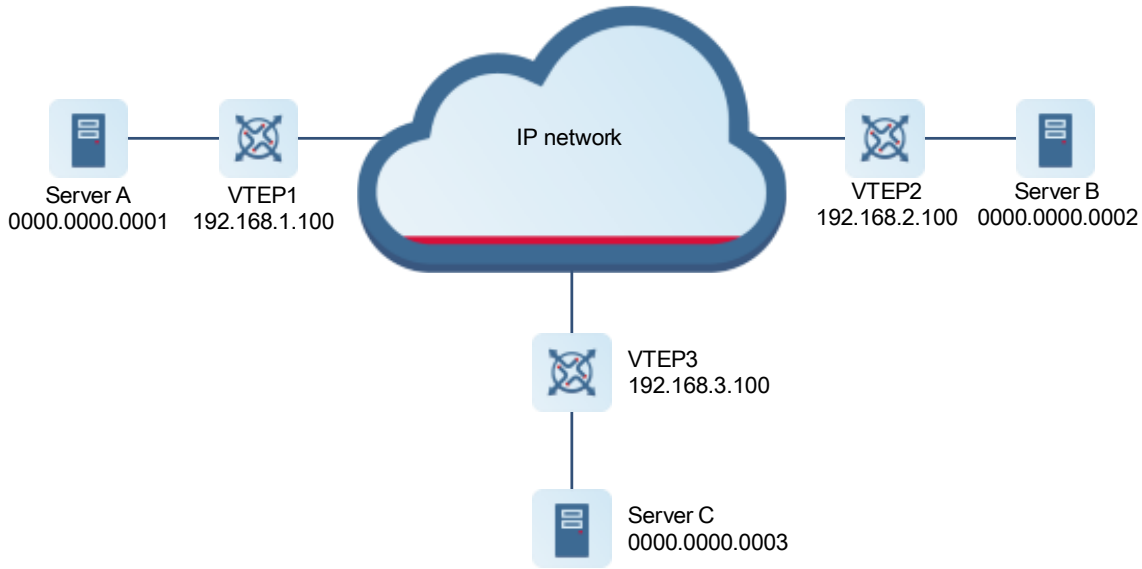
Figure 1-9 Packet Flooding Using VXLAN Tunnel Header Replication



2. VXLAN Layer 2 Forwarding

As shown in Figure 1-10, three servers interconnect to each other at Layer 2 through VXLAN carried over an IP network, and the VXLAN VNI is 100.

Figure 1-10 Topology of VXLAN Layer 2 Forwarding

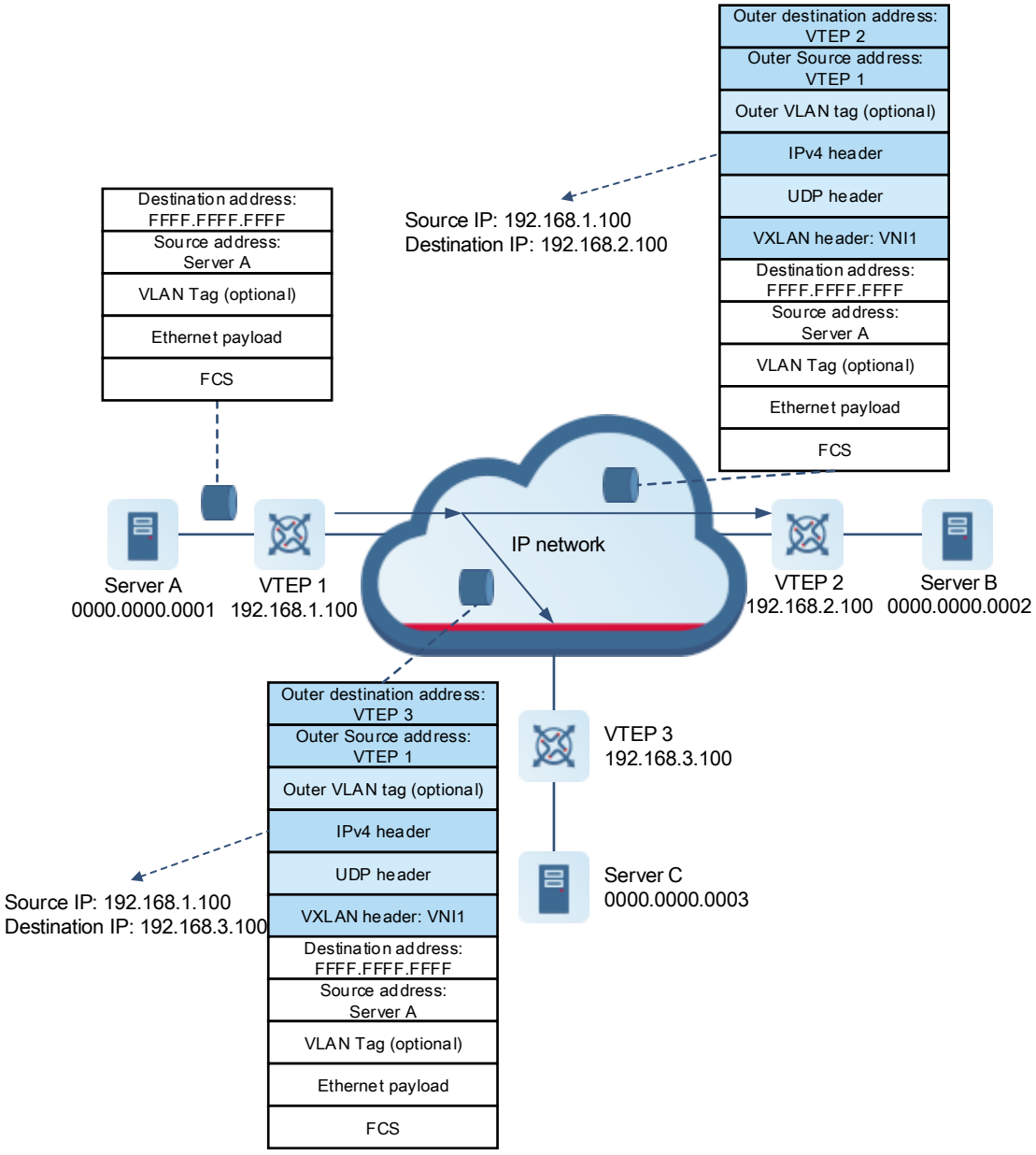


The following uses the process, in which server A sends an ARP request to server B and server B returns a ARP reply, as an example to illustrate the VXLAN packet forwarding.

Note

The **Next Hop MAC** and **Last Hop MAC** fields mentioned in this chapter refer to the next-hop and last-hop MAC addresses transmitted on the Underlay network. The values of the **Next Hop MAC** and **Last Hop MAC** fields vary with legends.

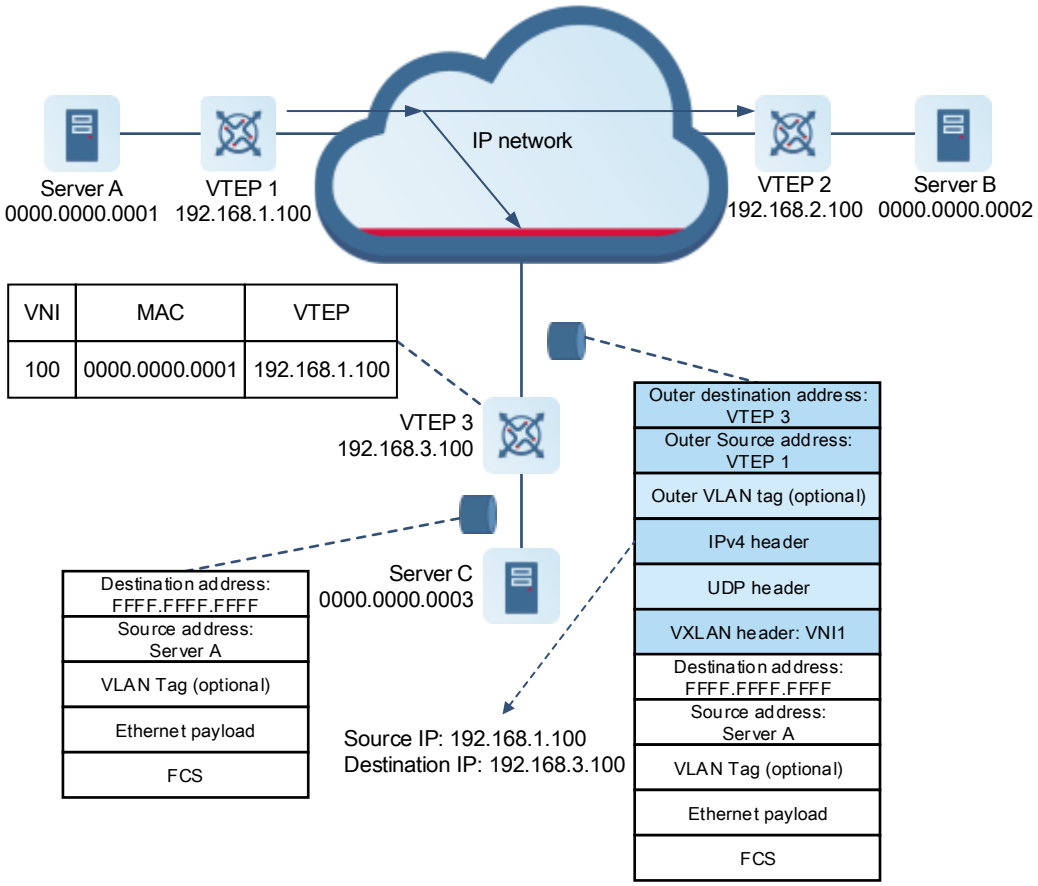
Figure 1-11 Sending an ARP Request from Server A to Server B Through a Static VXLAN Tunnel



(1) Server A sends an ARP request, which is a broadcast packet. After receiving the ARP request, VTEP 1 (switch) floods the broadcast packet in tunnel header replication mode, encapsulates it into two unicast packets, and sends them to VTEP 2 and VTEP 3 through tunnels.

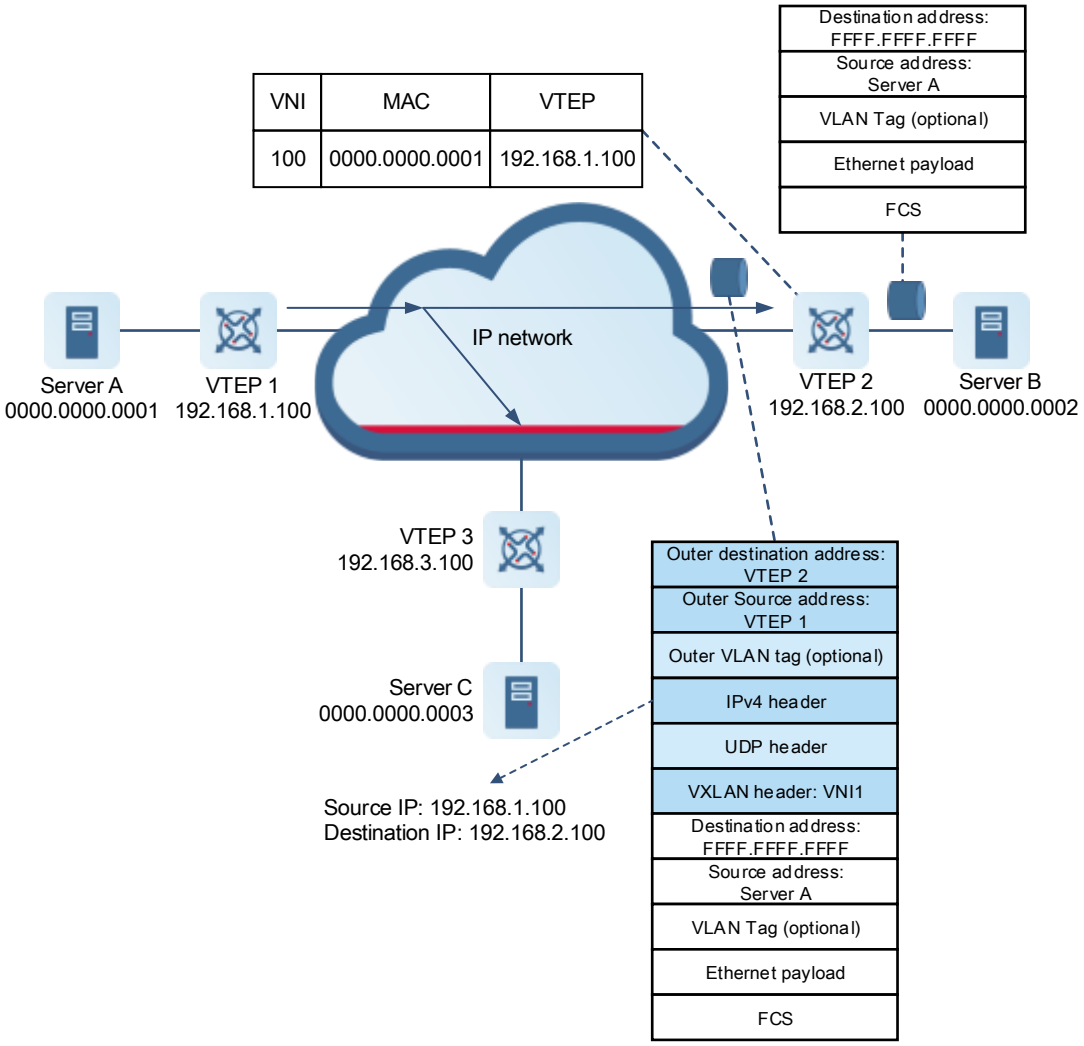
(2) The IP core network forwards the unicast VXLAN packets.

Figure 1-12 VTEP 3 Decapsulating the VXLAN Packet and Performing Address Learning



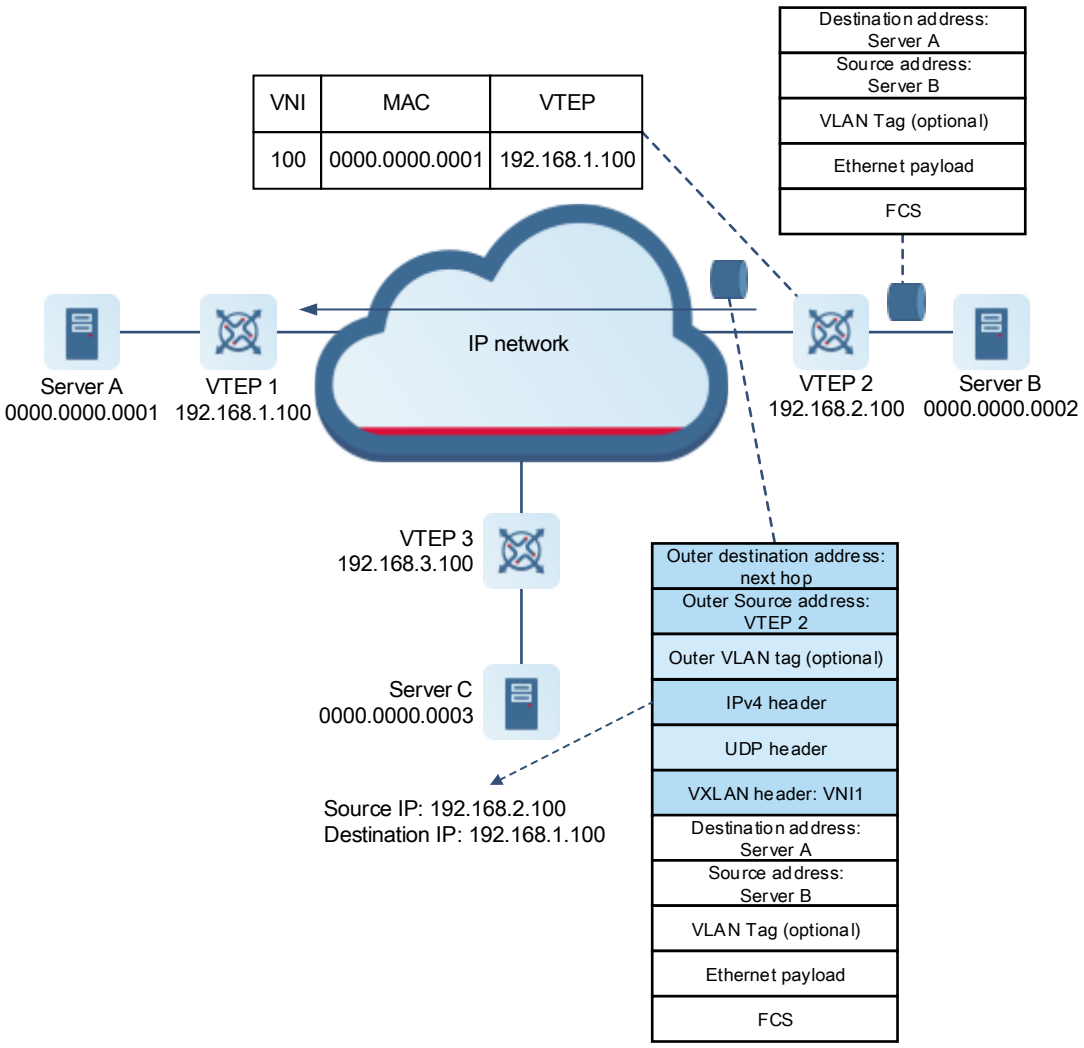
(3) After receiving the VXLAN packet, VTEP 3 decapsulates it to retrieve the Ethernet packet and performs VXLAN address learning (the VNI is 100, the MAC address is 0000.0000.0001, and the IP address is 192.168.1.100). Then, VTEP 3 floods the decapsulated packet to server C in broadcast mode.

Figure 1-13 VTEP 2 Decapsulating the VXLAN Packet and Performing Address Learning



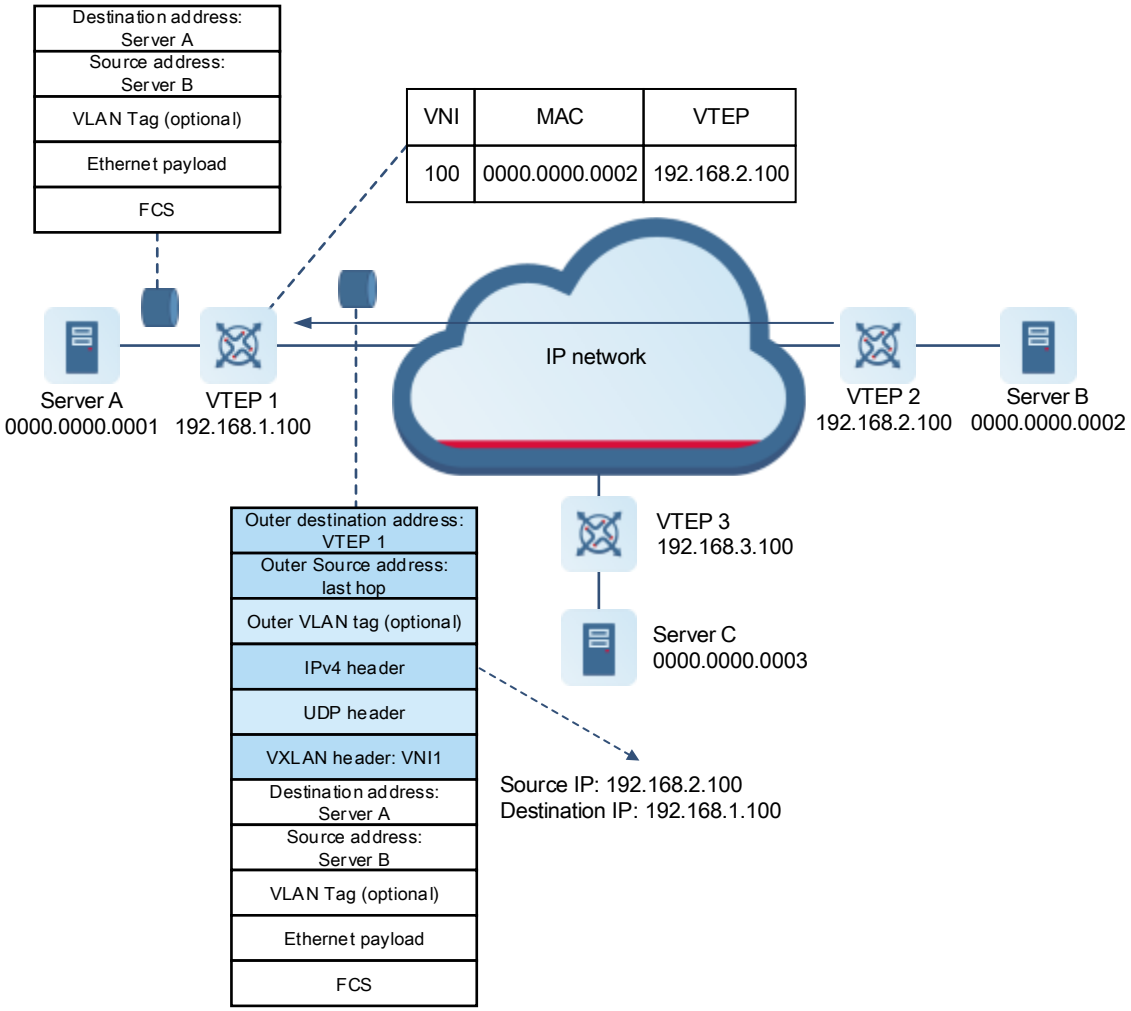
(4) After receiving the VXLAN packet, VTEP 2 decapsulates it to retrieve the Ethernet packet, learns the address of the Ethernet packet (the VNI is 100, the MAC address is 0000.0000.0001, and the IP address is 192.168.1.100), and forwards the packet. Server B gives a reply after receiving the ARP request.

Figure 1-14 VTEP 2 Searching the Address Table and the Switch Encapsulating the Packet into a Unicast VXLAN Packet



(5) After receiving the ARP reply from server B, VTEP 2 searches the address table and finds that the destination IP address of the outbound interface is 192.168.1.100. Then, the switch encapsulates the packet into a unicast VXLAN packet (the outer source IP address is 192.168.2.100) destined for the switch at 192.168.1.100.

Figure 1-15 VTEP 1 Decapsulating the Received VXLAN Packet and Performing Address Learning



(6) The IP core network forwards the VXLAN packet.

(7) After receiving the VXLAN-encapsulated ARP reply packet from server B, VTEP 1 decapsulates the packet to retrieve an Ethernet packet, performs address learning (VNI is 100, the MAC address is 0000.0000.0002, and the IP address is 192.168.2.100), and forwards the packet. Server A receives the ARP reply.

Note

After obtaining the MAC address of server B, server A interacts with server B in unicast mode. For more information about the interaction process, see the ARP response process of server B.

3. VXLAN Layer 3 Forwarding

To provide Layer 3 services for VMs at a VXLAN site, you are required to deploy a VXLAN gateway in the network so that the VMs at the site can make Layer 3 communication with external networks or VMs in other VXLAN networks through the VXLAN gateway.

Note

The **NH MAC** (**Next Hop MAC**, next-hop MAC address) field mentioned in this chapter refers to the next hop MAC address transmitted on the Underlay network. The value of the **Next Hop MAC** field varies with legends.

- Layer 3 forwarding between users in different VXLANs

  As shown in Figure 1-16, VTEP-1 and VTEP-2 implement VXLAN Layer 2 forwarding, and VTEP-3 serves as a VXLAN gateway for routing and forwarding. Configure one VNI 10 tunnel on VTEP-1, one VNI 20 tunnel on VTEP-2, and one VNI 10 tunnel and one VNI 20 tunnel on VTEP-3.

  VTEPs completes MAC and ARP information synchronization through EVPN. For existing ARP and MAC entries on each device, see Table 1-1, Table 1-2, and Table 1-3.

Table 1-1 VXLAN Tunnel Information on Devices

Device	VNI	Tunnel	SIP	DIP	Device	VNI	Tunnel	SIP	DIP
VTEP-1	10	Tunnel 3	IP-1	IP-3	VTEP-3	10	Tunnel 1	IP-3	IP-1
VTEP-2	20	Tunnel 3	IP-2	IP-3		20	Tunnel 2	IP-3	IP-2

Table 1-2 VXLAN MAC Entries on Devices

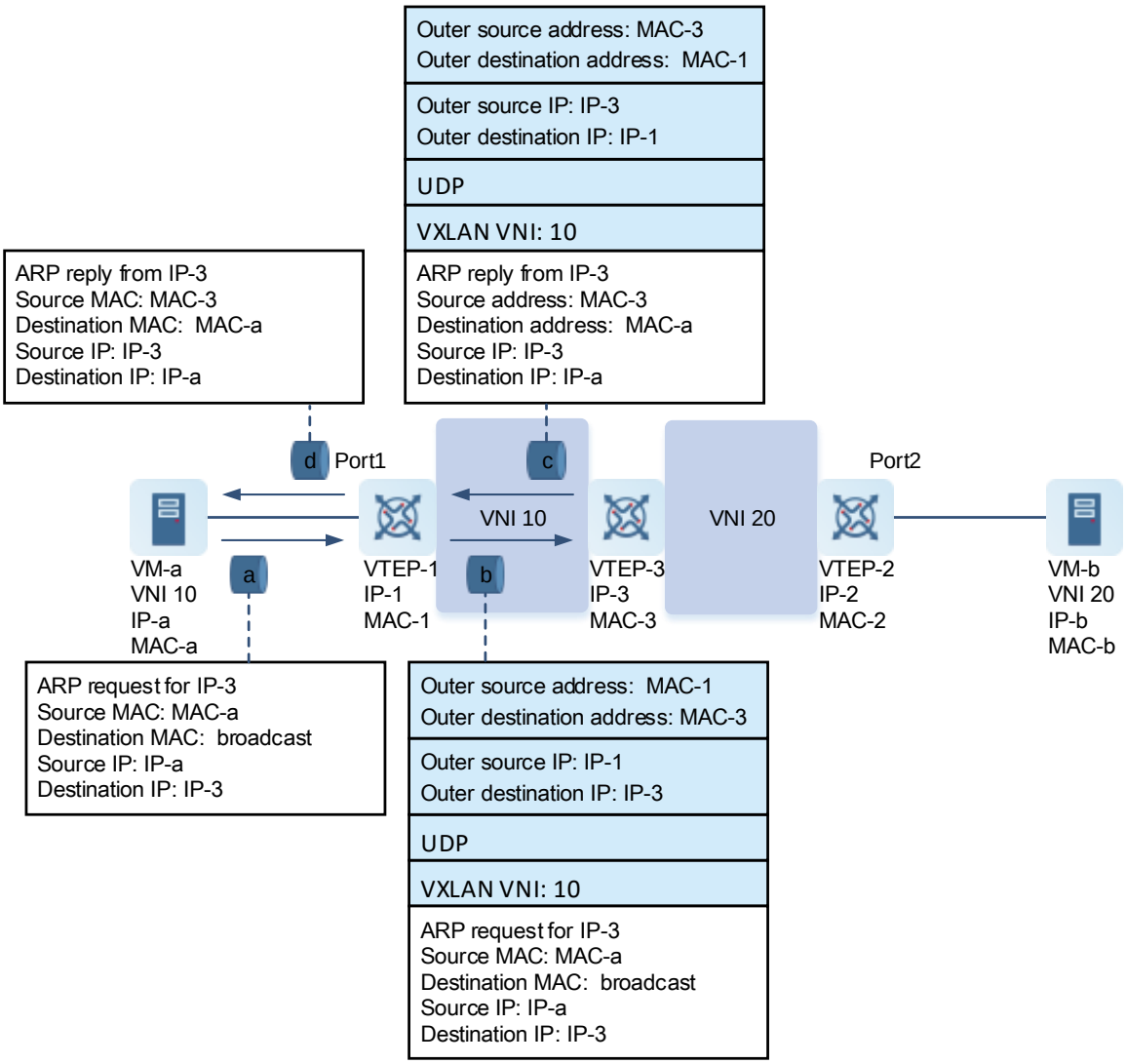
Device	MAC	VNI	Destination	Device	MAC	VNI	Destination
VTEP-1	MAC-a	10	Port1	VTEP-2	MAC-b	20	Port2
	MAC-3	10	Remote IP-3		MAC-3	20	Remote IP-3

Table 1-3 ARP Entries on VTEP-3

IP	VNI	Next Hop	MAC	IP	VNI	Next Hop	MAC
IP-a	10	Remote IP-1	MAC-a	IP-b	20	Remote IP-2	MAC-b

The communication between VM-a and VM-b is illustrated as follows:

Figure 1-16 VM-a Requesting the MAC Address of the Gateway



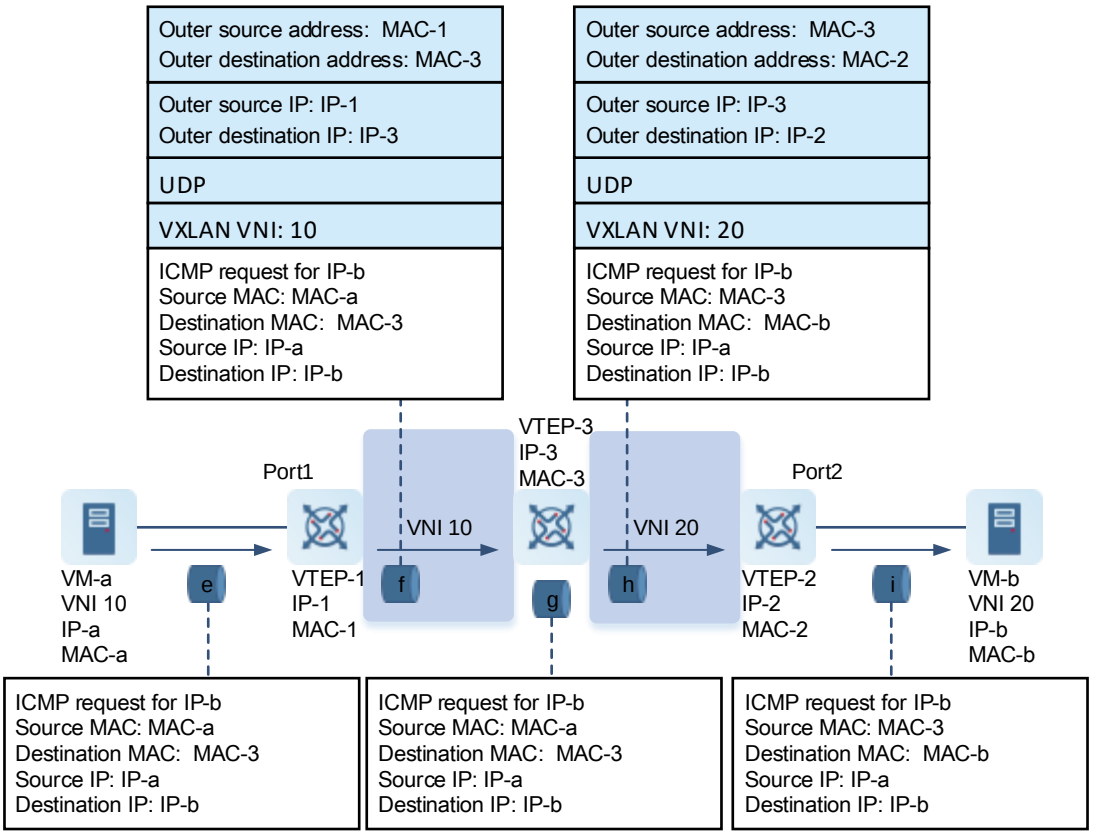
a VM-a and VM-b belong to different network segments. VM-a broadcasts an ARP packet to request the MAC address of the gateway IP-3.

b After receiving the ARP packet from VM-a, VTEP-1 encapsulates the VXLAN header (VNI 10) into the packet and sends the packet to VTEP-3 through a tunnel.

c VTEP-3 decapsulates the packet, finds that VM-a requests the MAC address of IP-3, and returns an ARP reply packet, in which the MAC address (MAC-3) of IP-3 is carried. VTEP-3 also encapsulates the VXLAN header (VNI 10) into the packet before transmitting the packet to VTEP-1.

d After receiving the VXLAN-encapsulated (VNI 10) ARP reply packet, VTEP-1 decapsulates the packet and sends the ARP reply packet to VM-a according to the MAC entry.

Figure 1-17 Forwarding an ICMP Request Between VMs in Different VXLANs



e After learning ARP information, VM-a sends an Internet Control Message Protocol (ICMP) packet to VM-b. Their communication traverses different network segments. Therefore, the destination MAC address of the packet is MAC-3 of the gateway VTEP-3.

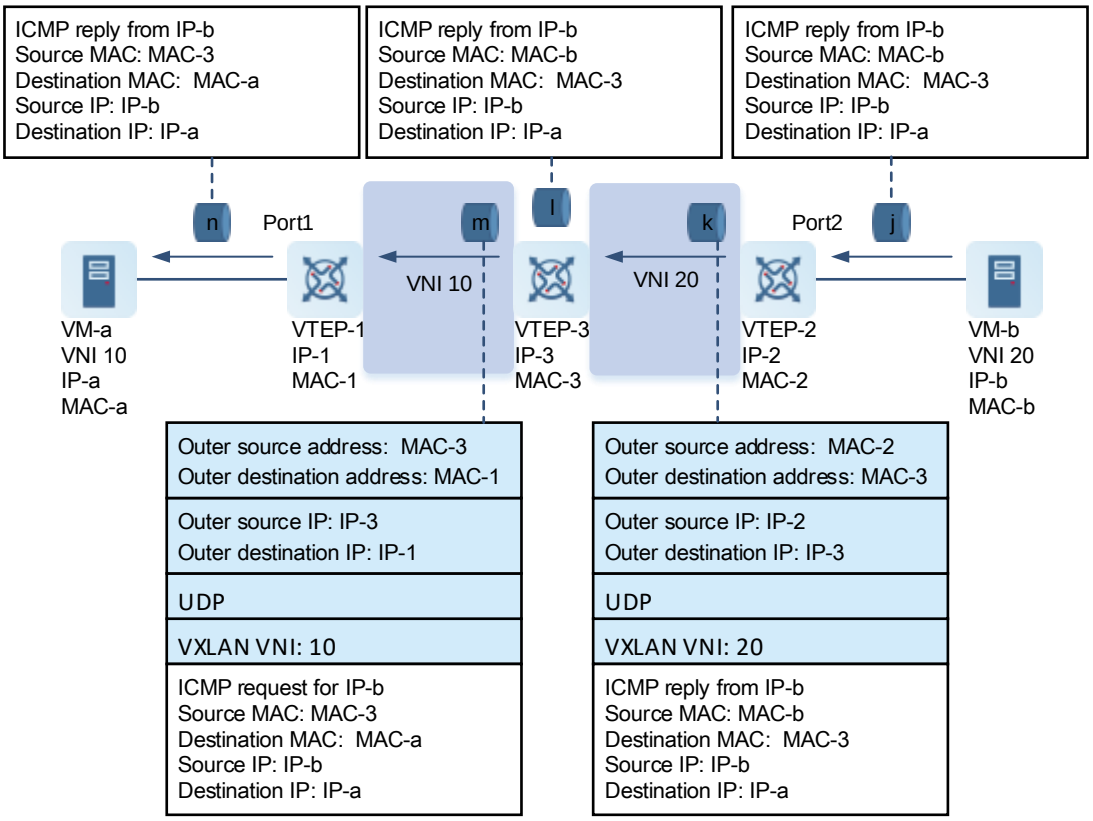
f After receiving the ICMP request, VTEP-1 searches the Layer 2 address table and finds that the packet matches an entry of MAC-3 + VNI 10 + egress of remote tunnel IP-3 in the address table. Therefore, VTEP-1 encapsulates the packet into a VXLAN packet and sends it to VTEP-3. The inner packet remains unchanged and the outer destination MAC address refers to the next-hop MAC address that corresponds to the destination IP address IP-3 in the routing table.

g After receiving the encapsulated VXLAN packet, VTEP-3 decapsulates the packet to retrieve the inner packet.

h VTEP-3 is a gateway. The decapsulated inner packet triggers Layer 3 forwarding and matches an entry with the egress of remote tunnel IP-2 in the routing table. VTEP-3 encapsulates the packet into a VXLAN packet and forwards it. The destination MAC address of the inner packet is changed to MAC-b and the outer VNI is VNI 10.

i After receiving the encapsulated VXLAN packet, VTEP-2 decapsulates the packet to retrieve the inner packet, and forwards the packet to VM-b at Layer 2 in unicast mode.

Figure 1-18 Forwarding an ICMP Reply Between VMs in Different VXLANs



j After receiving the ICMP request packet, VM-b judges that the packet is destined for itself and sends an ICMP reply packet.

k After receiving the ICMP reply packet, VTEP-2 encapsulates the packet into a VXLAN packet and forwards it at Layer 2 in unicast mode.

l After receiving the encapsulated VXLAN packet, VTEP-3 decapsulates the packet to retrieve the inner packet.

m VTEP-3 is a gateway. The decapsulated inner packet triggers Layer 3 forwarding and matches an entry with the egress of remote tunnel IP-1 in the routing table. Then, VTEP-3 encapsulates the packet into a VXLAN packet and forwards it. The destination MAC address of the inner packet is changed to MAC-a and the outer VNI is VNI 10.

n After receiving the encapsulated VXLAN packet, VTEP-1 decapsulates the packet to retrieve the inner packet, and forwards the packet to VM-a at Layer 2 in unicast mode. After receiving the packet, VM-a judges that the packet is destined for itself, indicating that the route to VM-b is reachable.

- Communication between a user in a VXLAN and a user outside the VXLAN

  As shown in Figure 1-19, VTEP-1 implements VXLAN Layer 2 forwarding, VTEP-2 serves as a VXLAN gateway, the external underlay network is connected to the virtualized overlay network through VTEP-2. The overlay network routing table in VTEP-2 is imported into the underlay network routing table so that VTEPs can communicate with each other and forward packets through the overlay network. VTEP-1 establishes one VNI 10 tunnel and VTEP-2 establishes one VNI 10 tunnel.

VTEPs completes MAC and ARP information synchronization through EVPN. For existing ARP and MAC entries on each device, see Table 1-4, Table 1-5, and Table 1-6.

Table 1-4 VXLAN Tunnel Information on Devices

Device	VNI	Tunnel	SIP	DIP	Device	VNI	Tunnel	SIP	DIP
VTEP-1	10	Tunnel 1	IP-1	IP-2	VTEP-2	10	Tunnel 1	IP-2	IP-1

Table 1-5 VXLAN MAC Entries on Devices

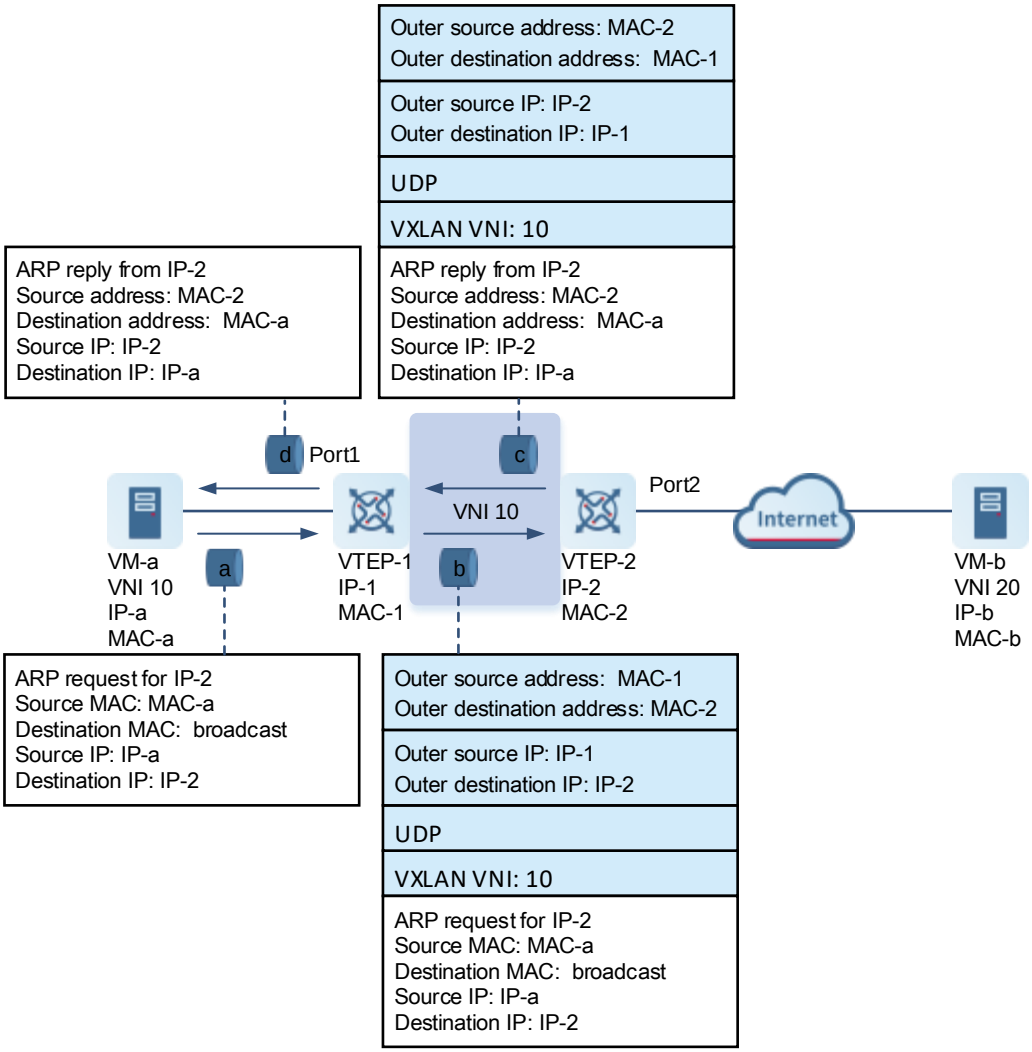
Device	MAC	VNI	Destination	Device	MAC	VNI	Destination
VTEP-1	MAC-a	10	Port1	VTEP-1	MAC-2	10	Remote IP-2

Table 1-6 ARP Entries on VTEP-2

IP	VNI	Next Hop	MAC	IP	VNI	Next Hop	MAC
IP-a	10	Remote IP-1	MAC-a	IP-b	-	Port2	Adj-MAC

The communication between VM-a and VM-b is illustrated as follows:

Figure 1-19 VM-a Requesting the MAC Address of the Gateway (External Network Communication of VXLAN)



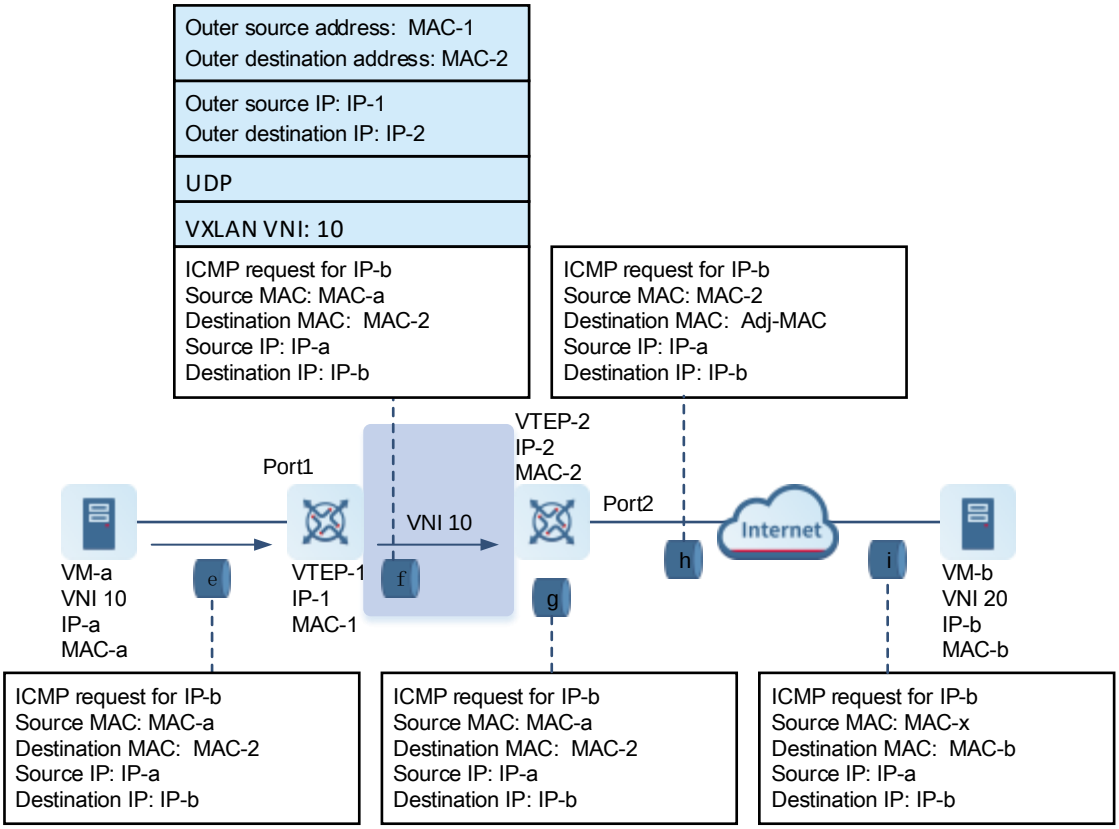
a VM-a and VM-b belong to different network segments. VM-a broadcasts an ARP packet to request the MAC address of the gateway IP-2.

b After receiving the ARP packet from VM-a, VTEP-1 encapsulates the VXLAN header (VNI 10) into the packet and sends the packet to VTEP-2 through a tunnel.

c VTEP-2 decapsulates the packet, finds that VM-a requests the MAC address of IP-2, and returns an ARP reply packet, in which the MAC address (MAC-2) of IP-2 is carried. VTEP-2 also encapsulates the VXLAN header (VNI 10) into the packet before transmitting the packet to VTEP-1.

d After receiving the VXLAN-encapsulated (VNI 10) ARP reply packet, VTEP-1 decapsulates the packet and sends the ARP reply packet to VM-a according to the MAC entry.

Figure 1-20 Forwarding an ICMP Request Between a VM Inside a VXLAN and a VM Outside (External Network Communication of VXLAN)



e After learning ARP information, VM-a sends an ICMP packet to VM-b. Their communication traverses different network segments. Therefore, the destination MAC address of the packet is MAC-2 of the gateway VTEP-2.

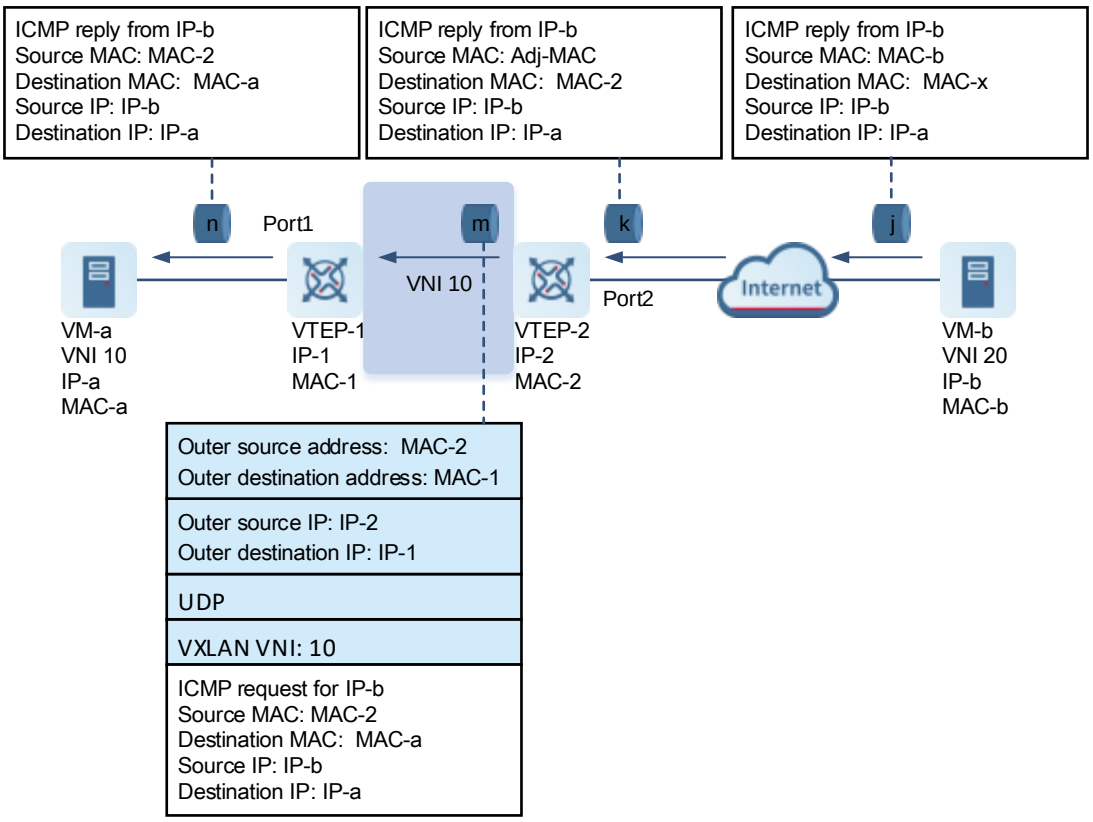
f After receiving the ICMP request, VTEP-1 searches the Layer 2 address table and finds that the packet matches an entry of MAC-2 + VNI 10 + egress of remote tunnel IP-2 in the address table. The match triggers the VXLAN packet forwarding mechanism of VTEP-1 to perform Layer 2 unicast forwarding. Therefore, VTEP-1 encapsulates the packet and sends it to VTEP-2. The inner packet remains unchanged and the outer destination MAC address is MAC-2.

g After receiving the VXLAN-encapsulated packet, VTEP-2 decapsulates the packet to retrieve the inner packet.

h VTEP-2 is a gateway. The decapsulated inner packet triggers Layer 3 forwarding and matches an entry with the egress of local port 2 in the routing table. Therefore, VTEP-2 routes and forwards the packet. The destination MAC address of the inner packet is changed to Adj-MAC.

i Then the packet is routed and forwarded as it is in a conventional network. The ICMP request packet reaches VM-b, the source MAC address is MAC-x of the previous-hop device and the destination MAC address is MAC-b.

Figure 1-21 Forwarding an ICMP Reply Between a VM Inside a VXLAN and a VM Outside (External Network Communication of VXLAN)



j After receiving the ICMP request packet, VM-b judges that the packet is destined for itself and sends an ICMP reply packet.

k Then the packet is routed and forwarded as it is in a conventional network. The ICMP reply packet reaches VTEP-2, the source MAC address is Adj-MAC, and the destination MAC address is MAC-2.

l VTEP-2 is a gateway. The received ICMP packet triggers Layer 3 forwarding and the destination IP address matches an entry with the egress of remote tunnel IP-1 in the routing table. Therefore, VTEP-2 encapsulates the packet into a VXLAN packet and forwards it. The destination MAC address of the inner packet is changed to MAC-a and the outer VNI is VNI 10.

m After receiving the encapsulated VXLAN packet, VTEP-1 decapsulates the packet to retrieve the inner packet, and forwards the packet to VM-a at Layer 2 in unicast mode. After receiving the packet, VM-a judges that the packet is destined for itself, indicating that the route to VM-b is reachable.

1.2.9 VXLAN Routing Deployment Model

Currently, VXLAN routing can be deployed in two modes: deployed in a centralized VXLAN gateway for small-scale data centers and in distributed VXLAN gateways for large-scale data centers.

1. Centralized Gateway Deployment

All VXLAN gateways in the network are deployed on a core device. Cross-VNI traffic must be forwarded by VXLAN gateways deployed on the core device. The core device needs to learn MAC addresses and ARP

information of all VMs in the topology. All access devices run only the VXLAN bridging function and do not support cross-VNI forwarding of VXLAN traffic.

2. Distributed Gateway Deployment

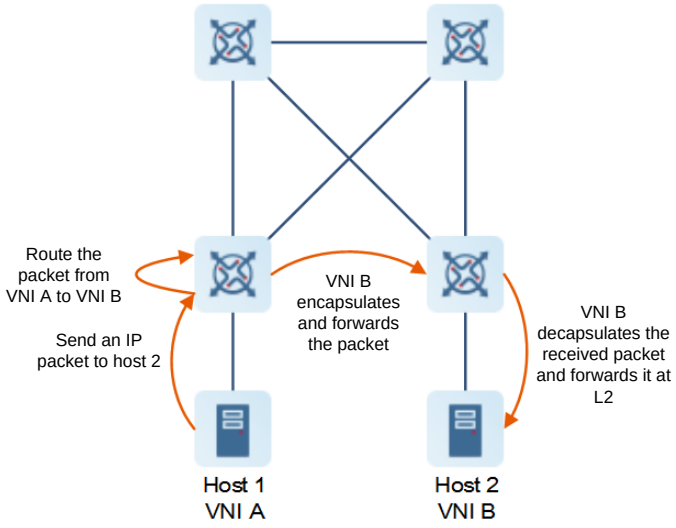
In distributed gateway deployment, each VTEP serves as the gateway of the local VM directly connected to the VTEP. Therefore, cross-VNI traffic does not need to be diverted to the centralized gateway. This deployment mode not only ensures the optimal paths, but also alleviates the traffic forwarding load on the centralized gateway. Distributed gateway deployment can be further classified into distributed asymmetric gateway deployment and distributed symmetric gateway deployment.

- Distributed asymmetric gateway deployment

  During cross-VNI forwarding of data packets, routes are looked up only on the ingress VTEP or egress VTEP. All VNIs must be configured on all VTEPs. Each VTEP can perform both VXLAN bridging and VXLAN routing and forwarding. VTEPs need to learn ARP and address entries of hosts in all VNIs, including non-local hosts. Therefore, the distributed asymmetric gateway deployment has poor scalability.

  The asymmetric gateway deployment provides the optimal path for packet forwarding and supports seamless migration of VMs. However, it is only suitable for small-scale data centers due to poor scalability.

Figure 1-22 Distributed Asymmetric Gateway Deployment



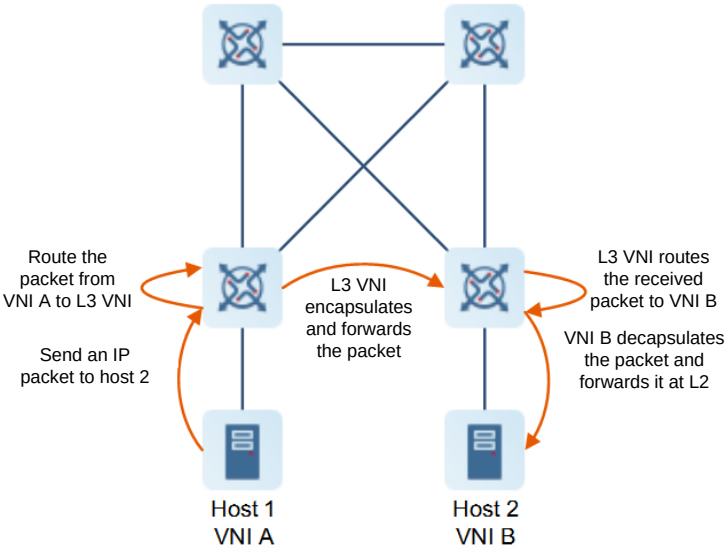
- Distributed symmetric gateway deployment

  The Layer 3 traffic of VMs triggers Layer 3 forwarding on the directly connected VTEPs, which encapsulate packets into VXLAN packets and send the packets to the peer VTEPs. After the peer VTEPs receive and decapsulate the packets, another Layer 3 forwarding is triggered. Compared with asymmetric gateway deployment, this mode is symmetric and called VXLAN symmetric gateway deployment.

  When the distributed symmetric gateway mode is deployed, if intra-VNI forwarding is needed, VXLAN Layer 2 forwarding is performed on the ingress VTEP. If cross-VNI communication is needed, traffic is routed from the ingress VTEP to Layer 3 VNI and then routed by the destination VTEP to the host based on the inner IP address. In this mode, VNI information of the destination host does not need to be configured on the ingress VTEP. Therefore, VTEPs do not need to learn the MAC address and ARP information of the remote hosts, reducing the resource consumption of the MAC address table and ARP adjacency table.

Distributed symmetric gateway deployment reduces the pressure in the capacity of the MAC address table on VTEPs and provides better scalability. However, the seamless VM migration is limited due to the single-side VNI deployment on VTEPs.

Figure 1-23 Distributed Symmetric Gateway Deployment



1.2.10 Anycast Gateway

Anycast gateway, also called all-active gateway, is a VXLAN gateway deployment mode. In this deployment mode, all gateways are active and any one of them supports cross-VNI routing and forwarding. Generally, the anycast gateway deployment requires two VXLAN gateways. The same virtual MAC address and gateway address must be deployed on the two gateways. Anycast gateway provides different benefits in centralized and distributed gateway deployment scenarios.

In a centralized gateway deployment scenario, the same VXLAN gateway and the same VTEP IP address are configured on multiple core devices to form an all-active anycast gateway based on the equal-cost multi-path routing (ECMP) of the underlay network. When one of the core devices malfunctions, cross-VNI traffic can be balanced to other core devices to prevent service interruption.

In a distributed gateway deployment scenario, if anycast gateway is not deployed, the cross-VNI traffic is forwarded to the only gateway in the network to trigger Layer 3 forwarding even if the VMs are directly connected to VTEPs. As a result, unnecessary detour is caused to most VXLAN packets in the network. After anycast gateway is configured, Layer 3 packets of VMs are directly forwarded by VTEPs directly connected to the VMs according to routes, and traffic forwarding paths in the VXLAN network are optimized. VM migration is also supported in anycast gateway deployment mode.

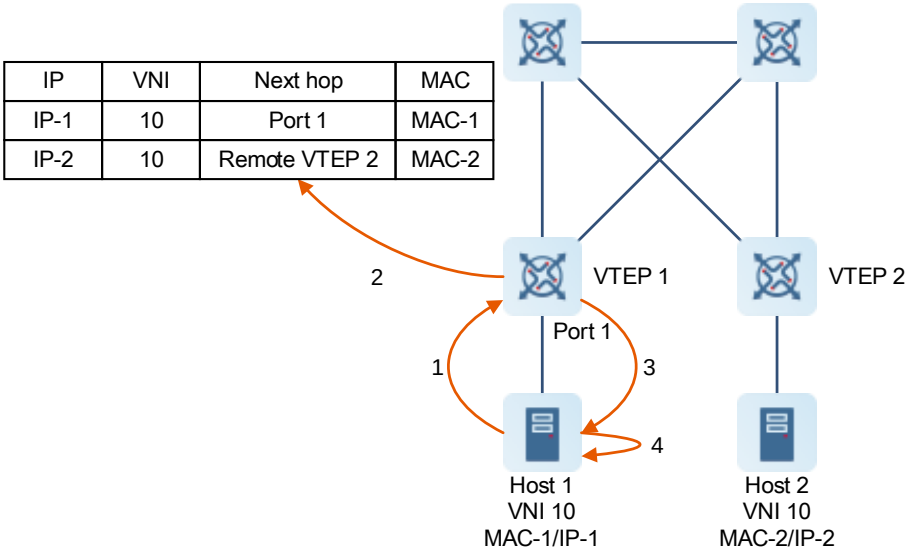
1.2.11 ARP/ND Flood Suppression

ARP/Neighbor Discovery (ND) flood suppression reduces broadcast traffic in the network by maintaining the ARP/ND cache table in a VNI and enabling VTEPs to respond to ARP/ND requests as a proxy. After ARP/ND flood suppression is enabled, a VTEP maintains the ARP/ND cache table in the VNI. When a host sends an ARP/neighbor solicitation (NS) request, the VTEP directly connected to the host intercepts the ARP/NS request, and responds to the ARP/NS request as a proxy or floods the request to other VTEPs. Dumb terminals (that do

not send any traffic before receiving a request) may exist in a network. If a VTEP does not have ARP/ND information of a dumb terminal (serving as a destination host), the ARP/NS request needs to be flooded to other VTEPs.

Most terminals send a Gratuitous Address Resolution Protocol (GARP), Reverse Address Resolution Protocol (RARP), or NS message upon going online. VTEPs directly connected to the terminals can immediately learn their MAC and IP addresses and spread the information to other VTEPs. Therefore, ARP suppression can reduce flood traffic caused by terminal ARP/ND learning. The following uses ARP suppression as an example to illustrate the flood suppression process.

Figure 1-24 ARP Suppression



(1) Host 1 knows that the IP address of host 2 is IP-2 and sends an ARP request to request the MAC address of host 2.

(2) After receiving the ARP request, VTEP 1 queries the ARP cache table of VNI 10 and finds that the MAC entry of host 2 exists and the MAC address of host 2 is MAC-2.

(3) VTEP 1 sends an ARP reply to host 1, which carries the MAC address (MAC-2) of host 2.

(4) Host 1 receives the ARP reply packet and establishes the mapping between IP-2 and MAC-2.

1.3 Protocols and Standards

- RFC 7348: Virtual eXtensible Local Area Network

2 Configuring VXLAN Basic Features

2.1 Overview

VXLAN supports centralized networking and distributed networking based on the VXLAN routing deployment mode. Centralized networking is classified into static tunnel networking and dynamic tunnel networking,

depending on the tunnel creation mode. Distributed networking is further classified into symmetric networking and asymmetric networking based on whether symmetric routing is adopted. The four networking modes above are the most basic VXLAN networking modes.

2.1.1 Centralized VXLAN Networking Using Static Tunnels

Centralized VXLAN networking using static tunnels is applicable to small- and medium-sized data centers:

- In a centralized networking scenario, all traffic must pass through the centralized gateway, and the core device needs to learn the MAC addresses and ARP information of all VMs in the topology. Therefore, the network size is limited by the core device.
- Tunnels are created statically. Configuration delivery or manual configuration is required for the establishment of each tunnel. Some routes and flow tables also need to be manually delivered. Once the network is oversized, the configuration workload increases sharply, which brings difficulties in network management and planning.

In this networking mode, VXLAN Layer 2 functions need to be configured on access devices to connect VMs to the VXLAN network, and VXLAN Layer 3 functions need to be configured on core devices to provide the VXLAN routing function and implement cross-VXLAN communication. VXLAN tunnels are generated via static configuration. VXLAN configurations can be delivered by the software-defined networking (SDN) controller through communication mechanisms such as NETCONF or manually completed on the command line interface (CLI).

Host routes and VXLAN forwarding flow tables can be delivered by the SDN controller, or routes and entries can be generated via automatic learning. When the SDN controller malfunctions, the automatic learning function ensures that the VXLAN works normally.

2.1.2 Centralized VXLAN Networking Using Dynamic Tunnels

Centralized VXLAN networking using dynamic tunnels is applicable to small- and medium-sized data centers. In a centralized networking scenario, all traffic must pass through the centralized gateway, and the core device needs to learn the MAC addresses and ARP information of all VMs in the topology. Therefore, the network size is limited by the core device.

In this networking mode, VXLAN Layer 2 functions need to be configured on access devices to connect VMs to the VXLAN network, and VXLAN Layer 3 functions need to be configured on core devices to provide the VXLAN routing function and implement cross-VXLAN communication. The EVPN control plane learning function needs to be enabled on access devices and core devices. They accomplish VXLAN tunnel learning, MAC address learning, and route learning via the control plane protocol, as well as VXLAN bridging and VXLAN routing to implement data communication between VXLANs and between a VXLAN and an external network.

2.1.3 Distributed Asymmetric VXLAN Networking

The distributed asymmetric VXLAN networking is applicable to small- and medium-sized data centers. This networking mode is no longer limited by the performance of the core device but it is still inapplicable to large-scale data centers. The reasons are that all VNIs need to be configured on each access device and the devices need to learn ARP and address entries of all hosts.

In this networking mode, VXLAN Layer 2 and Layer 3 functions need to be configured on access devices, to enable VMs to connect to the VXLAN network and provide the VXLAN routing function. The same overlay router gateway must be configured on all access devices. The EVPN control plane learning function needs to

be enabled on access devices. They accomplish VXLAN tunnel learning, MAC address learning, and route learning via the control plane protocol, as well as VXLAN bridging and VXLAN routing to implement data communication between VXLANs and between a VXLAN and an external network.

2.1.4 Distributed Symmetric VXLAN Networking

The distributed symmetric VXLAN networking is applicable to large-sized data centers. In symmetric networking, VNIs do not need to be configured on access devices, and they do not need to learn ARP and address entries of all hosts. Therefore, symmetric VXLAN networking has better scalability.

As in distributed asymmetric networking, VXLAN Layer 2 and Layer 3 functions and EVPN control plane learning need to be configured on access devices. Different overlay router gateways can be configured on access devices, but a unified Layer 3 VNI needs to be configured for Layer 3 traffic forwarding.

2.2 Restrictions and Guidelines

2.2.1 Centralized VXLAN Networking Using Static Tunnels

- VXLAN configurations can be delivered by the SDN controller through communication mechanisms such as NETCONF, or manually completed on the CLI. Only configuration delivery from the SDN controller is recommended in normal cases.
- The VXLAN requires the support from existing unicast routes in the network. Therefore, devices in the network must ensure the connectivity of the IP network through other routing protocols such as the Open Shortest Path First (OSPF).
- On the centralized anycast gateway, you need to add the anycast IP address of gateways to different routing domains to prevent IP address conflict.

2.2.2 Centralized VXLAN Networking Using Dynamic Tunnels

- The VXLAN requires the support from existing unicast routes in the network. Therefore, devices in the network must ensure the connectivity of the IP network through other routing protocols such as OSPF.
- The VXLAN completes VXLAN tunnel learning, MAC address learning, and route learning through the MP-BGP-EVPN protocol. Therefore, BGP configuration must be completed on VTEPs.

2.2.3 Distributed Asymmetric VXLAN Networking

- Layer 3 traffic is routed and forwarded once by the VXLAN gateway only on the ingress VTEP, and the egress VTEP forwards only Layer 2 traffic.
- The gateways of all VXLANs must be configured on each VTEP.
- VTEPs establish a BGP neighbor relationship and synchronize MAC address and ARP information through BGP-EVPN type-2 routes.
- VTEPs do not learn ARP information and MAC addresses from packets received from tunnels by default.

2.2.4 Distributed Symmetric VXLAN Networking

- Layer 3 forwarding is required on both the ingress VTEP and egress VTEP. Therefore, the overlay router interface of the VXLAN distributed gateway and the overlay router interface of the Layer 3 VNI need to be configured on all ingress VTEPs and egress VTEPs.
- Only one symmetric instance is allowed in a virtual routing and forwarding (VRF) instance.

2.3 Configuration Task Summary

2.3.1 Centralized VXLAN Networking Using Static Tunnels

The configuration of the centralized VXLAN networking using static tunnels includes the following tasks:

(1) Configuring a VXLAN Routed

(2) Configuring a Static Tunnel

(3) Configuring a VXLAN

(4) (Optional) Configuring an EVPN

a Configuring a BGP EVPN

b Configuring an EVI

(5) (Optional) Configuring an Anycast Gateway

(6) (Optional) Enabling the Anycast Gateway Synchronization Function

2.3.2 Centralized VXLAN Networking Using Dynamic Tunnels

The configuration of the centralized VXLAN networking using dynamic tunnels includes the following tasks:

(1) Configuring a VXLAN Routed

(2) Configuring a VXLAN

(3) Configuring a VTEP

(4) Configuring an EVPN

a Configuring a BGP EVPN

b Configuring an EVI

(5) Enabling ARP/ND Automatic Learning

(6) (Optional) Configuring an Anycast Gateway

2.3.3 Distributed Asymmetric VXLAN Networking

The configuration of the distributed asymmetric VXLAN networking includes the following tasks:

(1) Configuring a VXLAN Routed

(2) Configuring a VXLAN

(3) Configuring a VTEP

(4) Configuring an EVPN

a Configuring a BGP EVPN

b Configuring an EVI

(5) Configuring an Anycast Gateway

2.3.4 Distributed Symmetric VXLAN Networking

The configuration of the distributed symmetric VXLAN networking includes the following tasks:

(1) Configuring a VXLAN Routed

(2) Configuring a VXLAN

(3) Configuring a VTEP

(4) Configuring an EVPN

a Configuring a BGP EVPN

b Configuring an EVI

(5) Configuring an Anycast Gateway

(6) Configuring a Symmetric VXLAN

2.4 Configuring a VXLAN Routed Interface

2.4.1 Overview

This section describes how to create an overlay router interface and configure the VXLAN gateway IP address.

2.4.2 Restrictions and Guidelines

The configuration is required on the VXLAN gateway. The overlay router interface and gateway IP address are configured on the core device in a centralized scenario, and on access devices in a distributed scenario.

2.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create an overlay router interface and enter the overlay router interface configuration mode.

interface overlayrouter *interface-number*

(4) Configure an IP address for the overlay router interface.

(IPv4 network)

ip address *ipv4-address mask*

No IPv4 address is configured for an interface by default.

(IPv6 network)

ipv6 address *ipv6-address mask*

No IPv6 address is configured for an interface by default.

2.5 Configuring a Static Tunnel

2.5.1 Overview

This section describes how to create a VXLAN tunnel interface and configure the source and destination IP addresses for the interface so that a VXLAN static tunnel is established between devices.

2.5.2 Restrictions and Guidelines

- The source and destination addresses of a VXLAN tunnel interface must belong to the same IP protocol family.

- In scenarios of centralized VXLAN networking using static tunnels, the configuration is required on VXLAN access devices and VXLAN gateways.

2.5.3 Prerequisites

Before configuring a VXLAN tunnel interface, ensure that the route between devices, where the source and destination IP addresses of the VXLAN tunnel interface reside, is reachable. Otherwise, a tunnel cannot be established.

2.5.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a VXLAN tunnel interface and enter the VXLAN tunnel interface configuration mode.

interface overlaytunnel *interface-number*

(4) Configure a source address for the VXLAN tunnel.

overlaytunnel source *ipv4-address*

No source (local) IP address is configured for a VXLAN tunnel by default.

(5) Configure a destination address for the VXLAN tunnel.

overlaytunnel destination *ipv4-address*

No destination IP address is configured for a VXLAN tunnel by default.

2.6 Configuring a VXLAN Instance

2.6.1 Overview

This section describes how to create a VXLAN instance, and configure an overlay router interface, VXLAN static tunnel, and associated VLAN for the VXLAN instance.

2.6.2 Restrictions and Guidelines

- The **tunnel-interface** command must be associated with a created VXLAN tunnel interface.
- The **router-interface** command must be associated with a created overlay router interface.
- One overlay router interface can associate with only one VXLAN.
- Different VXLAN instances cannot be associated with the same VLAN.
- After a VLAN is associated with a VXLAN, all packets of the VLAN will be encapsulated into VXLAN packets. Therefore, an SVI cannot be used as the IP gateway of the VLAN on the device.
- In any VXLAN networking scenarios, the configuration is required on VXLAN access devices and VXLAN gateways.

2.6.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a VLAN and enter the VLAN configuration mode.

vlan *vlan-id*

Only one common static VLAN (VLAN 1) exists by default.

(4) Return to the global configuration mode.

exit

(5) Create a VXLAN instance and enter the VXLAN configuration mode.

vxlan *vni-number*

No VXLAN instance is configured by default.

(6) All the configuration tasks below are optional. Select one of the tasks for configuration based on the VXLAN networking mode and device role.

- Associate the VXLAN instance with an overlay router interface.

  router-interface overlayrouter *interface-number*

  The configuration is required on the VXLAN gateway.

  No VXLAN routing interface is configured by default.

- Associate the VXLAN instance with a VXLAN tunnel.

  tunnel-interface overlaytunnel *interface-number*

  In a scenario, in which a static tunnel is configured, the configuration is required on devices at both ends of the static tunnel.

  No VXLAN instance is configured to associate with a tunnel by default.

- Associate the VXLAN instance with a VLAN.

  extend-vlan *vlan-list*

  The configuration is required on VXLAN access devices.

  The VLAN to be associated with a VXLAN instance is not configured by default.

2.7 Configuring a VTEP

2.7.1 Overview

This section describes how to configure a VTEP.

2.7.2 Restrictions and Guidelines

- In scenarios of centralized VXLAN networking using dynamic tunnels, distributed asymmetric VXLAN networking, and distributed symmetric VXLAN networking, the configuration is required on VXLAN access devices and VXLAN gateways.
- You can specify an IPv4 address for a VTEP in either of the following ways. One VTEP can associate with one IPv4 address in only one way:
  - Associating the VTEP with a loopback interface to use the interface's IPv4 address

- Specifying an IPv4 address directly

- One VTEP can associate with only one IPv4 address.
- If the specified Layer 3 outbound interface is an overlay router interface e during static route configuration, the next-hop IP address cannot be set to the associated IP address of a VTEP.

2.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VTEP configuration mode.

vtep

(4) Configure an IP address for the VTEP by configuring any of the following tasks:

- Bind a loopback interface to the VTEP.

  source loopback *loopback-port-id*

  No loopback interface is bound to a VTEP by default.

- Bind an IPv4 address to the VTEP.

  source ipv4 *ipv4-address*

  No IPv4 address is configured for a VTEP by default.

2.8 Configuring an EVPN

2.8.1 Overview

The VXLAN implementation involves only the data plane but no control plane. Therefore, when no control plane is available, the establishment of a VXLAN tunnel and remote device address learning can be carried out only through the data plane flooding. This severely affects the normal forwarding of service traffic in the data center network. To solve this problem, VXLAN uses an EVPN as the control plane. After an EVPN is configured, the VXLAN can establish and maintain VXLAN tunnels and synchronize address information of the remote device through the EVPN.

Configuring an EVPN includes two parts:

- Create and configure a BGP EVPN peer relationship.
- Create and configure an EVAN instance (EVI).

2.8.2 Restrictions and Guidelines

- In scenarios of centralized VXLAN networking using SDN static tunnels, if the VXLAN anycast gateway synchronization function needs to be enabled, the configuration is required on VXLAN gateways.
- In scenarios of centralized VXLAN networking using dynamic tunnels, distributed asymmetric VXLAN networking, and distributed symmetric VXLAN networking, the configuration is required on VXLAN access devices and VXLAN gateways.

2.8.3 Configuration Tasks

EVPN configuration includes the following tasks:

(1) Configuring a BGP EVPN

(2) Configuring an EVI

2.8.4 Configuring a BGP EVPN

1. Prerequisites

Ensure that routes between devices are reachable before configuration.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a BGP instance and enter the BGP configuration mode.

router-protocol bgp *as-number*

BGP is disabled by default.

(4) Configure a BGP peer (group).

peer { *neighbor-address* | *peer-group-name* } **remote-as** *as-number*

A BGP peer is disabled by default.

(5) Configure a network interface to be used for establishing a BGP connection between IBGP peers.

peer { *neighbor-address* | *peer-group-name* } **update-source** { *interface-type interface-number* | *address* }

The optimal local interface is used as the output interface by default.

(6) Configure BGP neighbors to exchange L2VPN EVPN information and enter the EVPN address family configuration mode.

address-family-mode l2vpn evpn

The L2VPN address family is disabled by default.

(7) Activate neighbors or peer groups in current address mode.

peer { *neighbor-address* | *peer-group-name* } **activate** { **ipv4** | **ipv6** }

Neighbors or peer groups are activated in address family IPv4 unicast mode by default.

2.8.5 Configuring an EVI

1. Prerequisites

Ensure that routes between devices are reachable before configuration.

2. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the EVPN configuration mode.

evpn

(4) Create an EVI and enter the VNI configuration mode.

vni *vni-id*

No EVI is configured by default.

(5) Define a route distinguisher (RD) for the EVI.

rd { **auto** | *rd-value* }

No RD is configured by default.

(6) Define the route target (RT) attribute for the EVI.

route-target { **import** | **export** | **both** } { **auto** | *rt-value* }

No RT is configured by default.

(7) (Optional) Configure a routing policy rule for the local device to distribute EVPN routes to the remote device.

export map *routemap-name*

No routing policy rule is configured for the local device to distribute EVPN routes to the remote device by default.

(8) (Optional) Configure a routing policy rule of importing remote EVPN routes to the local VNI.

import map *routemap-name*

No routing policy rule is configured for importing remote EVPN routes to the local VNI by default.

2.9 Enabling ARP/ND Automatic Learning

2.9.1 Overview

After remote ARP/ND learning is configured, the switch learns ARP/ND entries from VXLAN-encapsulated ARP/ND packets received from VXLAN tunnels.

2.9.2 Restrictions and Guidelines

- You are advised to enable this function only on gateways in the centralized VXLAN topology. This function is not recommended in other scenarios.
- This function takes effect only on dynamic tunnels and is invalid to static tunnels. The device learns ARP entries from VXLAN packets received from static tunnels by default.

2.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VTEP configuration mode.

vtep

(4) Enable ARP/ND automatic learning.

(IPv4 network)

remote arp learn enable

The global remote ARP packet learning function is disabled by default.

(IPv6 network)

remote nd learn enable

The global remote IPv6 ND entry learning function is disabled by default.

2.10 Configuring an Anycast Gateway

2.10.1 Overview

This section describes how to configure VXLAN gateways as an anycast gateway.

2.10.2 Restrictions and Guidelines

- After the anycast attribute is configured, the device uses the global virtual anycast MAC address as the gateway MAC address.
- When multiple VTEPs are virtualized into one logical VTEP, the virtual anycast MAC address must be the same throughout the network. The anycast MAC address cannot be set to 0000.0000.0000, multicast MAC address, local MAC address, or MAC addresses of other VXLAN devices in the network, to ensure function availability.
- The anycast gateway IP address must be the same in the same VXLAN instance.
- If the device supports a network virtualization edge (NVE) interface, this command will associate with the NVE interface. When the corresponding NVE interface is deleted or configured to work in default state, the configuration of this command will be deleted accordingly.
- In centralized VXLAN networking scenarios, you can configure the settings based on the convergence requirements of the actual network. If a high network convergence requirement is raised, you can configure the settings on the VXLAN gateway.
- In distributed VXLAN networking scenarios, the configuration is required on VXLAN gateways.
- When a VXLAN is configured on the multichassis link aggregation group (M-LAG) system, the VXLAN gateway must be configured as the anycast gateway.

2.10.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure an anycast MAC address.

fabric anycast-gateway-mac *mac-address*

No global anycast virtual MAC address is configured by default.

The configuration is required on the centralized anycast gateway.

(4) Enter the overlay router interface configuration mode.

interface overlayrouter *interface-number*

(5) Configure the anycast attribute for the overlay router interface.

anycast-gateway

No anycast attribute is configured for the overlay router interface by default.

2.11 Enabling the Anycast Gateway Synchronization Function

2.11.1 Restrictions and Guidelines

- If a static tunnel is utilized to deploy the VXLAN centralized anycast gateway, you can use the VXLAN anycast gateway synchronization function to synchronize ARP, IPv6 ND, and MAC entries on all gateways, on which the static tunnel is configured.
- After VXLAN anycast gateway synchronization is enabled, the MAC, ARP, and IPv6 ND entries with the egress of a static tunnel are synchronized outward, and MAC and ARP entries that are synchronized from other devices and have the egress of the static tunnel also take effect.
- The gateway synchronization function can be enabled only on the anycast gateway. It does not need to be enabled on VXLAN access devices.
- To enable the gateway synchronization function on the anycast gateway, ensure that the function is enabled on each physical gateway. Otherwise, entry synchronization may be omitted, affecting the normal operation of the gateway.

2.11.2 Prerequisites

Before configuring synchronization on the anycast gateway, perform the following operations on the anycast gateway:

- Create a BGP instance and configure its peers. Configure the network interface used for establishing BGP connections and configure BGP neighbors to exchange L2VPN EVPN information.
- Create an EVI, and configure the VNI ID, EVI RD and RT attributes.

2.11.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VTEP configuration mode.

vtep

(4) Enable VXLAN all-active gateway synchronization.

vxlan gateway synchronize

No VXLAN all-active gateway synchronization is configured by default.

The interval for configuring the **vxlan gateway synchronize** command consecutively should be greater than 10s. Otherwise, the configuration will fail.

(5) Return to the global configuration mode.

exit

(6) Enter the VXLAN tunnel interface configuration mode.

interface overlaytunnel *interface-number*

(7) (Optional) Enable the function of synchronizing MAC entries with the egress of static tunnels to the remote VTEP.

evpn mac advertise enable

The function of synchronizing MAC entries with the egress of static tunnels to the remote VTEP is disabled by default.

(8) (Optional) Enable the function of synchronizing ARP/ND entries with the egress of static tunnels to the remote VTEP.

evpn mac-ip advertise enable

The function of synchronizing ARP/ND entries with the egress of static tunnels to the remote VTEP is disabled by default.

2.12 Configuring a Symmetric VXLAN

2.12.1 Overview

This section describes how to configure the symmetric attribute for a VXLAN instance. VXLAN Layer 3 traffic of the VRF instance where a symmetric overlay router interface resides is forwarded by the VXLAN routed interface associated with the instance.

2.12.2 Restrictions and Guidelines

- Symmetric instances need to be configured only in symmetric scenarios. Only one symmetric instance can be configured in each VRF instance. After a symmetric instance is configured in a VRF instance, Layer 3 forwarding in other asymmetric instances is switched to the symmetric instance.
- The symmetric VXLAN and anycast gateway synchronization functions are mutually exclusive and cannot be configured at the same time.
- In distributed symmetric VXLAN networking scenarios, the configuration is required on the VXLAN gateway.

2.12.3 Prerequisites

To configure the symmetric attribute for a VXLAN instance, complete the following operations:

- Create and configure a VRF instance.
- Create and configure an overlay router interface and associate it with a VRF instance.
- Associate the overlay router interface with the VXLAN instance, for which the symmetric attribute is to be configured.

2.12.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VXLAN configuration mode.

vxlan *vni-number*

No VXLAN instance is configured by default.

(4) Configure the symmetric attribute for the instance.

symmetric

No symmetric attribute is configured for a VXLAN instance by default.

3 Configuring VXLAN Extended Functions

3.1 Overview

This section describes VXLAN extended functions, which are a supplement to VXLAN basic functions. Configure appropriate extended functions as required.

3.2 Restrictions and Guidelines

VXLAN extended functions described in this section are not applicable to all VXLAN networking scenarios. For details, see the restrictions and guidelines of each function.

3.3 Configuration Task Summary

All the configuration tasks below are optional. Select the configuration tasks as required.

- Enabling Storm Control
- Configuring the VXLAN UDP Destination Port
- Configuring a Static VXLAN MAC Address Table
- Configuring Multi-Tenant Layer 3 Packet Isolation
- Configuring the EVPN Synchronization Attribute
- Configuring Rate Limits for a Tunnel Interface
- Configuring ARP/ND Suppression
- Configuring ARP/ND Proxy
- Creating VXLAN Instances in Batches
- Configuring a Data Center Interconnection Tunnel
- Configuring VNI Mapping
- Configuring a Layer 2 Ethernet Sub-interface to Access a VXLAN
- Configuring a Delay in the Protocol Status Change of a VXLAN Tunnel
- Configuring the Optimal Minimum Route Length for VXLAN Tunnels

3.4 Enabling Storm Control

3.4.1 Overview

This section describes how to configure storm control for a VXLAN instance. You can limit the rate of broadcast, multicast, or unicast VXLAN packets to implement storm control.

3.4.2 Restrictions and Guidelines

If VNI-based storm control is required, you can configure storm control rate limit for a VXLAN instance.

3.4.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VXLAN configuration mode.

vxlan *vni-number*

No VXLAN instance is configured by default.

(4) Configure storm control for the VXLAN instance.

storm-control-enable { **broadcast** | **multicast** | **unicast** } [**pps** *limit-value-pps* | **kbps** *limit-value-kbps* | **mbps** *limit-value-mbps*]

No storm control is configured for a VXLAN instance by default.

3.5 Configuring the VXLAN UDP Destination Port

3.5.1 Overview

This section describes how to manually configure the VXLAN UDP destination port ID. The VXLAN UDP destination port used by some devices is not 4789 before the implementation of RFC 7348: Virtual eXtensible Local Area Network. You can run this command to change the VXLAN UDP port ID of the device to ensure compatibility. You can also run this command to define the VXLAN UDP destination port.

3.5.2 Restrictions and Guidelines

After modifying the UDP destination port, save the configuration and restart the device for the configuration to take effect.

3.5.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the VXLAN UDP destination port.

vxlan udp-port *port-number*

The default VXLAN UDP destination port is **4789**.

3.6 Configuring a Static VXLAN MAC Address Table

3.6.1 Overview

Configured static VXLAN MAC forwarding entries are usually used in scenarios of centralized networking using static tunnels. The VXLAN MAC address table delivered by the SDN controller is considered as a static VXLAN MAC address table on the device.

Restrictions and Guidelines

- The next-hop interface must have been created before configuration.
- The next-hop interface cannot be a dynamic tunnel.

3.6.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure a static VXLAN MAC address table.

vxlan mac static *mac-address* **vni** *vni-number* [**interface** *interface-type interface-number* **vlan** *vlan-id* | **interface overlaytunnel** *interface-number*]

No static VXLAN MAC forwarding entry is configured by default.

3.7 Configuring Multi-Tenant Layer 3 Packet Isolation

3.7.1 Overview

The network of a cloud data center can be leased to multiple tenants and tenants cannot access each other to ensure the independence and security of the tenants' networks. Subnets of different VXLANs are Layer 2-isolated and can communicate with each other only through Layer 3 forwarding. Each tenant may have multiple VXLAN subnets with different VNIs. Therefore, it is necessary to ensure that different tenants can perform Layer 3 forwarding only in their own networks and cannot communicate with each other through Layer 3 forwarding.

Multi-tenant Layer 3 packet isolation can be configured to meet the above requirements. The essence of multi-tenant Layer 3 packet isolation is to utilize VRF instances to isolate Layer 3 packet forwarding. You can specify a VRF instance for the overlay router interface of each VXLAN subnet owned by a tenant to define the L3 forwarding range of the VXLAN subnet. This method of isolating multiple tenants by VRF instance not only ensures the independence and security of tenants' networks, but also reserves enough hierarchical space for internal network classification of tenants.

3.7.2 Restrictions and Guidelines

Perform the configuration on VXLAN gateways as required.

3.7.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a single-protocol VRF instance and enter the VRF instance configuration mode.

ip vrf *vrf-name*

No single-protocol VRF instance is configured by default.

(4) Define the RD value for the VRF instance.

rd *rd-value*

No RD value is configured by default.

(5) Configure the RT attribute for the VRF instance.

route-target { **import** | **export** | **both** } *route-target-value*

No RT value is defined by default.

(6) Enter the overlay routed interface configuration mode.

interface overlayrouter *interface-number*

(7) Add the overlay router interface to the single-protocol VRF instance.

ip vrf binding *vrf-name*

No interface is added to any single-protocol VRF by default.

3.8 Configuring the EVPN Synchronization Attribute

3.8.1 Overview

The EVPN standard defines five types of routes. VTEPs mutually synchronize five types of EVPN routes to automatically discover VTEPs, establish VXLAN tunnels, advertise host addresses and routes, and implement multi-homing access. Type-2 routes are used by an EVPN to advertise host information and are classified into MAC-only routes (for advertising host MAC addresses) and MAC/IP routes (for advertising host MAC addresses, host ARP information, host IPv6 ND information, and host routes).

By configuring the EVPN synchronization attribute, you can specify whether a VXLAN instance extracts entry information from type-2 route packets of EVPN.

3.8.2 Restrictions and Guidelines

The following commands can be configured only on Layer 2 VNIs (that is, VXLAN instances with the **symmetric** command not configured). They are unavailable on Layer 3 VNIs.

When you run the following commands on a VXLAN instance, only the EVPN entry resolution of the VXLAN instance is affected, and other VXLAN instances with the commands not configured are not affected.

- **evpn arp mac-learning enable**
  - When this command is configured, the device resolves one ARP entry and one MAC entry from a MAC-IP type-2 route synchronized from a VXLAN-EVPN neighbor.
  - By default, this command is not configured and the device resolves only one ARP entry but no MAC entry from a MAC-IP type-2 route synchronized from a VXLAN-EVPN neighbor.
  - This command can be used together with the **evpn mac advertise disable** command. That is, the network-wide VXLAN-EVPN neighbors synchronize only the MAC-IP type-2 routes but no MAC-only type-2 routes. All devices resolve and extract MAC entries from MAC-IP type-2 routes.
- **evpn nd mac-learning enable**
  - After this command is configured, the device resolves one IPv6 ND entry and one MAC entry from a MAC-IP type-2 route (IPv6 ND) synchronized from a VXLAN-EVPN neighbor.
  - By default, this command is not configured and the device resolves only one IPv6 ND entry but no MAC entry from a MAC-IP type-2 route synchronized from a VXLAN-EVPN neighbor.
  - This command can be used together with the **evpn mac advertise disable** command. That is, the network-wide VXLAN-EVPN neighbors synchronize only the MAC-IP type-2 routes but no MAC-only type-2 routes. All devices resolve and extract MAC entries from MAC-IP type-2 routes.
- **evpn mac advertise disable**
  - This command is not configured on a device by default. The device generates one MAC-only type-2 route based on a locally learned MAC entry according to the VXLAN-EVPN protocol, and synchronizes the type-2 route to an EVPN neighbor (that is, remote VTEP). Then, the remote VTEP learns the MAC entry from the MAC-only type-2 route.
  - After this command is configured, the device does not generate a VXLAN-EVPN MAC-only type-2 route based on the MAC entry. Therefore, the device will not advertise MAC-only type 2 routes to EVPN neighbors.
  - This command can be used together with the **evpn arp mac-learning enable** command. That is, the network-wide VXLAN-EVPN neighbors synchronize only the MAC-IP type-2 routes but no MAC-only type-2 routes. All devices resolve and extract MAC entries from MAC-IP type-2 routes.
- **evpn mac inactive**
  - After this command is configured, the device will not learn MAC entries from VXLAN-EVPN type-2 routes (MAC-IP or MAC-only type-2 routes) synchronized from neighbors.
  - This command is not configured on a device by default. The device learns MAC entries from VXLAN-EVPN type-2 routes synchronized from neighbors.
- **evpn rt-2 advertise disable**
  - This command is not configured on a device by default. The device generates one MAC-only type-2 route based on a locally learned MAC entry according to the VXLAN-EVPN protocol, and synchronizes the type-2 route to an EVPN neighbor (that is, remote VTEP). Then, the remote VTEP learns the MAC entry from the MAC-only type-2 route. In addition, the device generates one MAC-IP type-2 route based on a locally learned ARP entry according to the VXLAN-EVPN protocol and synchronizes the type-2 route to an EVPN neighbor. Then, the remote VTEP learns the ARP entry and host route from the MAC-IP type-2 route.
  - After this command is configured, the device does not generate VXLAN-EVPN type-2 routes (MAC-IP or

MAC-only routes) based on MAC entries and ARP entries and therefore, no type-2 routes are advertised to EVPN neighbors.

3.8.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VXLAN configuration mode.

vxlan *vni-number*

(4) Configure the EVPN synchronization attribute. Configure at least one of the tasks.

- Configure the function of extracting MAC entries from EVPN MAC-IP type-2 routes (ARP information).

  evpn arp mac-learning enable

  The function of extracting MAC entries from EVPN MAC-IP type-2 routes (ARP information) is not configured by default.

- Configure the function of extracting MAC entries from EVPN MAC-IP type-2 routes (IPv6 ND information).

  evpn nd mac-learning enable

  The function of extracting MAC entries from EVPN MAC-IP type-2 routes (IPv6 ND information) is not configured by default.

- Configure a Layer 2 VNI to disable the synchronization of the local MAC address to the remote VTEP through EVPN messages.

  evpn mac advertise disable

  A Layer 2 VNI is not configured to disable the synchronization of the local MAC address to the remote VTEP through EVPN messages by default.

- Configure a Layer 2 VNI to disable the delivery of MAC addresses remotely synchronized through EVPN messages to the local MAC address table.

  evpn mac inactive

  A Layer 2 VNI is not configured to disable the delivery of MAC addresses remotely synchronized through EVPN messages to the local MAC address table.

- Configure a Layer 2 VNI to stop generating EVPN type-2 routes.

  evpn rt-2 advertise disable

  A Layer 2 VNI is not configured to stop generating EVPN type-2 routes (MAC+ARP entries) by default.

3.9 Configuring Rate Limits for a Tunnel Interface

3.9.1 Overview

This section describes how to configure the input/output rate limits for a dynamic tunnel. If you need to limit the rate of a dynamic tunnel, configure the input/output rate limits.

3.9.2 Restrictions and Guidelines

N/A

3.9.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VTEP configuration mode.

vtep

(4) Configure the rate limits for a tunnel interface.

vxlan overlaytunnel dip *ipv4-address* **speed-limit** { **input** | **output** } *limit-per-second burst-size*

No rate limit is configured for a dynamic tunnel by default.

3.10 Configuring ARP/ND Suppression

3.10.1 Overview

This section describes how to configure global ARP/ND suppression or VNI-based ARP/ND suppression. ARP/ND suppression is generally enabled on VXLAN access devices in centralized deployment scenarios or on VXLAN gateways in distributed deployment scenarios.

- After ARP suppression is enabled, the VTEP responds to ARP requests from hosts as a proxy, to reduce ARP data flooding in the network.
- After IPv6 ND suppression is enabled, the VTEP responds to NS multicast packets from hosts as a proxy, to reduce NS multicast packet flooding in the network.

3.10.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure ARP/ND suppression. Configure at least one of the tasks.

- Run the following commands in sequence to configure global ARP/ND suppression.

  vtep

  Configure global ARP/ND suppression.

  (IPv4 network)

  arp suppress enable

  No global ARP suppression is configured by default.

  (IPv6 network)

  nd suppress enable

No global IPv6 ND suppression is configured by default.

- Run the following commands in sequence to configure VNI-based ARP/ND suppression.

  vxlan *vni-number*

  No VXLAN instance is configured by default.

  (IPv4 network)

  arp suppress enable

  No VNI-based ARP suppression is configured by default.

  (IPv6 network)

  nd suppress enable

  No VNI-based IPv6 ND suppression is configured by default.

3.11 Configuring ARP/ND Proxy

3.11.1 Overview

The ARP/ND proxy function implements the mutual access between hosts in the same VXLAN through gateway forwarding. After ARP/ND proxy is enabled, a VTEP responds to ARP/NS requests from hosts as a proxy and the MAC address replied by the proxy is that of the gateway on the VTEP. Therefore, the MAC address returned for ARP/NS requests is that of the VTEP, and the access traffic between hosts in the same VNI is implemented through Layer 3 forwarding.

3.11.2 Restrictions and Guidelines

- The **route-in-vni** command configures the ARP proxy and ND proxy functions.
- The ARP/ND proxy function enabled on an overlay router interface takes effect only after the ARP/ND suppression is enabled.
- ARP/ND proxy can be enabled only on VXLAN gateways, and generally on VXLAN gateways in distributed deployment scenarios.

3.11.3 Prerequisites

ARP/ND suppression has been enabled before the **route-in-vni** command is configured.

3.11.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the overlay router interface configuration mode.

interface overlayrouter *interface-number*

(4) Configure ARP/ND proxy.

route-in-vni

No ARP/ND proxy is configured by default.

(5) Return to the global configuration mode.

end

(6) Enter the VXLAN configuration mode.

vxlan *vni-number*

No VXLAN instance is configured by default.

(7) (Optional) Disable VLAN check in the case of ARP/ND proxy.

proxy check vlan disable

VLAN check is enabled in the case of ARP/ND proxy by default.

After VLAN check is enabled and the device serves as an ARP/ND proxy, if the Layer 2 egress and VLAN of a host requested by an ARP/NS packet are the same as those in the ARP/NS packet, the device discards the ARP/NS request and does not respond as a proxy.

3.12 Creating VXLAN Instances in Batches

3.12.1 Overview

This section describes how to create VXLAN instances in batches or enter the batch VXLAN instance configuration mode.

3.12.2 Restrictions and Guidelines

- If the number of VXLAN instances reaches the maximum value, the configuration fails.
- When configuring multiple VIDs, separate the VIDs with commas (,) or connect the first and last VIDs of continuous VIDs with a hyphen (-).

3.12.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create VXLAN instances in batches and enter the batch VXLAN instance configuration mode.

vxlan range *vxlan-list*

No VXLAN instance is configured by default.

3.13 Configuring a Data Center Interconnection Tunnel

3.13.1 Overview

A data center interconnection tunnel controls Layer 2 forwarding of packets between an internal tunnel and an interconnection tunnel of a data center. It does not affect Layer 3 forwarding. The packet processing behavior of the VXLAN gateway is as follows:

- After receiving packets from an internal tunnel of a data center, the gateway forwards them to the interconnection tunnel of the data center at Layer 2.

- After receiving packets from the interconnection tunnel of the data center, the gateway forwards the packets to an internal tunnel of the data center at Layer 2, but will not forward the packets to interconnection tunnels of other data centers.

3.13.2 Restrictions and Guidelines

- This function is usually configured on edge devices that connect data centers.
- Multiple data center interconnection tunnels can be configured on one device.

3.13.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VTEP configuration mode.

vtep

(4) Configure a data center interconnection tunnel.

vxlan outside center vtep-ip *ipv4-address*

No data center interconnection tunnel is configured by default.

3.14 Configuring VNI Mapping

3.14.1 Overview

In a scenario where multiple data centers are interworked, data centers can be deployed in the following mode: Each data center (hereinafter referred to as a point of delivery (POD)) maintains one independent VNI, and one public VNI is used for interaction across data centers. This deployment mode implements more flexible VNI deployment and maintenance in each center.

In this deployment mode, when an access request packet across data centers is sent out of a center, the VNI of the center is replaced with the cross-center public VNI in the packet. When the request reaches the peer center, the public VNI is replaced with the VNI of a related subnet in the peer center. The VNI replacement across data centers is called VNI mapping.

VNI mapping is classified into Layer 2 VNI mapping and Layer 3 VNI mapping based on the mapping method.

Layer 2 VNI mapping: Used for VNI replacement for cross-center Layer 2 traffic. The Layer 2 VNI of a data center is replaced with the public Layer 2 VNI across data centers.

Layer 3 VNI mapping: Used for VNI replacement for cross-center Layer 3 traffic. The Layer 3 VNI (VNI of a symmetric VXLAN instance) of a data center is replaced with the public Layer 3 VNI (VNI of a symmetric VXLAN instance) across data centers.

3.14.2 Restrictions and Guidelines

The function is explained as follows for your better understanding: If the **mapping vni** VNI-b or **l3mapping vni** VNI-b command is configured in a VNI-a instance, VNI-a is the original VNI and VNI-b is the post-mapping VNI.

- Data center interconnection based on VNI mapping is usually configured on edge devices that connect data

centers.

- After data center interconnection based on VNI mapping is configured, when receiving an internal VXLAN packet from a data center, the device replaces the original VNI in the packet with the post-mapping VNI. When receiving a VXLAN packet from another data center, the device replaces the post-mapping VNI in the packet with the original VNI.
- After data center interconnection based on VNI mapping is configured, the original VNI/Layer 3 VNI carried in the EVPN routes synchronized by the device to other data centers through BGP will be replaced with the post-mapping VNI/Layer 3 VNI.
- Data center interconnection based on VNI mapping takes effect only when both the original VNI and post-mapping VNI exist.
- Mapping cannot be configured on a post-mapping VNI.

3.14.3 Prerequisites

A data center interconnection tunnel has been configured before VNI mapping is configured.

3.14.4 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Create a VXLAN instance and enter the VXLAN configuration mode.

vxlan *vni-number*

No VXLAN instance is configured by default.

(4) Configure VNI mapping. Configure one of the following tasks.

- Configure a Layer 2 VNI mapping.

  mapping vni *vni-number*

  No post-mapping VNI is configured for a VXLAN instance.

  This command can be configured only on Layer 2 VNIs (instances with the **symmetric** command not configured), and post-mapping VNIs can be only Layer 2 VNIs.

  Configuring post-mapping VNIs and configuring associated VLANs are mutually exclusive on VXLANs. That is, a post-mapping VNI cannot be associated with a VLAN, and a VXLAN instance associated with a VLAN cannot be configured as a post-mapping VNI.

  Original VNIs and post-mapping VNIs are in one-to-one mappings.

- Configure an L3VNI mapping.

  l3mapping vni *vni-number*

  No post-mapping VNI is configured for symmetric instances by default.

  This command can be configured only on Layer 3 VNIs (instances configured with the **symmetric** command). That is, the original VNI must be an Layer 3 VNI and the post-mapping VNI can be an Layer 3 VNI or an Layer 2 VNI.

This command supports the mapping from one original VNI to multiple VNIs but one VNI cannot configured as the post-mapping VNI of multiple VNIs.

3.15 Configuring a Layer 2 Ethernet Sub-interface to Access a VXLAN

3.15.1 Overview

This section describes how to configure hosts to access a VXLAN through Layer 2 Ethernet sub-interfaces. Layer 2 Ethernet sub-interfaces can access a VXLAN in untagged VLAN encapsulation mode, tagged VLAN encapsulation mode, and QinQ encapsulation mode.

3.15.2 Restrictions and Guidelines

- The **extend-vlan** command configured for a VXLAN instance uses the same encapsulation rule as that configured for a sub-interface (tagged VLAN or untagged VLAN encapsulation). When a packet matches the encapsulation rule of a sub-interface, the encapsulation rule priority of the sub-interface is higher than that of **extend-vlan** configured for a VXLAN instance.
- Perform the configuration on VXLAN access devices based on actual networking conditions.

3.15.3 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the Layer 2 Ethernet sub-interface configuration mode.

interface *interface-type interface-number.subnumber*

(4) Configure VXLAN encapsulation rules for the sub-interface.

encapsulation vxlan *vni-number*

Configure VXLAN encapsulation rules for the specified sub-interface.

No VXLAN encapsulation rule is configured for a sub-interface by default.

(5) Configure the Layer 2 Ethernet sub-interface to access the VXLAN. Configure either of the following tasks.

- Configure the untagged or VLAN encapsulation rule for the sub-interface.

  encapsulation dot1q { **untag** | **s-vid** *vlan-id* }

  No untagged and VLAN encapsulation rules are configured for a sub-interface by default.

- Configure the QinQ encapsulation rule for the sub-interface.

  encapsulation qinq s-vid *svlan-id* **c-vid** *cvlan-id*

  No QinQ encapsulation rule is configured for a sub-interface by default.

3.16 Configuring a Delay in the Protocol Status Change of a VXLAN Tunnel

3.16.1 Overview

This section describes how to configure a delay in the protocol status change of a VXLAN tunnel from down state to up state.

3.16.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VTEP configuration mode.

vtep

(4) Configure a delay in the protocol status change of a VXLAN tunnel.

vxlan overlaytunnel carrier-delaytime up *delay-time*

The default delay in the protocol status change of VXLAN tunnels is 2000 ms.

3.17 Configuring the Optimal Minimum Route Length for VXLAN Tunnels

3.17.1 Overview

You can configure the optimal minimum IPv4 route length for tunnels to control the optimal IPv4 DIP routes for the tunnels.

3.17.2 Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enter the VTEP configuration mode.

vtep

(4) Configure the optimal minimum IPv4/IPv6 route length for tunnels.

vxlan overlaytunnel nexthop resolution prefix-length minimum *minimum-length-ipv4*

The default optimal minimum IPv4 route length of tunnels is **0**.

4 Monitoring

Run the **display** command to check the running status of a configured function to verify the configuration effect.

Run the **cleanup** command to clear information.

⚠ Caution

Running the **cleanup** command may interrupt services due to loss of important information.

Table 4-1 VXLAN Monitoring

Command	Purpose
display vxlan *vni-number*	Displays the VXLAN configuration and status of a device.
display vxlan mac-addr-table [**address** *mac-address*] [**vni** *vni-number*] [**remote** | **local**]	Displays the MAC addresses learned by the VXLAN configured on the device.
display vxlan mac-active [**vni** *vni-number*]	Displays active VXLAN MAC entries obtained by the VXLAN protocol module from the local device.
display vxlan mac-bgp [**vni** *vni-number*]	Displays remote VXLAN MAC entries synchronized via BGP.
display vxlan mac-log { **all** | **vni** *vni-number* | **mac** *mac-address* | **vni** *vni-number* **mac** *mac-address* | **status** }	Displays operation logs of VXLAN MAC entries.
display arp vxlan	Displays ARP entries learned by the VXLAN module of the device.
display ipv6 neighbors-table	Displays IPv6 ND entries learned by the VXLAN module of the device.
display vxlan global	Displays the VTEP-IP address and anycast MAC address of the device.
display vxlan route-active [**count** | **vrf** *vrf-name* | **global-vrf** | **vni** *vni-number* | **local** | **remote**]	Displays active VXLAN ARP entries.
display vxlan route-active ipv6 [**count** | **vrf** *vrf-name* | **global-vrf** | **vni** *vni-number* | **local** | **remote**]	Displays active VXLAN IPv6 ND entries.
display vxlan vrf { **all** | **global-vrf** | **vrf-name** *vrf-name* }	Displays VXLAN instances of a VRF instance.

Command	Purpose
display vxlan arp suppress	Displays the enabling status of VXLAN ARP suppression on the device.
display vxlan nd suppress	Displays the enabling status of VXLAN IPv6 ND suppression on the device.
display vxlan udp-port	Displays the VXLAN UDP destination port of the device.
display vxlan vni-statistics { **all** | **vni** *vni-number* }	Displays traffic data of a VNI.
display interface	Displays the traffic statistics of a tunnel interface.
statistics enable	Enables traffic statistics collection on a Layer 2 Ethernet sub-interface in interface configuration mode.
display interface	Displays traffic statistics of a Layer 2 Ethernet sub-interface.
display vxlan debugging detail	Displays specified VXLAN debugging information.
display vxlan ethernet { **interface** | **vlan** [*vlan-id*] | **vni** [*vni-number*] }	Displays VXLAN Ethernet interface information.
display vxlan log	Displays key run logs of the VXLAN process.
display vxlan overlaytunnel route [**interface** *interface-number*] [detail]	Displays routing information of a tunnel.
display vxlan route-bgp [**count** | **vrf** vrf-name | **global-vrf** | **vni** *vni-number*]	Displays VXLAN ARP entries synchronized via BGP.
display vxlan route-bgp ipv6 [**count** | **vrf** vrf-name | **global-vrf** | **vni** *vni-number*]	Displays VXLAN ND entries synchronized via BGP.
display vxlan route-log { **all** | **vni** *vni-number* | **ip** *ipv4-address* | **ipv6** *ipv6-address* | **vni** *vni-number* { **ip** *ipv4-address* | **ipv6** *ipv6-address* } | **status** }	Displays operation logs of VXLAN ARP/IPv6 ND entries.
display vxlan server [**database** | **router** | **tunnel** | **client-info** *client-name*]	Displays details about the VXLAN process.
display vxlan storm-control-enable [**vni** *vni-number*]	Displays the configuration of VNI-based storm control.
display vxlan subport { **all** | **vni** *vni-number* | **interface** *interface-type interface-number* }	Displays VXLAN Layer 2 sub-interface information.
display vxlan vtep-log { **all** | **vni** *vni-number* | **ip** *ip-address* | **vni** *vni-number* **ip** *ip-address* | **status** }	Displays operation logs related to a VXLAN tunnel.

Command	Purpose
display vxlan vtep-nbr [**dip** *ipv4-address*]	Displays VTEP neighbor information.
cleanup vxlan config all	Clears all VXLAN-related configuration.
cleanup vxlan mac dynamic [**vni** *vni-number* | **address** *mac-address* **vni** *vni-number*] [**force**]	Clears VXLAN MAC entries.
cleanup vxlan vni-statistics	Clears VNI traffic data.
cleanup vxlan log	Clears key run logs of the VXLAN process.
vni-counting enable	Enables VNI traffic statistics collection (globally or on a VXLAN instance).
vni-counting period sample *sample-time* **load** *load-time*	Configures the sampling period and calculation period of VNI traffic statistics collection.
vxlan overlaytunnel counting enable	Enables tunnel traffic statistics collection globally.
vxlan vtep-log { **disable** | **limit** *limit-num* | **type** { **event** | **nstime** | **timestamp** } }	Configures VXLAN tunnel logging parameters.
vxlan route-log { **disable** | **limit** *limit-num* | **type** { **event** | **nstime** | **timestamp** } }	Configures VXLAN ARP/ND entry logging.
vxlan mac-log { **disable** | **limit** *limit-num* | **type** { **event** | **nstime** | **timestamp** } }	Configures VXLAN MAC entry logging.

5 Configuration Examples

5.1 Configuring a VXLAN SDN

5.1.1 Requirements

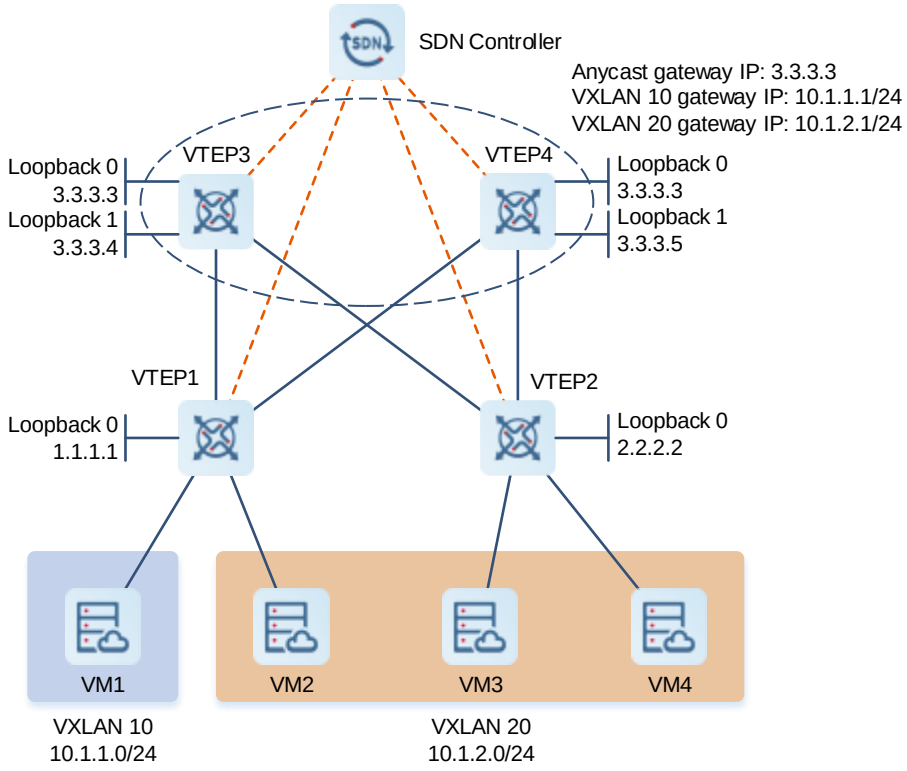
In the topology as shown in Figure 5-1, the following requirements need to be met:

- VMs on the network are classified into several VXLAN subnets, which can communicate with each other through Layer 3 forwarding.
- The SDN controller is used to manage and maintain VXLAN configuration.
- Fast convergence can be performed on VXLAN traffic when the gateway is faulty.

Centralized VXLAN networking using static tunnels needs to be configured in this topology.

5.1.2 Topology

Figure 5-1 Configuring a VXLAN SDN



5.1.3 Notes

- On VTEP 1, VTEP 2, VTEP 3, and VTEP 4, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that unicast routes between the devices are reachable (omitted).
- On VTEP 1, VTEP 2, VTEP 3, and VTEP 4, configure a loopback interface and configure an IP address for the loopback interface. The loopback interfaces distribute packets using the unicast routing protocol. The IP addresses of CORE-1 and CORE-2 are the same.
- On VTEP 3 and VTEP 4, configure the anycast gateway, the same VXLAN gateway, and the same VTEP IP address.
- Configure the VXLAN and gateway address on each VM (omitted).
- Create a BGP neighbor relationship between VTEP 3 and VTEP 4 and configure BGP-EVPN.

The following content can be manually configured or delivered by the SDN controller. In this typical configuration example, manual configuration is adopted.

- On VTEP 1, create VXLAN instances VXLAN 10 and VXLAN 20, and associate them with VLAN 10 and VLAN 20 respectively. Disable MAC address auto-learning. Configure two VXLAN tunnels to connect to VTEP 2 and the anycast gateway. Associate VXLAN 10 and VXLAN 20 with the two tunnels respectively.
- On VTEP 2, create a VXLAN instance VXLAN 20 and associate it with VLAN 20. Disable MAC address auto-learning. Configure two VXLAN tunnels to connect to VTEP 1 and anycast gateway. Associate VXLAN 20 with the two tunnels.

- On VTEP 3, create VXLAN instances VXLAN 10 and VXLAN 20. Disable MAC address auto-learning. Configure the anycast MAC address. Configure two overlay router interface and set their IP addresses to 10.1.1.1/24 and 10.1.2.1/24 respectively. Associate VXLAN 10 with the overlay router interface whose IP address is 10.1.1.1/24. Associate VXLAN 20 with the overlay router interface whose IP address is 10.1.2.1/24. Configure two VXLAN tunnels to connect to VTEP 1 and VTEP 2 respectively. Associate VXLAN 10 and VXLAN 20 with the two tunnels respectively.
- The configurations on VTEP 4 are basically the same as those on VTEP 3.
- On VTEP 3 and VTEP 4, configure Loopback 1 and set the IP address to 3.3.3.4 and 3.3.3.5 respectively. Configure the two devices to establish a BGP neighbor relationship through Loopback 1 and activate the BGP neighbor relationship in the L2VPN EVPN address family.

5.1.4 Procedure

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 0
VTEP1(cfg-if-Loopback 0)# ip address 1.1.1.1 255.255.255.255
VTEP1(cfg-if-Loopback 0)# exit
VTEP1(cfg)# interface overlaytunnel 1
VTEP1(cfg-if-OverlayTunnel 1)# overlaytunnel source 1.1.1.1
VTEP1(cfg-if-OverlayTunnel 1)# overlaytunnel destination 2.2.2.2
VTEP1(cfg-if-OverlayTunnel 1)# exit
VTEP1(cfg)# interface overlaytunnel 2
VTEP1(cfg-if-OverlayTunnel 2)# overlaytunnel source 1.1.1.1
VTEP1(cfg-if-OverlayTunnel 2)# overlaytunnel destination 3.3.3.3
VTEP1(cfg-if-OverlayTunnel 2)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP1(cfg-vxlan)# tunnel-interface overlaytunnel 2
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# end
VTEP1(cfg)# vxlan 20
VTEP1(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP1(cfg-vxlan)# tunnel-interface overlaytunnel 2
VTEP1(cfg-vxlan)# extend-vlan 20
VTEP1(cfg-vxlan)# end
```

(2) Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP2(cfg)# interface loopback 0
VTEP2(cfg-if-Loopback 0)# ip address 2.2.2.2 255.255.255.255
VTEP2(cfg-if-Loopback 0)# exit
VTEP2(cfg)# interface overlaytunnel 1
VTEP2(cfg-if-OverlayTunnel 1)# overlaytunnel source 2.2.2.2
```

```
VTEP2(cfg-if-OverlayTunnel 1)# overlaytunnel destination 1.1.1.1
VTEP2(cfg-if-OverlayTunnel 1)# exit
VTEP2(cfg)# interface overlaytunnel 2
VTEP2(cfg-if-OverlayTunnel 2)# overlaytunnel source 2.2.2.2
VTEP2(cfg-if-OverlayTunnel 2)# overlaytunnel destination 3.3.3.3
VTEP2(cfg-if-OverlayTunnel 2)# exit
VTEP2(cfg)# vxlan 20
VTEP2(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP2(cfg-vxlan)# tunnel-interface overlaytunnel 2
VTEP2(cfg-vxlan)# extend-vlan 20
VTEP2(cfg-vxlan)# end
```

(3) Configure VTEP 3.

```
VTEP3> privilege
VTEP3# config-system terminal
VTEP3(cfg)# fabric anycast-gateway-mac 0000.1234.5678
VTEP3(cfg)# vtep
VTEP3(cfg-vtep)# vxlan gateway synchronize
VTEP3(cfg-vtep)# exit
VTEP3(cfg)# interface loopback 0
VTEP3(cfg-if-Loopback 0)# ip address 3.3.3.3 255.255.255.255
VTEP3(cfg-if-Loopback 0)# exit
VTEP3(cfg)# interface loopback 1
VTEP3(cfg-if-Loopback 1)# ip address 3.3.3.4 255.255.255.255
VTEP3(cfg-if-Loopback 1)# exit
VTEP3(cfg)# router-protocol bgp 10000
VTEP3(cfg-router)# peer 3.3.3.5 remote-as 10000
VTEP3(cfg-router)# peer 3.3.3.5 update-source loopback 1
VTEP3(cfg-router)# address-family-mode l2vpn evpn
VTEP3(cfg-router-af)# peer 3.3.3.5 activate
VTEP3(cfg-router-af)# peer 3.3.3.5 send-community extended
VTEP3(cfg-router-af)# exit-address-family-mode
VTEP3(cfg-router)# exit
VTEP3(cfg)# interface overlaytunnel 1
VTEP3(cfg-if-OverlayTunnel 1)# overlaytunnel source 3.3.3.3
VTEP3(cfg-if-OverlayTunnel 1)# overlaytunnel destination 2.2.2.2
VTEP3(cfg-if-OverlayTunnel 1)# exit
VTEP3(cfg)# interface overlaytunnel 2
VTEP3(cfg-if-OverlayTunnel 2)# overlaytunnel source 3.3.3.3
VTEP3(cfg-if-OverlayTunnel 2)# overlaytunnel destination 1.1.1.1
VTEP3(cfg-if-OverlayTunnel 2)# exit
VTEP3(cfg)# interface overlayrouter 1
VTEP3(cfg-if-OverlayRouter 1)# ip address 10.1.1.1/24
VTEP3(cfg-if-OverlayRouter 1)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 1)# exit
VTEP3(cfg)# interface overlayrouter 2
VTEP3(cfg-if-OverlayRouter 2)# ip address 10.1.2.1/24
```

```
VTEP3(cfg-if-OverlayRouter 2)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 2)# exit
VTEP3(cfg)# vxlan 10
VTEP3(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP3(cfg-vxlan)# tunnel-interface overlaytunnel 2
VTEP3(cfg-vxlan)# router-interface overlayrouter 1
VTEP3(cfg-vxlan)# end
VTEP3(cfg)# vxlan 20
VTEP3(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP3(cfg-vxlan)# tunnel-interface overlaytunnel 2
VTEP3(cfg-vxlan)# router-interface overlayrouter 2
VTEP3(cfg-vxlan)# end
```

(4) Configure VTEP 4.

```
VTEP4> privilege
VTEP4# config-system terminal
VTEP4(cfg)# fabric anycast-gateway-mac 0000.1234.5678
VTEP4(cfg)# vtep
VTEP4(cfg-vtep)# vxlan gateway synchronize
VTEP4(cfg-vtep)# exit
VTEP4(cfg)# interface loopback 0
VTEP4(cfg-if-Loopback 0)# ip address 3.3.3.3 255.255.255.255
VTEP4(cfg-if-Loopback 0)# exit
VTEP4(cfg)# interface loopback 1
VTEP4(cfg-if-Loopback 1)# ip address 3.3.3.5 255.255.255.255
VTEP4(cfg-if-Loopback 1)# exit
VTEP4(cfg)# router-protocol bgp 10000
VTEP4(cfg-router)# peer 3.3.3.4 remote-as 10000
VTEP4(cfg-router)# peer 3.3.3.4 update-source loopback 1
VTEP4(cfg-router)# address-family-mode l2vpn evpn
VTEP4(cfg-router-af)# peer 3.3.3.4 activate
VTEP4(cfg-router-af)# peer 3.3.3.4 send-community extended
VTEP4(cfg-router-af)# exit-address-family-mode
VTEP4(cfg-router)#exit
VTEP4(cfg)# interface overlaytunnel 1
VTEP4(cfg-if-OverlayTunnel 1)# overlaytunnel source 3.3.3.3
VTEP4(cfg-if-OverlayTunnel 1)# overlaytunnel destination 2.2.2.2
VTEP4(cfg-if-OverlayTunnel 1)# exit
VTEP4(cfg)# interface overlaytunnel 2
VTEP4(cfg-if-OverlayTunnel 2)# overlaytunnel source 3.3.3.3
VTEP4(cfg-if-OverlayTunnel 2)# overlaytunnel destination 1.1.1.1
VTEP4(cfg-if-OverlayTunnel 2)# exit
VTEP4(cfg)# interface overlayrouter 1
VTEP4(cfg-if-OverlayRouter 1)# ip address 10.1.1.1/24
VTEP4(cfg-if-OverlayRouter 1)# anycast-gateway
VTEP4(cfg-if-OverlayRouter 1)# exit
VTEP4(cfg)# interface overlayrouter 2
```

```
VTEP4(cfg-if-OverlayRouter 2)# ip address 10.1.2.1/24
VTEP4(cfg-if-OverlayRouter 2)# anycast-gateway
VTEP4(cfg-if-OverlayRouter 2)# exit
VTEP4(cfg)# vxlan 10
VTEP4(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP4(cfg-vxlan)# tunnel-interface overlaytunnel 2
VTEP4(cfg-vxlan)# router-interface overlayrouter 1
VTEP4(cfg-vxlan)# end
VTEP4(cfg)# vxlan 20
VTEP4(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP4(cfg-vxlan)# tunnel-interface overlaytunnel 2
VTEP4(cfg-vxlan)# router-interface overlayrouter 2
VTEP4(cfg-vxlan)# end
```

5.1.5 Verification

- After the commands above are run and the centralized VXLAN networking using static tunnels is configured, VM 1, VM 2, VM 3, and VM 4 can mutually ping each other.
- Run the **display vxlan** command to display the VXLAN instance and tunnel configuration on the devices. The following uses VTEP 1 as an example.

```
VTEP1# display vxlan
VXLAN Total Count: 2
VXLAN Capacity     : 8000

VXLAN 10
      Symmetric property    : FALSE
      Routed interface      : -
      Extend VLAN           : 10
      VTEP Adjacency Count: 2
      VTEP Adjacency List :
      Interface                 Source IP                                   Destination IP
Type
      ----------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 1           1.1.1.1                                     2.2.2.2
static
      OverlayTunnel 2           1.1.1.1                                     3.3.3.3
static

VXLAN 20
      Symmetric property    : FALSE
      Routed interface      : -
      Extend VLAN           : 20
      VTEP Adjacency Count: 2
      VTEP Adjacency List :
      Interface                 Source IP                                   Destination IP
Type
```

```
    ---------------------- ------------------------------------------- ------------------------------------------- -------
    OverlayTunnel 1        1.1.1.1                                     2.2.2.2
static
    OverlayTunnel 2        1.1.1.1                                     3.3.3.3
static
```

5.1.6 Configuration Files

- VTEP 1 configuration file

```
!
interface Loopback 0
 ip address 1.1.1.1 255.255.255.255
!
interface OverlayTunnel 1
 overlaytunnel source 1.1.1.1
 overlaytunnel destination 2.2.2.2
!
interface OverlayTunnel 2
 overlaytunnel source 1.1.1.1
 overlaytunnel destination 3.3.3.3
!
vxlan 10
 tunnel-interface OverlayTunnel 1
 tunnel-interface OverlayTunnel 2
 extend-vlan 10
!
vxlan 20
 tunnel-interface OverlayTunnel 1
 tunnel-interface OverlayTunnel 2
 extend-vlan 20
!
```

- VTEP 2 configuration file

```
!
interface Loopback 0
 ip address 2.2.2.2 255.255.255.255
!
interface OverlayTunnel 1
 overlaytunnel source 2.2.2.2
 overlaytunnel destination 1.1.1.1
!
interface OverlayTunnel 2
 overlaytunnel source 2.2.2.2
 overlaytunnel destination 3.3.3.3
!
vxlan 10
 tunnel-interface OverlayTunnel 1
```

```
 tunnel-interface OverlayTunnel 2
 extend-vlan 10
!
vxlan 20
 tunnel-interface OverlayTunnel 1
 tunnel-interface OverlayTunnel 2
 extend-vlan 20
!
```

- VTEP 3 configuration file

```
!
fabric anycast-gateway-mac 0000.1234.5678
!
interface Loopback 0
 ip address 3.3.3.3 255.255.255.255
!
interface Loopback 1
 ip address 3.3.3.4 255.255.255.255
!
interface OverlayTunnel 1
 overlaytunnel source 3.3.3.3
 overlaytunnel destination 2.2.2.2
!
interface OverlayTunnel 2
 overlaytunnel source 3.3.3.3
 overlaytunnel destination 1.1.1.1
!
interface OverlayRouter 1
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 2
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
!
router-protocol bgp 10000
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 3.3.3.5 remote-as 10000
 peer 3.3.3.5 update-source Loopback 1
 address-family-mode ipv4
  peer 3.3.3.5 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 3.3.3.5 activate
```

```
  peer 3.3.3.5 send-community extended
  exit-address-family-mode
!
vxlan 10
 router-interface OverlayRouter 1
 tunnel-interface OverlayTunnel 1
 tunnel-interface OverlayTunnel 2
!
vxlan 20
 router-interface OverlayRouter 2
 tunnel-interface OverlayTunnel 1
 tunnel-interface OverlayTunnel 2
!
```

- VTEP 4 configuration file

```
!
fabric anycast-gateway-mac 0000.1234.5678
!
interface Loopback 0
 ip address 3.3.3.3 255.255.255.255
!
interface Loopback 1
 ip address 3.3.3.5 255.255.255.255
!
interface OverlayTunnel 1
 overlaytunnel source 3.3.3.3
 overlaytunnel destination 2.2.2.2
!
interface OverlayTunnel 2
 overlaytunnel source 3.3.3.3
 overlaytunnel destination 1.1.1.1
!
interface OverlayRouter 1
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 2
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
!
router-protocol bgp 10000
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 3.3.3.4 remote-as 10000
 peer 3.3.3.4 update-source Loopback 1
```

```
 address-family-mode ipv4
  peer 3.3.3.4 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 3.3.3.4 activate
  peer 3.3.3.4 send-community extended
  exit-address-family-mode
!
vxlan 10
 router-interface OverlayRouter 1
 tunnel-interface OverlayTunnel 1
 tunnel-interface OverlayTunnel 2
!
vxlan 20
 router-interface OverlayRouter 2
 tunnel-interface OverlayTunnel 1
 tunnel-interface OverlayTunnel 2
!
```

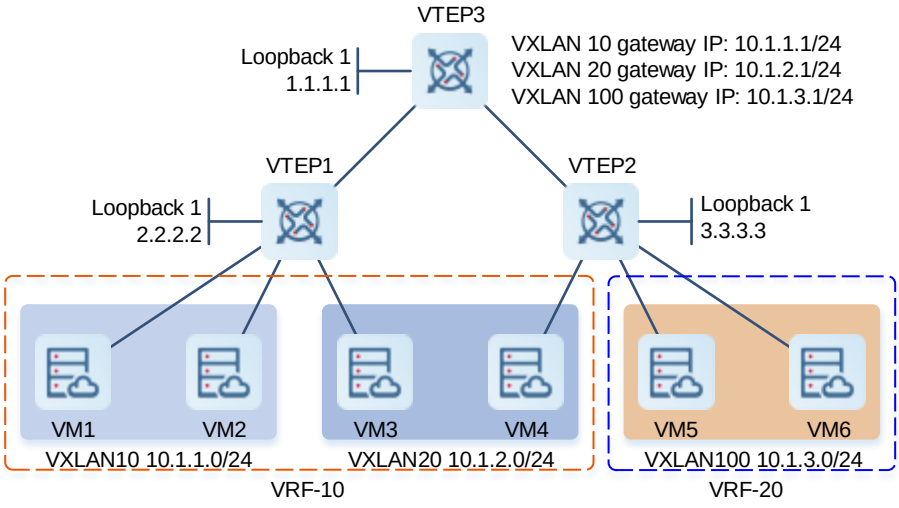
5.2 Configuring EVPN-based Multi-Tenant Centralized VXLAN Networking

5.2.1 Requirements

In the topology as shown in Figure 5-2, centralized VXLAN networking using dynamic tunnels needs to be configured on VTEPs to support multi-tenant application in the data center, and VRF instances need to be configured to classify VXLANs by tenant. In this way, the VXLANs of the same tenant can communicate with each other through Layer 3 routes, but the VXLANs of different tenants cannot communicate with each other.

5.2.2 Topology

Figure 5-2 Configuring EVPN-based Multi-Tenant Centralized VXLAN Networking



5.2.3 Notes

- On each VTEP, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that the unicast routes are reachable (omitted).
- On each VTEP, configure the BGP-EVPN routing protocol to ensure that the devices establish BGP neighbor relationships and exchange L2VPN EVPN information.
- On each VTEP, configure an EVI for BGP-EVPN.
- On the virtualized servers, configure VXLANs and specify the gateway address for the VMs.
- On each VTEP, configure and associate a loopback interface for the establishment of a VXLAN tunnel.
- On each VTEP, create a VXLAN instance and associate it with a VLAN.
- On VTEP 3, create an overlay router interface and configure the VXLAN gateway IP address. Configure different VRF instances for the overlay router interface and identify tenants by VRF instance.
- On VTEP 3, associate the VXLAN instance with the overlay router interface to implement Layer 3 routing between different VNI networks.
- On VTEP 3, configure remote ARP packet learning so that VXLAN routing entries are dynamically generated.
- (Optional) Configure ARP suppression on VTEP 1 and VTEP 2 to reduce ARP packets flowing to VXLANs.

5.2.4 Procedure

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 2.2.2.2/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# vtep
VTEP1(cfg-vtep)# source loopback 1
VTEP1(cfg-vtep)# arp suppress enable
VTEP1(cfg-vtep)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# vxlan 20
VTEP1(cfg-vxlan)# extend-vlan 20
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# router-protocol bgp 64512
VTEP1(cfg-router)# peer 1.1.1.1 remote-as 64512
VTEP1(cfg-router)# peer 1.1.1.1 update-source loopback 1
VTEP1(cfg-router)# address-family-mode l2vpn evpn
VTEP1(cfg-router-af)# peer 1.1.1.1 activate
```

```
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# exit
VTEP1(cfg)# evpn
VTEP1(cfg-evpn)# vni 10
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 20
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
```

(2) Configure VTEP 2.

```
VTEP> privilege
VTEP2# config-system terminal
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 3.3.3.3/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 1
VTEP2(cfg-vtep)# arp suppress enable
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# vxlan 100
VTEP2(cfg-vxlan)# extend-vlan 100
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# vxlan 20
VTEP2(cfg-vxlan)# extend-vlan 20
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 64512
VTEP2(cfg-router)# peer 1.1.1.1 remote-as 64512
VTEP2(cfg-router)# peer 1.1.1.1 update-source loopback 1
VTEP2(cfg-router)# address-family-mode l2vpn evpn
VTEP2(cfg-router)# peer 1.1.1.1 activate
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# exit
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 20
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 100
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
```

(3) Configure VTEP 3.

```
VTEP3> privilege
VTEP3# config-system terminal
VTEP3(cfg)# interface loopback 1
VTEP3(cfg-if-Loopback 1)# ip address 1.1.1.1/32
VTEP3(cfg-if-Loopback 1)# exit
VTEP3(cfg)# vtep
VTEP3(cfg-vtep)# source loopback 1
VTEP3(cfg-vtep)# remote arp learn enable
VTEP3(cfg-vtep)# exit
VTEP3(cfg)# ip vrf vrf-10
VTEP3(cfg-vrf)# rd 10:10
VTEP3(cfg-vrf)# route-target both 1000:1000
VTEP3(cfg-vrf)# exit
VTEP3(cfg)# interface overlayrouter 10
VTEP3(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP3(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP3(cfg-if-OverlayRouter 10)# exit
VTEP3(cfg)# vxlan 10
VTEP3(cfg-vxlan)# router-interface overlayrouter 10
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# interface overlayrouter 20
VTEP3(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
VTEP3(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP3(cfg-if-OverlayRouter 20)# exit
VTEP3(cfg)# vxlan 20
VTEP3(cfg-vxlan)# router-interface overlayrouter 20
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# ip vrf vrf-20
VTEP3(cfg-vrf)# rd 20:20
VTEP3(cfg-vrf)# route-target both 2000:2000
VTEP3(cfg-vrf)# exit
VTEP3(cfg)# interface overlayrouter 100
VTEP3(cfg-if-OverlayRouter 100)# ip vrf binding vrf-20
VTEP3(cfg-if-OverlayRouter 100)# ip address 10.1.3.1/24
VTEP3(cfg-if-OverlayRouter 100)# exit
VTEP3(cfg)# vxlan 100
VTEP3(cfg-vxlan)# router-interface overlayrouter 100
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# router-protocol bgp 64512
VTEP3(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP3(cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP3(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP3(cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP3(cfg-router)# address-family-mode l2vpn evpn
VTEP3(cfg-router-af)# peer 2.2.2.2 activate
VTEP3(cfg-router-af)# peer 3.3.3.3 activate
```

```
VTEP3(cfg-router-af)# peer 2.2.2.2 route-reflector-client
VTEP3(cfg-router-af)# peer 3.3.3.3 route-reflector-client
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP3(cfg-router-af)# network-advertise 10.1.1.0 mask 255.255.255.0
VTEP3(cfg-router-af)# network-advertise 10.1.2.0 mask 255.255.255.0
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# address-family-mode ipv4 vrf vrf-20
VTEP3(cfg-router-af)# network-advertise 10.1.3.0 mask 255.255.255.0
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# exit
VTEP3(cfg)# evpn
VTEP3(cfg-evpn)# vni 10
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# exit
VTEP3(cfg-evpn)# vni 20
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# exit
VTEP3(cfg-evpn)# vni 100
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# exit
```

5.2.5 Verification

- After the commands above are run and the centralized VXLAN networking using dynamic tunnels is configured:
  - VMs in VRF-10 can ping each other successfully.
  - VMs in VRF-20 can ping each other successfully.
  - VMs in VRF-10 cannot ping those in VRF-20 and vice versa.
  - VMs can be migrated in the same VXLAN, they can access the Internet service successfully after migration, and the configurations do not need to be modified.
- Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 1 as an example.

```
VTEP1# display vxlan
VXLAN Total Count: 2
VXLAN Capacity     : 8000

VXLAN 10
     Symmetric property    : FALSE
     Routed interface      : -
     Extend VLAN           : 10
     VTEP Adjacency Count: 1
```

```
    VTEP Adjacency List :
    Interface                Source IP                                    Destination IP
Type
    ---------------------- ---------------------------------------------- ---------------------------------------------- -------
    OverlayTunnel 6145       2.2.2.2                                      1.1.1.1
dynamic

VXLAN 20
    Symmetric property   : FALSE
    Routed interface     : -
    Extend VLAN          : 20
    VTEP Adjacency Count: 2
    VTEP Adjacency List :
    Interface                Source IP                                    Destination IP
Type
    ---------------------- ---------------------------------------------- ---------------------------------------------- -------
    OverlayTunnel 6145       2.2.2.2                                      1.1.1.1
dynamic
    OverlayTunnel 6146       2.2.2.2                                      3.3.3.3
dynamic
```

5.2.6 Configuration Files

- VTEP 1 configuration file

```
!
vtep
 source loopback 1
 arp suppress enable
!
interface Loopback 1
 ip address 2.2.2.2 255.255.255.255
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.1 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 1.1.1.1 activate
  peer 1.1.1.1 send-community extended
  exit-address-family-mode
!
```

```
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
!
vxlan 10
 extend-vlan 10
 arp suppress enable
!
vxlan 20
 extend-vlan 20
 arp suppress enable
!
```

- VTEP 2 configuration file

```
!
vtep
 source loopback 1
 arp suppress enable
!
interface Loopback 1
 ip address 3.3.3.3 255.255.255.255
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.1 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 1.1.1.1 activate
  peer 1.1.1.1 send-community extended
  exit-address-family-mode
!
evpn
 vni 20
  rd auto
  route-target both auto
 vni 100
  rd auto
```

```
   route-target both auto
!
vxlan 20
 extend-vlan 20
 arp suppress enable
!
vxlan 100
 extend-vlan 100
 arp suppress enable
!
```

- VTEP 3 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target both 1000:1000
!
ip vrf vrf-20
 rd 20:20
 route-target both 2000:2000
!
vtep
 source loopback 1
 remote arp learn enable
!
interface Loopback 1
 ip address 1.1.1.1 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 10.1.2.1 255.255.255.0
!
interface OverlayRouter 100
 ip vrf binding vrf-20
 ip address 10.1.3.1 255.255.255.0
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
```

```
 peer 3.3.3.3 remote-as 64512
 peer 3.3.3.3 update-source Loopback 1
 address-family-mode ipv4
  peer 2.2.2.2 activate
  peer 3.3.3.3 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 2.2.2.2 activate
  peer 2.2.2.2 route-reflector-client
  peer 2.2.2.2 send-community extended
  peer 3.3.3.3 activate
  peer 3.3.3.3 route-reflector-client
  peer 3.3.3.3 send-community extended
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-10
  network-advertise 10.1.1.0 mask 255.255.255.0
  network-advertise 10.1.2.0 mask 255.255.255.0
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-20
  network-advertise 10.1.3.0 mask 255.255.255.0
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
!
vxlan 20
 router-interface OverlayRouter 20
!
vxlan 100
 router-interface OverlayRouter 100
!
```

5.3 Configuring EVPN-based Multi-Tenant Centralized Anycast Gateway Deployment

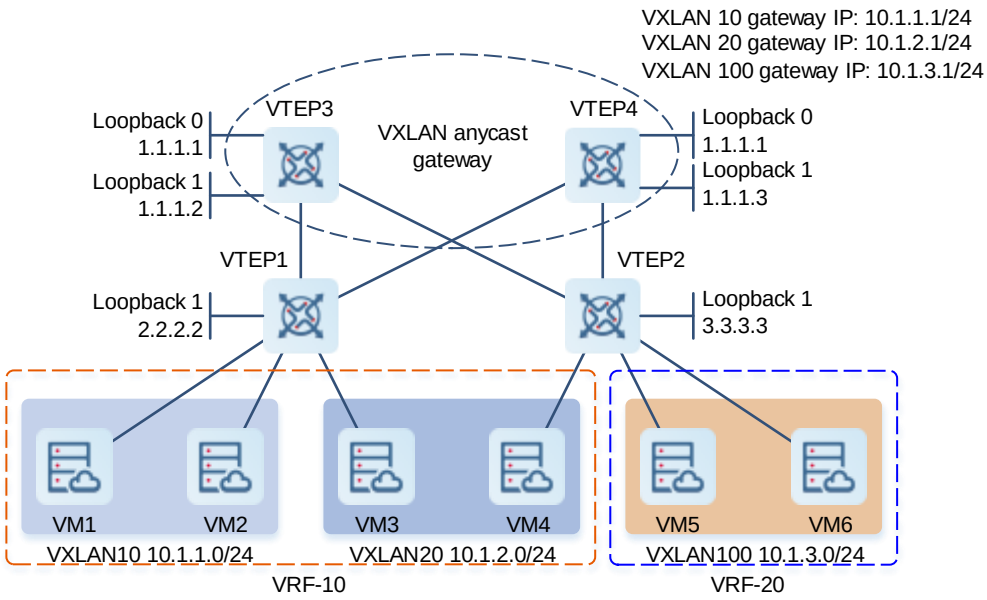
5.3.1 Requirements

In the topology as shown in Figure 5-3, the following configurations are needed to support multi-tenant application in the data center and ensure that fast traffic convergence can be implemented on the deployed VXLANs when the gateway is faulty:

- Configure centralized VXLAN networking using dynamic tunnels on VTEPs, configure VRF instances to configure multiple VXLANs for each tenant so that VXLANs of the same tenant can access each other through Layer 3 routing while VXLANs of different tenants cannot.
- Configure the VXLAN gateways as the anycast gateway so that two VXLAN gateways can be virtualized into one VXLAN gateway. When one VXLAN gateway malfunctions, VXLAN traffic in the topology can be still forwarded normally.

5.3.2 Topology

Figure 5-3 Configuring EVPN-based Multi-Tenant Centralized Anycast All-Alive Gateway Deployment



5.3.3 Notes

- On each VTEP, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that the unicast routes are reachable (omitted).
- On each VTEP, configure the BGP-EVPN routing protocol to ensure that the four VTEPs establish BGP neighbor relationships and exchange L2VPN EVPN information.
- On each VTEP, configure an EVI for BGP-EVPN.
- On the virtualized servers, configure VXLANs and specify the gateway address for the VMs.
- On each VTEP, configure and associate a loopback interface for the establishment of a VXLAN tunnel.

- On VTEP 3, create an overlay router interface and configure the VXLAN gateway IP address. Configure different VRF instances for the overlay router interface and identify tenants by VRF instance.
- On each VTEP, create a VXLAN instance and associate it with a VLAN.
- On VTEP 3 and VTEP 4, create an overlay router interface and configure the VXLAN gateway IP address. Configure different VRF instances for the overlay router interfaces and identify tenants by VRF instance. Associate the VXLAN instance with the overlay router interface to implement VXLAN routing.
- On VTEP 3 and VTEP 4, configure the same anycast gateway MAC address and configure the anycast attribute for all overlay router interfaces.
- On VTEP 3 and VTEP 4, configure remote ARP packet learning so that VXLAN routing entries are dynamically generated.
- (Optional) Configure ARP suppression on VTEP 1 and VTEP 2 to reduce ARP packets flowing to VXLANs.

5.3.4 Procedure

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 2.2.2.2/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# vtep
VTEP1(cfg-vtep)# source loopback 1
VTEP1(cfg-vtep)# arp suppress enable
VTEP1(cfg-vtep)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# vxlan 20
VTEP1(cfg-vxlan)# extend-vlan 20
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# router-protocol bgp 64512
VTEP1 (cfg-router)# peer 1.1.1.2 remote-as 64512
VTEP1 (cfg-router)# peer 1.1.1.2 update-source loopback 1
VTEP1 (cfg-router)# peer 1.1.1.3 remote-as 64512
VTEP1 (cfg-router)# peer 1.1.1.3 update-source loopback 1
VTEP1 (cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP1 (cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP1(cfg-router)# address-family-mode l2vpn evpn
VTEP1(cfg-router-af)# peer 1.1.1.2 activate
VTEP1(cfg-router-af)# peer 1.1.1.3 activate
VTEP1(cfg-router-af)# peer 3.3.3.3 activate
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# exit
```

```
VTEP1(cfg)# evpn
VTEP1(cfg-evpn)# vni 10
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 20
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
```

(2) Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 3.3.3.3/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 1
VTEP2(cfg-vtep)# arp suppress enable
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# vxlan 100
VTEP2(cfg-vxlan)# extend-vlan 100
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# vxlan 20
VTEP2(cfg-vxlan)# extend-vlan 20
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 64512
VTEP2 (cfg-router)# peer 1.1.1.2 remote-as 64512
VTEP2 (cfg-router)# peer 1.1.1.2 update-source loopback 1
VTEP2 (cfg-router)# peer 1.1.1.3 remote-as 64512
VTEP2 (cfg-router)# peer 1.1.1.3 update-source loopback 1
VTEP2 (cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP2 (cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP2(cfg-router)# address-family-mode l2vpn evpn
VTEP2(cfg-router-af)# peer 1.1.1.2 activate
VTEP2(cfg-router-af)# peer 1.1.1.3 activate
VTEP2(cfg-router-af)# peer 2.2.2.2 activate
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# exit
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 20
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 100
```

```
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
```

(3) Configure VXLANs on VTEP 3 and VTEP 4. The configurations on the two devices are the same. The following uses VTEP 3 as an example.

```
VTEP3> privilege
VTEP3# config-system terminal
VTEP3(cfg)# interface loopback 0
VTEP3(cfg-if- Loopback 0)# ip address 1.1.1.1/32
VTEP3(cfg-if- Loopback 0)# exit
VTEP3(cfg)# interface loopback 1
VTEP3(cfg-if-Loopback 1)# ip address 1.1.1.2/32
VTEP3(cfg-if-Loopback 1)# exit
VTEP3(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP3(cfg)# vtep
VTEP3(cfg-vtep)# source loopback 0
VTEP3(cfg-vtep)# remote arp learn enable
VTEP3(cfg-vtep)# exit
VTEP3(cfg)# ip vrf vrf-10
VTEP3(cfg-vrf)# rd 10:10
VTEP3(cfg-vrf)# route-target both 1000:1000
VTEP3(cfg-vrf)# exit
VTEP3(cfg)# interface overlayrouter 10
VTEP3(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP3(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP3(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 10)# exit
VTEP3(cfg)# vxlan 10
VTEP3(cfg-vxlan)# router-interface overlayrouter 10
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# interface overlayrouter 20
VTEP3(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
VTEP3(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP3(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 20)# exit
VTEP3(cfg)# vxlan 20
VTEP3(cfg-vxlan)# router-interface overlayrouter 20
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# ip vrf vrf-20
VTEP3(cfg-vrf)# rd 20:20
VTEP3(cfg-vrf)# route-target both 2000:2000
VTEP3(cfg-vrf)# exit
VTEP3(cfg)# interface overlayrouter 100
VTEP3(cfg-if-OverlayRouter 100)# ip vrf binding vrf-20
VTEP3(cfg-if-OverlayRouter 100)# ip address 10.1.3.1/24
VTEP3(cfg-if-OverlayRouter 100)# anycast-gateway
```

```
VTEP3(cfg-if-OverlayRouter 100)# exit
VTEP3(cfg)# vxlan 100
VTEP3(cfg-vxlan)# router-interface overlayrouter 100
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# router-protocol bgp 64512
VTEP3(cfg-router)# peer 1.1.1.3 remote-as 64512
VTEP3(cfg-router)# peer 1.1.1.3 update-source loopback 1
VTEP3(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP3(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP3(cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP3(cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP3(cfg-router)# address-family-mode l2vpn evpn
VTEP3(cfg-router)# peer 1.1.1.3 activate
VTEP3(cfg-router)# peer 2.2.2.2 activate
VTEP3(cfg-router)# peer 3.3.3.3 activate
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP3(cfg-router-af)# network-advertise 10.1.1.0 mask 255.255.255.0
VTEP3(cfg-router-af)# network-advertise 10.1.2.0 mask 255.255.255.0
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# address-family-mode ipv4 vrf vrf-20
VTEP3(cfg-router-af)# network-advertise 10.1.3.0 mask 255.255.255.0
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# exit
VTEP3(cfg)# evpn
VTEP3(cfg-evpn)# vni 10
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# exit
VTEP3(cfg-evpn)# vni 20
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# exit
VTEP3(cfg-evpn)# vni 100
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# exit
```

(4) Configure VTEP 4.

```
VTEP4> privilege
VTEP4# config-system terminal
VTEP4(cfg)# interface loopback 0
VTEP4(cfg-if- Loopback 0)# ip address 1.1.1.1/32
VTEP4(cfg-if- Loopback 0)# exit
VTEP4(cfg)# interface loopback 1
VTEP4(cfg-if-Loopback 1)# ip address 1.1.1.3/32
VTEP4(cfg-if-Loopback 1)# exit
```

```
VTEP4(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP4(cfg)# vtep
VTEP4(cfg-vtep)# source loopback 0
VTEP4(cfg-vtep)# remote arp learn enable
VTEP4(cfg-vtep)# exit
VTEP4(cfg)# ip vrf vrf-10
VTEP4(cfg-vrf)# rd 10:10
VTEP4(cfg-vrf)# route-target both 1000:1000
VTEP4(cfg-vrf)# exit
VTEP4(cfg)# interface overlayrouter 10
VTEP4(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP4(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP4(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP4(cfg-if-OverlayRouter 10)# exit
VTEP4(cfg)# vxlan 10
VTEP4(cfg-vxlan)# router-interface overlayrouter 10
VTEP4(cfg-vxlan)# exit
VTEP4(cfg)# interface overlayrouter 20
VTEP4(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
VTEP4(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP4(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP4(cfg-if-OverlayRouter 20)# exit
VTEP4(cfg)# vxlan 20
VTEP4(cfg-vxlan)# router-interface overlayrouter 20
VTEP4(cfg-vxlan)# exit
VTEP4(cfg)# ip vrf vrf-20
VTEP4(cfg-vrf)# rd 20:20
VTEP4(cfg-vrf)# route-target both 2000:2000
VTEP4(cfg-vrf)# exit
VTEP4(cfg)# interface overlayrouter 100
VTEP4(cfg-if-OverlayRouter 100)# ip vrf binding vrf-20
VTEP4(cfg-if-OverlayRouter 100)# ip address 10.1.3.1/24
VTEP4(cfg-if-OverlayRouter 100)# anycast-gateway
VTEP4(cfg-if-OverlayRouter 100)# exit
VTEP4(cfg)# vxlan 100
VTEP4(cfg-vxlan)# router-interface overlayrouter 100
VTEP4(cfg-vxlan)# exit
VTEP4(cfg)# router-protocol bgp 64512
VTEP4(cfg-router)# peer 1.1.1.2 remote-as 64512
VTEP4(cfg-router)# peer 1.1.1.2 update-source loopback 1
VTEP4(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP4(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP4(cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP4(cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP4(cfg-router)# address-family-mode l2vpn evpn
VTEP4(cfg-router)# peer 1.1.1.2 activate
```

```
VTEP4(cfg-router)# peer 2.2.2.2 activate
VTEP4(cfg-router)# peer 3.3.3.3 activate
VTEP4(cfg-router-af)# exit
VTEP4(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP4(cfg-router-af)# network-advertise 10.1.1.0 mask 255.255.255.0
VTEP4(cfg-router-af)# network-advertise 10.1.2.0 mask 255.255.255.0
VTEP4(cfg-router-af)# exit
VTEP4(cfg-router)# address-family-mode ipv4 vrf vrf-20
VTEP4(cfg-router-af)# network-advertise 10.1.3.0 mask 255.255.255.0
VTEP4(cfg-router-af)# exit
VTEP4(cfg-router)# exit
VTEP4(cfg)# evpn
VTEP4(cfg-evpn)# vni 10
VTEP4(cfg-evpn-vni)# rd auto
VTEP4(cfg-evpn-vni)# route-target both auto
VTEP4(cfg-evpn-vni)# exit
VTEP4(cfg-evpn)# vni 20
VTEP4(cfg-evpn-vni)# rd auto
VTEP4(cfg-evpn-vni)# route-target both auto
VTEP4(cfg-evpn-vni)# exit
VTEP4(cfg-evpn)# vni 100
VTEP4(cfg-evpn-vni)# rd auto
VTEP4(cfg-evpn-vni)# route-target both auto
VTEP4(cfg-evpn-vni)# exit
```

5.3.5 Verification

- After the commands above are run and the centralized VXLAN networking using dynamic tunnels is configured:
  - VMs in VRF-10 can ping each other successfully.
  - VMs in VRF-20 can ping each other successfully.
  - VMs in VRF-10 cannot ping those in VRF-20 and vice versa.
  - VMs can be migrated in the same VXLAN, they can access the Internet service successfully after migration, and the configurations do not need to be modified.
- Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 1 and VTEP 3 as an example.

  On VTEP 1, run the **display vxlan** command to display VXLAN instances on the device.

```
VTEP1# display vxlan
VXLAN Total Count: 2
VXLAN Capacity     : 8000

VXLAN 10
     Symmetric property    : FALSE
     Routed interface      : -
```

```
    Extend VLAN              : 10
    VTEP Adjacency Count: 1
    VTEP Adjacency List :
    Interface                Source IP                                   Destination IP
Type
    ---------------------- ------------------------------------------- ------------------------------------------- -------
    OverlayTunnel 6145       2.2.2.2                                     1.1.1.1
dynamic

VXLAN 20
    Symmetric property   : FALSE
    Routed interface     : -
    Extend VLAN              : 20
    VTEP Adjacency Count: 2
    VTEP Adjacency List :
    Interface                Source IP                                   Destination IP
Type
    ---------------------- ------------------------------------------- ------------------------------------------- -------
    OverlayTunnel 6145       2.2.2.2                                     1.1.1.1
dynamic
    OverlayTunnel 6146       2.2.2.2                                     3.3.3.3
dynamic
```

On VTEP 3, run the **display vxlan** command to display VXLAN instances on the device.

```
VTEP3# display vxlan
VXLAN Total Count: 3
VXLAN Capacity     : 8000

VXLAN 10
    Symmetric property   : FALSE
    Routed interface     : OverlayRouter 10 (anycast)
    Extend VLAN              : 10
    VTEP Adjacency Count: 1
    VTEP Adjacency List :
    Interface                Source IP                                   Destination IP
Type
    ---------------------- ------------------------------------------- ------------------------------------------- -------
    OverlayTunnel 6147       1.1.1.1                                     2.2.2.2
dynamic

VXLAN 20
    Symmetric property   : FALSE
    Routed interface     : OverlayRouter 20 (anycast)
    Extend VLAN              : 20
    VTEP Adjacency Count: 2
    VTEP Adjacency List :
```

```
      Interface                 Source IP                                 Destination IP
Type
      ----------------------- ---------------------------------------- ---------------------------------------- -------
      OverlayTunnel 6147        1.1.1.1                                   2.2.2.2
dynamic
      OverlayTunnel 6148        1.1.1.1                                   3.3.3.3
dynamic

VXLAN 100
      Symmetric property    : FALSE
      Routed interface      : OverlayRouter 100 (anycast)
      Extend VLAN             : 100
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                 Source IP                                 Destination IP
Type
      ----------------------- ---------------------------------------- ---------------------------------------- -------
      OverlayTunnel 6148        1.1.1.1                                   3.3.3.3
dynamic
```

5.3.6 Configuration Files

- VTEP 1 configuration file

```
!
vtep
 source loopback 1
 arp suppress enable
!
interface Loopback 1
 ip address 2.2.2.2 255.255.255.255
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.1 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 1.1.1.1 activate
  peer 1.1.1.1 send-community extended
  exit-address-family-mode
!
evpn
```

```
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
!
vxlan 10
 extend-vlan 10
 arp suppress enable
!
vxlan 20
 extend-vlan 20
 arp suppress enable
!
```

- VTEP 2 configuration file

```
!
vtep
 source loopback 1
 arp suppress enable
!
interface Loopback 1
 ip address 3.3.3.3 255.255.255.255
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.1 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 1.1.1.1 activate
  peer 1.1.1.1 send-community extended
  exit-address-family-mode
!
evpn
 vni 20
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
```

```
!
vxlan 20
 extend-vlan 20
 arp suppress enable
!
vxlan 100
 extend-vlan 100
 arp suppress enable
!
```

- VTEP 3 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target both 1000:1000
!
ip vrf vrf-20
 rd 20:20
 route-target both 2000:2000
!
vtep
 source loopback 0
 remote arp learn enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 0
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 1
 ip address 1.1.1.2 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-20
 ip address 10.1.3.1 255.255.255.0
 anycast-gateway
!
```

```
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.3 remote-as 64512
 peer 1.1.1.3 update-source Loopback 1
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
 peer 3.3.3.3 remote-as 64512
 peer 3.3.3.3 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.3 activate
  peer 2.2.2.2 activate
  peer 3.3.3.3 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 1.1.1.3 activate
  peer 1.1.1.3 send-community extended
  peer 2.2.2.2 activate
  peer 2.2.2.2 send-community extended
  peer 3.3.3.3 activate
  peer 3.3.3.3 send-community extended
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-10
  network 10.1.1.0 mask 255.255.255.0
  network 10.1.2.0 mask 255.255.255.0
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-20
  network 10.1.3.0 mask 255.255.255.0
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
!
vxlan 10
```

```
 router-interface OverlayRouter 10
!
vxlan 20
 router-interface OverlayRouter 20
!
vxlan 100
 router-interface OverlayRouter 100
!
```

- VTEP 4 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target both 1000:1000
!
ip vrf vrf-20
 rd 20:20
 route-target both 2000:2000
!
vtep
 source loopback 0
 remote arp learn enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 0
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 1
 ip address 1.1.1.3 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-20
 ip address 10.1.3.1 255.255.255.0
 anycast-gateway
!
router-protocol bgp 64512
```

```
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.2 remote-as 64512
 peer 1.1.1.2 update-source Loopback 1
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
 peer 3.3.3.3 remote-as 64512
 peer 3.3.3.3 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.2 activate
  peer 2.2.2.2 activate
  peer 3.3.3.3 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 1.1.1.2 activate
  peer 1.1.1.2 send-community extended
  peer 2.2.2.2 activate
  peer 2.2.2.2 send-community extended
  peer 3.3.3.3 activate
  peer 3.3.3.3 send-community extended
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-10
  network-advertise 10.1.1.0 mask 255.255.255.0
  network-advertise 10.1.2.0 mask 255.255.255.0
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-20
  network-advertise 10.1.3.0 mask 255.255.255.0
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
```

```
!
vxlan 20
 router-interface OverlayRouter 20
!
vxlan 100
 router-interface OverlayRouter 100
!
```

5.4 Configuring EVPN-based Multi-Tenant Distributed VXLAN Networking (Asymmetric Deployment)

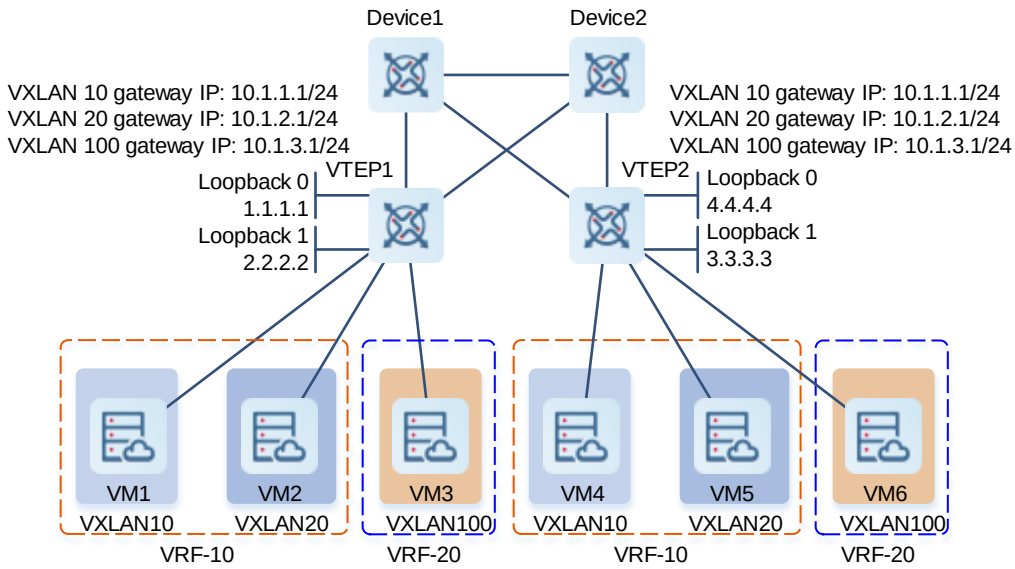
5.4.1 Requirements

In the topology as shown in Figure 5-4, the following configurations are needed to support multi-tenant application in the data center and improve the network scalability:

- Configure VTEPs on access devices. Configure distributed VXLAN networking using dynamic tunnels on the VTEPs, configure VRF instances to assign multiple VXLANs to each tenant so that VXLANs of the same tenant can access each other through Layer 3 routing while VXLANs of different tenants cannot.
- Configure VXLAN gateways as the anycast gateway.
- VXLAN configuration is not required on core devices, and the VXLAN access and Layer 3 routing are accomplished on access devices.

5.4.2 Topology

Figure 5-4 Configuring EVPN-based Multi-Tenant Distributed VXLAN Networking (Asymmetric Deployment)



5.4.3 Notes

- On device 1, device 2, and VTEPs, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that

the unicast routes are reachable (omitted).

- On device 1, device 2, and VTEPs, configure the BGP-EVPN routing protocol to ensure that the four devices establish BGP neighbor relationships and exchange L2VPN EVPN information.
- On each VTEP, configure an EVI for BGP-EVPN.
- On the virtualized servers, configure VXLANs and specify the gateway address for the VMs.
- On each VTEP, configure and associate a loopback interface for the establishment of a tunnel.
- On each VTEP, configure the anycast gateway MAC address.
- On each VTEP, create a VXLAN instance and associate it with a VLAN.
- On each VTEP, create an overlay router interface and configure the VXLAN gateway IP address. Configure different VRF instances for the overlay router interfaces and identify tenants by VRF instance. Configure the anycast attribute for the overlay router interface so that all VXLAN gateways in the network use the same IP address/MAC address.
- On each VTEP, associate the VXLAN instance with the overlay router interface to implement VXLAN routing.
- (Optional) Configure ARP suppression on each VTEP to reduce ARP packets flowing into the VXLANs.
- (Optional) Configure ARP proxy on the overlay router interface of each VTEP. The VXLAN gateway responds with the gateway MAC address as a proxy and VXLANs perform only Layer 3 forwarding.

5.4.4 Procedure

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 0
VTEP1(cfg-if-Loopback 0)# ip address 1.1.1.1/32
VTEP1(cfg-if-Loopback 0)# exit
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 2.2.2.2/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP1(cfg)# vtep
VTEP1(cfg-vtep)# source loopback 0
VTEP1(cfg-vtep)# arp suppress enable
VTEP1(cfg-vtep)# exit
VTEP1(cfg)# ip vrf vrf-10
VTEP1(cfg-vrf)# rd 10:10
VTEP1(cfg-vrf)# route-target both 1000:1000
VTEP1(cfg-vrf)# exit
VTEP1(cfg)# interface overlayrouter 10
VTEP1(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP1(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 10)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP1(cfg-if-OverlayRouter 10)# exit
```

```
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# router-interface overlayrouter 10
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 20
VTEP1(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP1(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 20)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP1(cfg-if-OverlayRouter 20)# exit
VTEP1(cfg)# vxlan 20
VTEP1(cfg-vxlan)# extend-vlan 20
VTEP1(cfg-vxlan)# router-interface overlayrouter 20
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# ip vrf vrf-20
VTEP1(cfg-vrf)# rd 20:20
VTEP1(cfg-vrf)# route-target both 2000:2000
VTEP1(cfg-vrf)# exit
VTEP1(cfg)# interface overlayrouter 100
VTEP1(cfg-if-OverlayRouter 100)# ip vrf binding vrf-20
VTEP1(cfg-if-OverlayRouter 100)# ip address 10.1.3.1/24
VTEP1(cfg-if-OverlayRouter 100)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 100)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP1(cfg-if-OverlayRouter 100)# exit
VTEP1(cfg)# vxlan 100
VTEP1(cfg-vxlan)# extend-vlan 100
VTEP1(cfg-vxlan)# router-interface overlayrouter 100
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# router-protocol bgp 64512
VTEP1(cfg-router)# bgp router-id 2.2.2.2
VTEP1(cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP1(cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP1(cfg-router)# address-family-mode l2vpn evpn
VTEP1(cfg-router-af)# peer 3.3.3.3 activate
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# exit
VTEP1(cfg)# evpn
VTEP1(cfg-evpn)# vni 10
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
```

```
VTEP1(cfg-evpn)# vni 20
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 100
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
```

Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP1(cfg)# interface loopback 0
VTEP1(cfg-if-Loopback 0)# ip address 4.4.4.4/32
VTEP1(cfg-if-Loopback 0)# exit
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 3.3.3.3/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 0
VTEP2(cfg-vtep)# arp suppress enable
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# ip vrf vrf-10
VTEP2(cfg-vrf)# rd 10:10
VTEP2(cfg-vrf)# route-target both 1000:1000
VTEP2(cfg-vrf)# exit
VTEP2(cfg)# interface overlayrouter 10
VTEP2(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP2(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 10)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP2(cfg-if-OverlayRouter 10)# exit
VTEP2(cfg)# vxlan 10
VTEP2(cfg-vxlan)# extend-vlan 10
VTEP2(cfg-vxlan)# router-interface overlayrouter 10
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 20
VTEP2(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP2(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 20)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP2(cfg-if-OverlayRouter 20)# exit
VTEP2(cfg)# vxlan 20
```

```
VTEP2(cfg-vxlan)# extend-vlan 20
VTEP2(cfg-vxlan)# router-interface overlayrouter 20
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# ip vrf vrf-20
VTEP2(cfg-vrf)# rd 20:20
VTEP2(cfg-vrf)# route-target both 2000:2000
VTEP2(cfg-vrf)# exit
VTEP2(cfg)# interface overlayrouter 100
VTEP2(cfg-if-OverlayRouter 100)# ip vrf binding vrf-20
VTEP2(cfg-if-OverlayRouter 100)# ip address 10.1.3.1/24
VTEP2(cfg-if-OverlayRouter 100)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 100)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP2(cfg-if-OverlayRouter 100)# exit
VTEP2(cfg)# vxlan 100
VTEP2(cfg-vxlan)# extend-vlan 100
VTEP2(cfg-vxlan)# router-interface overlayrouter 100
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 64512
VTEP1(cfg-router)# bgp-instance router-id 3.3.3.3
VTEP2(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP2(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP2(cfg-router)# address-family-mode l2vpn evpn
VTEP2(cfg-router-af)# peer 2.2.2.2 activate
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# exit
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 10
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 20
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 100
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
```

5.4.5 Verification

- After the commands above are run and the distributed asymmetric VXLAN networking is configured:
  - VMs in VRF-10 can ping each other successfully.
  - VMs in VRF-20 can ping each other successfully.

- VMs in VRF-10 cannot ping those in VRF-20 and vice versa.
- VMs can be migrated in the same VXLAN, they can access the Internet service successfully after migration, and the configurations do not need to be modified.

- Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 1 as an example.

```
VTEP1# display vxlan
VXLAN Total Count: 3
VXLAN Capacity     : 8000

VXLAN 10
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 10 (anycast)
      Extend VLAN           : 10
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                 Source IP                                   Destination IP
Type
      ----------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145        1.1.1.1                                     4.4.4.4
dynamic

VXLAN 20
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 20 (anycast)
      Extend VLAN           : 20
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                 Source IP                                   Destination IP
Type
      ----------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145        1.1.1.1                                     4.4.4.4
dynamic

VXLAN 100
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 100 (anycast)
      Extend VLAN           : 100
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                 Source IP                                   Destination IP
Type
      ----------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145        1.1.1.1                                     4.4.4.4
dynamic
```

5.4.6 Configuration Files

- VTEP 1 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target both 1000:1000
!
ip vrf vrf-20
 rd 20:20
 route-target both 2000:2000
!
vtep
 source loopback 0
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 0
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 1
 ip address 2.2.2.2 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
interface OverlayRouter 100
 ip vrf binding vrf-20
 ip address 10.1.3.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
router-protocol bgp 64512
 bgp-instance router-id 2.2.2.2
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
```

```
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 3.3.3.3 remote-as 64512
 peer 3.3.3.3 update-source Loopback 1
 address-family-mode ipv4
  peer 3.3.3.3 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 3.3.3.3 activate
  peer 3.3.3.3 send-community extended
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 extend-vlan 10
 arp suppress enable
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
 arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 extend-vlan 100
 arp suppress enable
!
```

- VTEP 2 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target both 1000:1000
!
ip vrf vrf-20
 rd 20:20
```

```
 route-target both 2000:2000
!
vtep
 source loopback 0
arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 0
 ip address 4.4.4.4 255.255.255.255
!
interface Loopback 1
 ip address 3.3.3.3 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
route-in-vni
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
route-in-vni
!
interface OverlayRouter 100
 ip vrf binding vrf-20
 ip address 10.1.3.1 255.255.255.0
 anycast-gateway
route-in-vni
!
router-protocol bgp 64512
 bgp-instance router-id 3.3.3.3
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
 address-family-mode ipv4
  peer 2.2.2.2 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 2.2.2.2 activate
  peer 2.2.2.2 send-community extended
```

```
   exit-address-family-mode
!
evpn
 vni 10
   rd auto
   route-target both auto
 vni 20
   rd auto
   route-target both auto
 vni 100
   rd auto
   route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 extend-vlan 10
arp suppress enable
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 extend-vlan 100
arp suppress enable
!
```

5.4.7 Common Errors

- In an asymmetric deployment scenario, a VRF instance is configured on a VTEP but no gateway is configured for all VXLANs in the VRF instance. As a result, the VXLANs with the gateway not configured cannot perform Layer 3 forwarding with other VXLANs.
- The anycast gateway MAC address conflicts with the MAC address of a device in VXLANs, resulting in a forwarding abnormality.

5.5 Configuring EVPN-based Multi-Tenant Distributed VXLAN Networking (Symmetric Deployment)

5.5.1 Requirements

In the topology as shown in Figure 5-5, the following configurations are needed to support multi-tenant application in the data center, improve the network scalability, and meet networking requirements of large data centers:

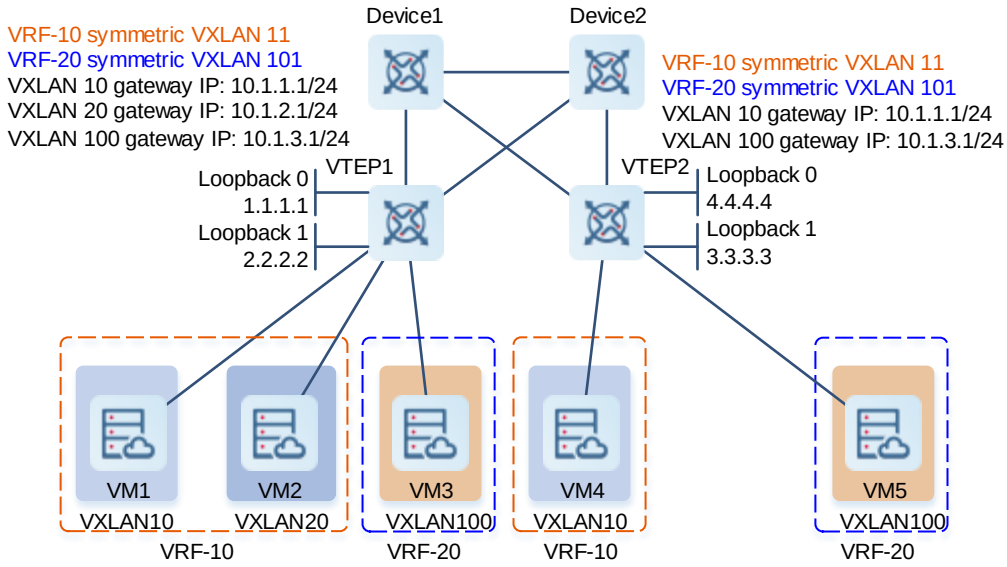
- Configure VTEPs on access devices. Configure distributed VXLAN networking using dynamic tunnels on the

VTEPs, configure VRF instances to assign multiple VXLANs to each tenant so that VXLANs of the same tenant can access each other through Layer 3 routing while VXLANs of different tenants cannot.

- Configure VXLAN gateways as the anycast gateway.
- VXLAN configuration is not required on core devices, and the VXLAN access and Layer 3 routing are accomplished on access devices.
- Deploy symmetric VXLANs on VTEPs.

5.5.2 Topology

Figure 5-5 Configuring EVPN-based Multi-Tenant Distributed VXLAN Networking (Symmetric Deployment)



5.5.3 Notes

- On device 1, device 2, and VTEPs, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that the unicast routes are reachable (omitted).
- On device 1, device 2, and VTEPs, configure the BGP-EVPN routing protocol to ensure that the four devices establish BGP neighbor relationships and exchange L2VPN EVPN information.
- On each VTEP, configure an EVI for BGP-EVPN.
- On the virtualized servers, configure VXLANs and specify the gateway address for the VMs.
- On each VTEP, configure and associate a loopback interface for the establishment of a tunnel.
- On each VTEP, configure the anycast gateway MAC address.
- On VTEP 1, create VXLAN instances 10, 20, and 100 and associate them with VLANs.
- On VTEP 2, create VXLAN instances 10 and 100, and associate them with VLANs.
- On each VTEP, create an overlay router interface and configure the VXLAN gateway IP address. Configure different VRF instances for the overlay router interfaces and identify tenants by VRF instance. Configure the anycast attribute for the overlay router interface so that all VXLAN gateways in the network use the same IP address/MAC address.

- On VTEP 1 and VTEP 2, create both VXLAN instances 11 and 101, configure the symmetric attribute, and configure the symmetric VXLANs as Layer 3 routing VXLANs in the VRF instances.
- On VTEP 1 and VTEP 2, create overlay router interfaces for VXLAN instances 11 and 101, configure different VRF instances for the overlay router interfaces, and configure VXLAN 11 and VXLAN 101 as the symmetric VXLANs of the VRF instances.
- On each VTEP, associate a VXLAN instance with the VXLAN routed interface to implement VXLAN routing.
- (Optional) Configure ARP suppression on VTEP 1 and VTEP 2 to reduce ARP packets flowing to VXLANs.
- (Optional) Configure ARP proxy on the overlay router interfaces associated with VXLAN instances 10, 20, and 100 on VTEP 1 and VTEP 2 so that the traffic of hosts in the same VXLAN is forwarded at Layer 3. This function needs to be enabled together with ARP suppression.
- (Optional) In ARP proxy deployment scenarios, configure the reduction of EVPN entry synchronization. On VXLAN instances 10, 20, and 100 on VTEP 1 and VTEP 2, configure the devices not to advertise or receive MAC-only EVPN type-2 routes.
  - Configure the devices not to advertise EVPN MAC-only type-2 routes to reduce EVPN route synchronization between devices.
  - Configure the devices not to deliver MAC entries synchronized by EVPNs to reduce the occupancy of hardware entry resources.

  There is no dependency between the two configurations. You can perform the configuration separately as required.
- (Optional) In scenarios in which ARP proxy is not deployed, configure the reduction of EVPN entry synchronization. On Layer 2 VNI VXLAN instances 10, 20, and 100 on VTEP 1 and VTEP 2, configure the devices not to advertise EVPN MAC-only type-2 routes and configure devices to extract MAC addresses from MAC-IP type-2 routes:
  - Configure the devices not to advertise EVPN MAC-only type-2 routes to reduce EVPN route synchronization between devices.
  - Configure the devices to extract MAC addresses from MAC-IP type-2 routes so that MAC entries can be learned even if the neighbor does not advertise MAC-only type-2 routes.

  The two configurations above must be used in combination. Otherwise, the devices cannot learn MAC entries from neighbors and Layer 2 forwarding cannot be implemented in VXLANs.

5.5.4 Procedure

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 0
VTEP1(cfg-if-Loopback 1)# ip address 1.1.1.1/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 2.2.2.2/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP1(cfg)# vtep
```

```
VTEP1(cfg-vtep)# source loopback 0
VTEP1(cfg-vtep)# arp suppress enable
VTEP1(cfg-vtep)# exit
VTEP1(cfg)# ip vrf vrf-10
VTEP1(cfg-vrf)# rd 10:10
VTEP1(cfg-vrf)# route-target both 1000:1000
VTEP1(cfg-vrf)# exit
VTEP1(cfg)# interface overlayrouter 10
VTEP1(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP1(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 10)# route-in-vni    //Optional. It is used to enable ARP proxy and needs to be used together with the arp suppress enable command.
VTEP1(cfg-if-OverlayRouter 10)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# router-interface overlayrouter 10
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# evpn mac advertise disable    //Optional.
VTEP1(cfg-vxlan)# evpn mac inactive    //Optional. In ARP proxy deployment scenarios, it is used to reduce the synchronization of EVPN entries.
VTEP1(cfg-vxlan)# evpn arp mac-learning enable    //Optional. In scenarios in which ARP proxy is not deployed, it is used to reduce the synchronization of EVPN entries.
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 20
VTEP1(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP1(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 20)# route-in-vni    //Optional. It is used to enable ARP proxy and needs to be used together with the arp suppress enable command.
VTEP1(cfg-if-OverlayRouter 20)# exit
VTEP1(cfg)# vxlan 20
VTEP1(cfg-vxlan)# extend-vlan 20
VTEP1(cfg-vxlan)# router-interface overlayrouter 20
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# evpn mac advertise disable    //Optional.
VTEP1(cfg-vxlan)# evpn mac inactive    //Optional. In ARP proxy deployment scenarios, it is used to reduce the synchronization of EVPN entries.
VTEP1(cfg-vxlan)# evpn arp mac-learning enable    //Optional. In scenarios with ARP proxy not deployed, it is used to reduce the synchronization of EVPN entries.
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 11
VTEP1(cfg-if-OverlayRouter 11)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 11)# exit
VTEP1(cfg)# vxlan 11
VTEP1(cfg-vxlan)# symmetric
```

```
VTEP1(cfg-vxlan)# router-interface overlayrouter 11
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# ip vrf vrf-20
VTEP1(cfg-vrf)# rd 20:20
VTEP1(cfg-vrf)# route-target both 2000:2000
VTEP1(cfg-vrf)# exit
VTEP1(cfg)# interface overlayrouter 100
VTEP1(cfg-if-OverlayRouter 100)# ip vrf binding vrf-20
VTEP1(cfg-if-OverlayRouter 100)# ip address 10.1.3.1/24
VTEP1(cfg-if-OverlayRouter 100)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 100)# route-in-vni //Optional. It is used to enable ARP proxy and needs to be used together with the arp suppress enable command.
VTEP1(cfg-if-OverlayRouter 100)# exit
VTEP1(cfg)# vxlan 100
VTEP1(cfg-vxlan)# extend-vlan 100
VTEP1(cfg-vxlan)# router-interface overlayrouter 100
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# evpn mac advertise disable    //Optional.
VTEP1(cfg-vxlan)# evpn mac inactive    //Optional. In ARP proxy deployment scenarios, it is used to reduce the synchronization of EVPN entries.
VTEP1(cfg-vxlan)# evpn arp mac-learning enable    //Optional. In scenarios in which ARP proxy is not deployed, it is used to reduce the synchronization of EVPN entries.
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 101
VTEP1(cfg-if-OverlayRouter 101)# ip vrf binding vrf-20
VTEP1(cfg-if-OverlayRouter 101)# exit
VTEP1(cfg)# vxlan 101
VTEP1(cfg-vxlan)# symmetric
VTEP1(cfg-vxlan)# router-interface overlayrouter 101
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# router-protocol bgp 64512
VTEP1(cfg-router)# bgp-instance router-id 2.2.2.2
VTEP1(cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP1(cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP1(cfg-router)# address-family-mode l2vpn evpn
VTEP1(cfg-router-af)# peer 3.3.3.3 activate
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# exit
VTEP1(cfg)# evpn
VTEP1(cfg-evpn)# vni 10
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 11
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
```

```
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 20
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 100
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 101
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
```

(2) Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP2(cfg)# interface loopback 0
VTEP2(cfg-if-Loopback 1)# ip address 4.4.4.4/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 3.3.3.3/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 0
VTEP2(cfg-vtep)# arp suppress enable
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# ip vrf vrf-10
VTEP2(cfg-vrf)# rd 10:10
VTEP2(cfg-vrf)# route-target both 1000:1000
VTEP2(cfg-vrf)# exit
VTEP2(cfg)# interface overlayrouter 10
VTEP2(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP2(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 10)# route-in-vni //Optional. It is used to enable ARP proxy and needs to be used together with the arp suppress enable command.
VTEP2(cfg-if-OverlayRouter 10)# exit
VTEP2(cfg)# vxlan 10
VTEP2(cfg-vxlan)# extend-vlan 10
VTEP2(cfg-vxlan)# router-interface overlayrouter 10
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# evpn mac advertise disable    //Optional.
VTEP2(cfg-vxlan)# evpn mac inactive    //Optional. In ARP proxy deployment scenarios, it is used to reduce the synchronization of EVPN entries.
```

VTEP2(cfg-vxlan)# evpn arp mac-learning enable //Optional. In scenarios in which ARP proxy is not deployed, it is used to reduce the synchronization of EVPN entries.
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 11
VTEP2(cfg-if-OverlayRouter 11)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 11)# exit
VTEP2(cfg)# vxlan 11
VTEP2(cfg-vxlan)# symmetric
VTEP2(cfg-vxlan)# router-interface overlayrouter 11
VTEP2(cfg-vxlan)# exit
VTEP2config)# ip vrf vrf-20
VTEP2(cfg-vrf)# rd 20:20
VTEP2(cfg-vrf)# route-target both 2000:2000
VTEP2(cfg-vrf)# exit
VTEP2(cfg)# interface overlayrouter 100
VTEP2(cfg-if-OverlayRouter 100)# ip vrf binding vrf-20
VTEP2(cfg-if-OverlayRouter 100)# ip address 10.1.3.1/24
VTEP2(cfg-if-OverlayRouter 100)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 100)# route-in-vni //Optional. It is used to enable ARP proxy and needs to be used together with the **arp suppress enable** command.
VTEP2(cfg-if-OverlayRouter 100)# exit
VTEP2(cfg)# vxlan 100
VTEP2(cfg-vxlan)# extend-vlan 100
VTEP2(cfg-vxlan)# router-interface overlayrouter 100
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# evpn mac advertise disable //Optional.
VTEP2(cfg-vxlan)# evpn mac inactive //Optional. In ARP proxy deployment scenarios, it is used to reduce the synchronization of EVPN entries.
VTEP2(cfg-vxlan)# evpn arp mac-learning enable //Optional. In scenarios in which ARP proxy is not deployed, it is used to reduce the synchronization of EVPN entries.
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 101
VTEP2(cfg-if-OverlayRouter 101)# ip vrf binding vrf-20
VTEP2(cfg-if-OverlayRouter 101)# exit
VTEP2(cfg)# vxlan 101
VTEP2(cfg-vxlan)# symmetric
VTEP2(cfg-vxlan)# router-interface overlayrouter 101
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 64512
VTEP2(cfg-router)# bgp-instance router-id 3.3.3.3
VTEP2(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP2(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP2(cfg-router)# address-family-mode l2vpn evpn
VTEP2(cfg-router-af)# peer 2.2.2.2 activate
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# exit

```
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 10
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 11
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 100
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 101
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
```

5.5.5 Verification

- After the commands above are run and the distributed symmetric VXLAN networking using dynamic tunnels is configured:
  - VMs in VRF-10 can ping each other successfully.
  - VMs in VRF-20 can ping each other successfully.
  - VMs in VRF-10 cannot ping those in VRF-20 and vice versa.
  - VMs can be migrated in the same VXLAN, they can access the Internet service successfully after migration, and the configurations do not need to be modified.
- Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 1 as an example.

```
VTEP1#display vxlan
VXLAN Total Count: 5
VXLAN Capacity     : 8000

VXLAN 10
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 10 (anycast)
      Extend VLAN           : 10
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                Source IP                                        Destination IP
Type
      ---------------------- ------------------------------------------------ ------------------------------------------------ -------
      OverlayTunnel 6145       1.1.1.1                                          4.4.4.4
dynamic
```

```
VXLAN 11
      Symmetric property   : TRUE
      Routed interface      : overlayrouter 11 (non-anycast)
      Extend VLAN           : -
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                  Source IP                                   Destination IP
Type
      ---------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145       1.1.1.1                                     4.4.4.4
dynamic

VXLAN 20
      Symmetric property   : FALSE
      Routed interface      : overlayrouter 20 (anycast)
      Extend VLAN           : 20
      VTEP Adjacency Count: 0

VXLAN 100
      Symmetric property   : FALSE
      Routed interface      : overlayrouter 100 (anycast)
      Extend VLAN           : 100
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                  Source IP                                   Destination IP
Type
      ---------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145       1.1.1.1                                     4.4.4.4
dynamic

VXLAN 101
      Symmetric property   : TRUE
      Routed interface      : overlayrouter 101 (non-anycast)
      Extend VLAN           : -
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                  Source IP                                   Destination IP
Type
      ---------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145       1.1.1.1                                     4.4.4.4
dynamic
```

5.5.6 Configuration Files

- VTEP 1 configuration file

```
!
ip vrf vrf-10
```

```
 rd 10:10
 route-target both 1000:1000
!
ip vrf vrf-20
 rd 20:20
 route-target both 2000:2000
!
vtep
 source loopback 0
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 0
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 1
 ip address 2.2.2.2 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
 route-in-vni    //Optional. It is used to enable ARP proxy and needs to be used together with the arp suppress enable command.
!
interface OverlayRouter 11
 ip vrf binding vrf-10
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
 route-in-vni    //Optional. It is used to enable ARP proxy and needs to be used together with the arp suppress enable command.
!
interface OverlayRouter 100
 ip vrf binding vrf-20
 ip address 10.1.3.1 255.255.255.0
 anycast-gateway
 route-in-vni    //Optional. It is used to enable ARP proxy and needs to be used together with the arp suppress enable command.
!
interface OverlayRouter 101
 ip vrf binding vrf-20
!
```

```
router-protocol bgp 64512
 bgp-instance router-id 2.2.2.2
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 3.3.3.3 remote-as 64512
 peer 3.3.3.3 update-source Loopback 1
 address-family-mode ipv4
  peer 3.3.3.3 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 3.3.3.3 activate
  peer 3.3.3.3 send-community extended
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 11
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
 vni 101
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 extend-vlan 10
 evpn mac advertise disable    //Optional.
 evpn mac inactive    //Optional. In ARP proxy deployment scenarios, it is used to reduce the
synchronization of EVPN entries.
 evpn arp mac-learning enable    //Optional. In scenarios in which ARP proxy is not deployed, it is used to
reduce the synchronization of EVPN entries.
 arp suppress enable
!
vxlan 11
 router-interface OverlayRouter 11
 symmetric
```

```
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
 evpn mac advertise disable    //Optional.
 evpn mac inactive    //Optional. In ARP proxy deployment scenarios, it is used to reduce the synchronization of EVPN entries.
 evpn arp mac-learning enable    //Optional. In scenarios in which ARP proxy is not deployed, it is used to reduce the synchronization of EVPN entries.
 arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 extend-vlan 100
 evpn mac advertise disable    //Optional.
 evpn mac inactive    //Optional. In ARP proxy deployment scenarios, it is used to reduce the synchronization of EVPN entries.
 evpn arp mac-learning enable    //Optional. In scenarios in which ARP proxy is not deployed, it is used to reduce the synchronization of EVPN entries.
 arp suppress enable
!
vxlan 101
 router-interface OverlayRouter 101
 symmetric
!
```

- VTEP 2 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target both 1000:1000
!
ip vrf vrf-20
 rd 20:20
 route-target both 2000:2000
!
vtep
 source loopback 0
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 0
 ip address 4.4.4.4 255.255.255.255
!
interface Loopback 1
 ip address 3.3.3.3 255.255.255.255
```

```
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
 route-in-vni    //Optional. It is used to enable ARP proxy and needs to be used together with the arp suppress enable command.
!
interface OverlayRouter 11
 ip vrf binding vrf-10
!
interface OverlayRouter 100
 ip vrf binding vrf-20
 ip address 10.1.3.1 255.255.255.0
 anycast-gateway
 route-in-vni    //Optional. It is used to enable ARP proxy and needs to be used together with the arp suppress enable command.
!
interface OverlayRouter 101
 ip vrf binding vrf-20
!
router-protocol bgp 64512
 bgp-instance router-id 3.3.3.3
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
 address-family-mode ipv4
  peer 2.2.2.2 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 2.2.2.2 activate
  peer 2.2.2.2 send-community extended
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 11
  rd auto
  route-target both auto
vni 100
  rd auto
```

```
   route-target both auto
  vni 101
   rd auto
   route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 extend-vlan 10
 evpn mac advertise disable    //Optional.
 evpn mac inactive    //Optional. In ARP proxy deployment scenarios, it is used to reduce the
synchronization of EVPN entries.
 evpn arp mac-learning enable    //Optional. In scenarios in which ARP proxy is not deployed, it is used to
reduce the synchronization of EVPN entries.
 arp suppress enable
!
vxlan 11
 router-interface OverlayRouter 11
 symmetric
!
vxlan 100
 router-interface OverlayRouter 100
 extend-vlan 100
 evpn mac advertise disable    //Optional.
 evpn mac inactive    //Optional. In ARP proxy deployment scenarios, it is used to reduce the
synchronization of EVPN entries.
 evpn arp mac-learning enable    //Optional. In scenarios in which ARP proxy is not deployed, it is used to
reduce the synchronization of EVPN entries.
 arp suppress enable
!
vxlan 101
 router-interface OverlayRouter 101
 symmetric
!
```

5.5.7 Common Errors

The anycast gateway MAC address conflicts with the MAC address of a device in VXLANs, resulting in a forwarding abnormality.

5.6 Configuring EVPN-based Single-Tenant VXLAN Routing

5.6.1 Requirements

In the topology as shown in Figure 5-6, VMs, access devices, and edge switches form a simple data center network:

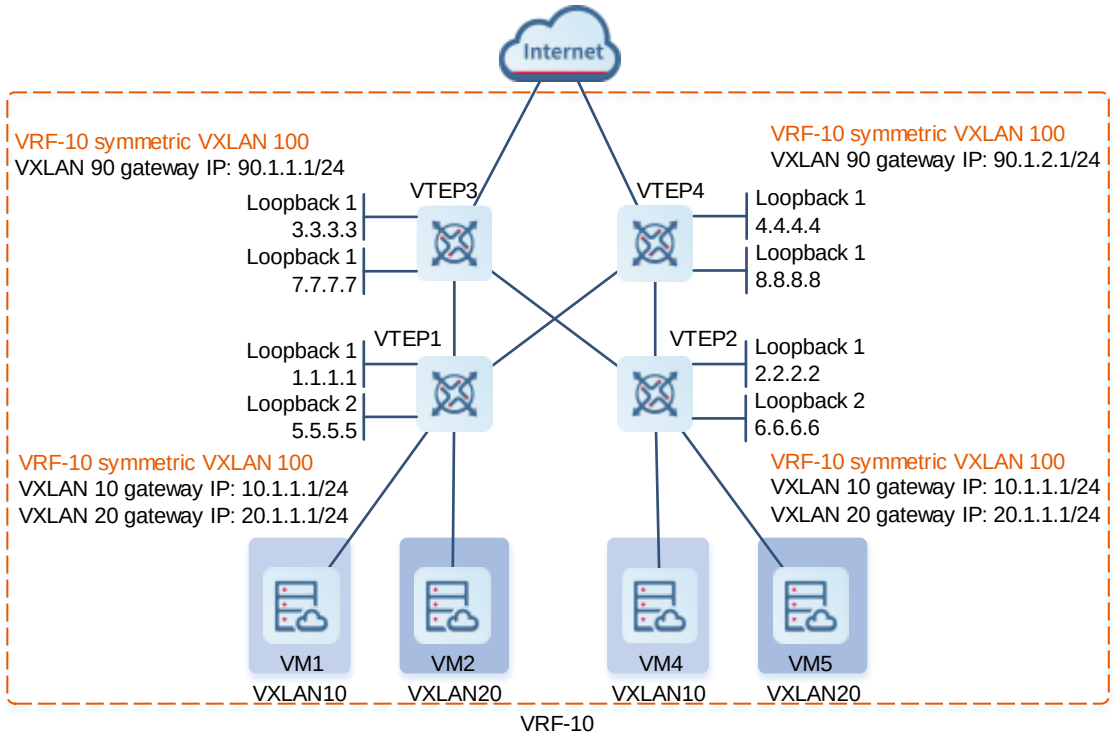
- VMs connect to the VXLAN through access devices VTEP 1 and VTEP 2.
- The data center network connects to the Internet through VTEP 3 and VTEP 4.

The following configurations are needed to support single-tenant application in the data center, improve the network scalability, and meet the networking requirements of large data centers:

- Configure distributed VXLAN networking using dynamic tunnels on VTEP 1 and VTEP 2, and configure VRF instances to allocate the VXLANs to tenants.
- Configure the anycast gateway on VTEP 1 and VTEP 2.
- Configure gateways on VTEP 3 and VTEP 4 for Layer 3 forwarding between VXLAN networks and non-VXLAN networks.
- Configure symmetric VXLAN networking on each VTEP, and configure VTEP 3 and VTEP 4 to import network routes to the VTEPs through the symmetric VXLAN.

5.6.2 Topology

Figure 5-6 Configuring EVPN-based Single-Tenant VXLAN Routing



5.6.3 Notes

- On each VTEP, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that the unicast routes are reachable (omitted).
- On each VTEP, configure the BGP-EVPN routing protocol to ensure that the devices establish BGP neighbor relationships (not between VTEP 3 and VTEP 4) and exchange L2VPN EVPN information.
- On each VTEP, configure an EVI for BGP-EVPN.
- On the virtualized servers, configure VXLANs and specify the gateway address for the VMs.
- On each VTEP, configure and associate a loopback interface for the establishment of a tunnel.
- On VTEP 1 and VTEP 2, configure the anycast gateway MAC address.

- On VTEP 1 and VTEP 2, create VXLAN instances 10 and 20 and associate them with VLANs.
- On VTEP 3 and VTEP 4, create VXLAN instance 90 and associate it with a VLAN.
- On VTEP 1 and VTEP 2, create overlay router interfaces for VXLAN instances 10 and 20 and configure the VXLAN gateway IP address. Configure the same VRF instance for the overlay router interfaces and configure the anycast attribute.
- On each VTEP, create VXLAN instance 100, configure the symmetric attribute, and configure the VXLAN instance as the Layer 3 routing VXLAN of the VRF instance.
- On each VTEP, create an overlay router interface for VXLAN instance 100, configure the same VRF instance for the overlay router interfaces, and configure VXLAN instance 100 as the symmetric VXLAN of the VRF instance.
- On each VTEP, associate a VXLAN instance with the overlay router interface to implement VXLAN routing.
- (Optional) Configure ARP suppression on VTEP 1 and VTEP 2 to reduce ARP packets flowing to VXLANs.

5.6.4 Procedure

(1) Configure VTEP 1.

```
VTEP> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 1.1.1.1/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# interface loopback 2
VTEP1(cfg-if-Loopback 2)# ip address 5.5.5.5/32
VTEP1(cfg-if-Loopback 2)# exit
VTEP1(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP1(cfg)# vtep
VTEP1(cfg-vtep)# source loopback 2
VTEP1(cfg-vtep)# arp suppress enable
VTEP1(cfg-vtep)# exit
VTEP1(cfg)# ip vrf vrf-10
VTEP1(cfg-vrf)# rd 10:10
VTEP1(cfg-vrf)# route-target export 1000:1000
VTEP1(cfg-vrf)# exit
VTEP1(cfg)# interface overlayrouter 10
VTEP1(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP1(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 10)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# router-interface overlayrouter 10
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 20
VTEP1(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
```

```
VTEP1(cfg-if-OverlayRouter 20)# ip address 20.1.1.1/24
VTEP1(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 20)# exit
VTEP1(cfg)# vxlan 20
VTEP1(cfg-vxlan)# extend-vlan 20
VTEP1(cfg-vxlan)# router-interface overlayrouter 20
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 100
VTEP1(cfg-if-OverlayRouter 100)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 100)# exit
VTEP1(cfg)# vxlan 100
VTEP1(cfg-vxlan)# symmetric
VTEP1(cfg-vxlan)# router-interface overlayrouter 100
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# router-protocol bgp 64512
VTEP1(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP1(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP1(cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP1(cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP1(cfg-router)# peer 4.4.4.4 remote-as 64512
VTEP1(cfg-router)# peer 4.4.4.4 update-source loopback 1
VTEP1(cfg-router)# address-family-mode l2vpn evpn
VTEP1(cfg-router-af)# peer 2.2.2.2 activate
VTEP1(cfg-router-af)# peer 3.3.3.3 activate
VTEP1(cfg-router-af)# peer 4.4.4.4 activate
VTEP1(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP1(cfg-router-af)# redistribute-route connected
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# exit
VTEP1(cfg)# evpn
VTEP1(cfg-evpn)# vni 10
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 20
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 100
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# route-target import 1000:1000
VTEP1(cfg-evpn-vni)# exit
```

(2) Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 2.2.2.2/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# interface loopback 2
VTEP2(cfg-if-Loopback 2)# ip address 6.6.6.6/32
VTEP2(cfg-if-Loopback 2)# exit
VTEP2(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 2
VTEP2(cfg-vtep)# arp suppress enable
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# ip vrf vrf-10
VTEP2(cfg-vrf)# rd 10:10
VTEP2(cfg-vrf)# route-target export 1000:1000
VTEP2(cfg-vrf)# exit
VTEP2(cfg)# interface overlayrouter 10
VTEP2(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP2(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 10)# exit
VTEP2(cfg)# vxlan 10
VTEP2(cfg-vxlan)# extend-vlan 10
VTEP2(cfg-vxlan)# router-interface overlayrouter 10
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 20
VTEP2(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 20)# ip address 20.1.1.1/24
VTEP2(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 20)# exit
VTEP2(cfg)# vxlan 20
VTEP2(cfg-vxlan)# extend-vlan 20
VTEP2(cfg-vxlan)# router-interface overlayrouter 20
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 100
VTEP2(cfg-if-OverlayRouter 100)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 100)# exit
VTEP2(cfg)# vxlan 100
VTEP2(cfg-vxlan)# symmetric
VTEP2(cfg-vxlan)# router-interface overlayrouter 100
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 64512
```

```
VTEP2(cfg-router)# peer 1.1.1.1 remote-as 64512
VTEP2(cfg-router)# peer 1.1.1.1 update-source loopback 1
VTEP2(cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP2(cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP2(cfg-router)# peer 4.4.4.4 remote-as 64512
VTEP2(cfg-router)# peer 4.4.4.4 update-source loopback 1
VTEP2(cfg-router)# address-family-mode l2vpn evpn
VTEP2(cfg-router-af)# peer 1.1.1.1 activate
VTEP2(cfg-router-af)# peer 3.3.3.3 activate
VTEP2(cfg-router-af)# peer 4.4.4.4 activate
VTEP2(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP2(cfg-router-af)# redistribute-route connected
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# exit
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 10
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 20
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 100
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# route-target import 1000:1000
VTEP2(cfg-evpn-vni)# exit
```

(3) Configure VTEP 3.

```
VTEP3> privilege
VTEP3# config-system terminal
VTEP3(cfg)# interface loopback 1
VTEP3(cfg-if-Loopback 1)# ip address 3.3.3.3/32
VTEP3(cfg-if-Loopback 1)# exit
VTEP3(cfg)# interface loopback 2
VTEP3(cfg-if-Loopback 2)# ip address 7.7.7.7/32
VTEP3(cfg-if-Loopback 2)# exit
VTEP3(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP3(cfg)# vtep
VTEP3(cfg-vtep)# source loopback 2
VTEP3(cfg-vtep)# arp suppress enable
VTEP3(cfg-vtep)# exit
VTEP3(cfg)# ip vrf vrf-10
VTEP3(cfg-vrf)# rd 10:10
```

```
VTEP3(cfg-vrf)# route-target export 1000:1000
VTEP3(cfg-vrf)# exit
VTEP3(cfg)# interface overlayrouter 90
VTEP3(cfg-if-OverlayRouter 90)# ip vrf binding vrf-10
VTEP3(cfg-if-OverlayRouter 90)# ip address 90.1.1.1/24
VTEP3(cfg-if-OverlayRouter 90)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 90)# exit
VTEP3(cfg)# vxlan 90
VTEP3(cfg-vxlan)# extend-vlan 90
VTEP3(cfg-vxlan)# router-interface overlayrouter 90
VTEP3(cfg-vxlan)# arp suppress enable
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# interface overlayrouter 100
VTEP3(cfg-if-OverlayRouter 100)# ip vrf binding vrf-10
VTEP3(cfg-if-OverlayRouter 100)# exit
VTEP3(cfg)# vxlan 100
VTEP3(cfg-vxlan)# symmetric
VTEP3(cfg-vxlan)# router-interface overlayrouter 100
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# router-protocol bgp 64512
VTEP3(cfg-router)# peer 1.1.1.1 remote-as 64512
VTEP3(cfg-router)# peer 1.1.1.1 update-source loopback 1
VTEP3(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP3(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP3(cfg-router)# address-family-mode l2vpn evpn
VTEP3(cfg-router-af)# peer 1.1.1.1 activate
VTEP3(cfg-router-af)# peer 2.2.2.2 activate
VTEP3(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP3(cfg-router-af)# redistribute-route static
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# exit
VTEP3(cfg)# evpn
VTEP3(cfg-evpn)# vni 90
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn)# vni 100
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# route-target import 1000:1000
VTEP3(cfg-evpn-vni)# exit
```

(4) Configure VTEP 4.

```
VTEP4> privilege
VTEP4# config-system terminal
VTEP4(cfg)# interface loopback 1
```

```
VTEP4(cfg-if-Loopback 1)# ip address 4.4.4.4/32
VTEP4(cfg-if-Loopback 1)# exit
VTEP4(cfg)# interface loopback 2
VTEP4(cfg-if-Loopback 1)# ip address 8.8.8.8/32
VTEP4(cfg-if-Loopback 1)# exit
VTEP4(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP4(cfg)# vtep
VTEP4(cfg-vtep)# source loopback 2
VTEP4(cfg-vtep)# arp suppress enable
VTEP4(cfg-vtep)# exit
VTEP4(cfg)# ip vrf vrf-10
VTEP4(cfg-vrf)# rd 10:10
VTEP4(cfg-vrf)# route-target export 1000:1000
VTEP4(cfg-vrf)# exit
VTEP4(cfg)# interface overlayrouter 90
VTEP4(cfg-if-OverlayRouter 90)# ip vrf binding vrf-10
VTEP4(cfg-if-OverlayRouter 90)# ip address 90.1.2.1/24
VTEP4(cfg-if-OverlayRouter 90)# exit
VTEP4(cfg)# vxlan 90
VTEP4(cfg-vxlan)# extend-vlan 90
VTEP4(cfg-vxlan)# router-interface overlayrouter 90
VTEP4(cfg-vxlan)# arp suppress enable
VTEP4(cfg-vxlan)# exit
VTEP4(cfg)# interface overlayrouter 100
VTEP4(cfg-if-OverlayRouter 100)# ip vrf binding vrf-10
VTEP4(cfg-if-OverlayRouter 100)# exit
VTEP4(cfg)# vxlan 100
VTEP4(cfg-vxlan)# symmetric
VTEP4(cfg-vxlan)# router-interface overlayrouter 100
VTEP4(cfg-vxlan)# exit
VTEP4(cfg)# router-protocol bgp 64512
VTEP4(cfg-router)# peer 1.1.1.1 remote-as 64512
VTEP4(cfg-router)# peer 1.1.1.1 update-source loopback 1
VTEP4(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP4(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP4(cfg-router)# address-family-mode l2vpn evpn
VTEP4(cfg-router-af)# peer 1.1.1.1 activate
VTEP4(cfg-router-af)# peer 2.2.2.2 activate
VTEP4(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP4(cfg-router-af)# exit
VTEP4(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP4(cfg-router-af)# redistribute-route static
VTEP4(cfg-router-af)# exit
VTEP4(cfg-router)# exit
VTEP4(cfg)# evpn
VTEP4(cfg-evpn)#vni 90
```

```
VTEP4(cfg-evpn-vni)# rd auto
VTEP4(cfg-evpn-vni)# route-target both auto
VTEP4(cfg-evpn)#vni 100
VTEP4(cfg-evpn-vni)# rd auto
VTEP4(cfg-evpn-vni)# route-target both auto
VTEP4(cfg-evpn-vni)# route-target import 1000:1000
VTEP4(cfg-evpn-vni)# exit
```

5.6.5 Verification

- VM 1, VM 2, VM 3, and VM 4 can ping each other.
- VMs can be migrated in the same VXLAN, they can access the Internet service successfully after migration, and the configurations do not need to be modified.
- Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 3 as an example.

```
VTEP3# display vxlan
VXLAN Total Count: 2
VXLAN Capacity     : 8000

VXLAN 90
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 90 (anycast)
      Extend VLAN           : 90
      VTEP Adjacency Count: 0

VXLAN 100
      Symmetric property    : TRUE
      Routed interface      : overlayrouter 100 (non-anycast)
      Extend VLAN           : -
      VTEP Adjacency Count: 2
      VTEP Adjacency List :
      Interface                Source IP                                    Destination IP
Type
      ---------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145       7.7.7.7                                      5.5.5.5
dynamic
      OverlayTunnel 6146       7.7.7.7                                      6.6.6.6
dynamic
```

5.6.6 Configuration Files

- VTEP 1 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target export 1000:1000
!
```

```
vtep
 source loopback 2
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 1
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 2
 ip address 5.5.5.5 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 20.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-10
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
 peer 3.3.3.3 remote-as 64512
 peer 3.3.3.3 update-source Loopback 1
 peer 4.4.4.4 remote-as 64512
 peer 4.4.4.4 update-source Loopback 1
 address-family-mode ipv4
  peer 2.2.2.2 activate
  peer 3.3.3.3 activate
  peer 4.4.4.4 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  advertise ipv4 unicast-af-modifier
  peer 2.2.2.2 activate
  peer 2.2.2.2 send-community extended
  peer 3.3.3.3 activate
```

```
   peer 3.3.3.3 send-community extended
   peer 4.4.4.4 activate
   peer 4.4.4.4 send-community extended
   exit-address-family-mode
   !
  address-family-mode ipv4 vrf vrf-10
   redistribute-route connected
   exit-address-family-mode
!
evpn
  vni 10
   rd auto
   route-target both auto
  vni 20
   rd auto
   route-target both auto
  vni 100
   rd auto
   route-target both auto
   route-target import 1000:1000
!
vxlan 10
  router-interface OverlayRouter 10
  extend-vlan 10
  arp suppress enable
!
vxlan 20
  router-interface OverlayRouter 20
  extend-vlan 20
  arp suppress enable
!
vxlan 100
  router-interface OverlayRouter 100
  symmetric
!
```

- VTEP 2 configuration file

```
!
ip vrf vrf-10
  rd 10:10
  route-target export 1000:1000
!
vtep
  source loopback 2
  arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
```

```
!
interface Loopback 1
 ip address 2.2.2.2 255.255.255.255
!
interface Loopback 2
 ip address 6.6.6.6 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 20.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-10
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 peer 3.3.3.3 remote-as 64512
 peer 3.3.3.3 update-source Loopback 1
 peer 4.4.4.4 remote-as 64512
 peer 4.4.4.4 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.1 activate
  peer 3.3.3.3 activate
  peer 4.4.4.4 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  advertise ipv4 unicast-af-modifier
  peer 1.1.1.1 activate
  peer 1.1.1.1 send-community extended
  peer 3.3.3.3 activate
  peer 3.3.3.3 send-community extended
  peer 4.4.4.4 activate
  peer 4.4.4.4 send-community extended
  exit-address-family-mode
  !
```

```
 address-family-mode ipv4 vrf vrf-10
  redistribute-route connected
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
  route-target import 1000:1000
!
vxlan 10
 router-interface OverlayRouter 10
 extend-vlan 10
 arp suppress enable
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
 arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 symmetric
!
```

- VTEP 3 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target export 1000:1000
!
vtep
 source loopback 2
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 1
 ip address 3.3.3.3 255.255.255.255
!
interface Loopback 2
```

```
 ip address 7.7.7.7 255.255.255.255
!
interface OverlayRouter 90
 ip vrf binding vrf-10
 ip address 90.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-10
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
address-family-mode ipv4
   peer 1.1.1.1 activate
   peer 2.2.2.2 activate
   exit-address-family-mode
 address-family-mode l2vpn evpn
   advertise ipv4 unicast-af-modifier
   peer 1.1.1.1 activate
   peer 1.1.1.1 send-community extended
   peer 2.2.2.2 activate
   peer 2.2.2.2 send-community extended
   exit-address-family-mode
   !
 address-family-mode ipv4 vrf vrf-10
   redistribute-route static
   exit-address-family-mode
!
evpn
 vni 90
   rd auto
   route-target both auto
 vni 100
   rd auto
   route-target both auto
   route-target import 1000:1000
!
vxlan 90
 router-interface OverlayRouter 90
```

```
 extend-vlan 90
 arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 symmetric
!
```

- VTEP 4 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target export 1000:1000
!
vtep
 source loopback 2
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 1
 ip address 4.4.4.4 255.255.255.255
!
interface Loopback 2
 ip address 8.8.8.8 255.255.255.255
!
interface OverlayRouter 90
 ip vrf binding vrf-10
 ip address 90.1.2.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-10
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
address-family-mode ipv4
   peer 1.1.1.1 activate
   peer 2.2.2.2 activate
   exit-address-family-mode
```

```
 address-family-mode l2vpn evpn
  advertise ipv4 unicast-af-modifier
  peer 1.1.1.1 activate
  peer 1.1.1.1 send-community extended
  peer 2.2.2.2 activate
  peer 2.2.2.2 send-community extended
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-10
  redistribute-route static
  exit-address-family-mode
!
evpn
 vni 90
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
  route-target import 1000:1000
!
vxlan 90
 router-interface OverlayRouter 90
 extend-vlan 90
 arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 symmetric
!
```

5.7 Configuring EVPN-based Multi-Tenant VXLAN Routing

5.7.1 Requirements

In the topology as shown in Figure 5-7, VMs, access devices, and edge switches form a simple data center network:

- VMs connect to the VXLAN through access devices VTEP 1 and VTEP 2.
- The data center network connects to the Internet through VTEP 3 and VTEP 4.

The following configurations are needed to support multi-tenant application in the data center, improve the network scalability, and meet the networking requirements of large data centers:

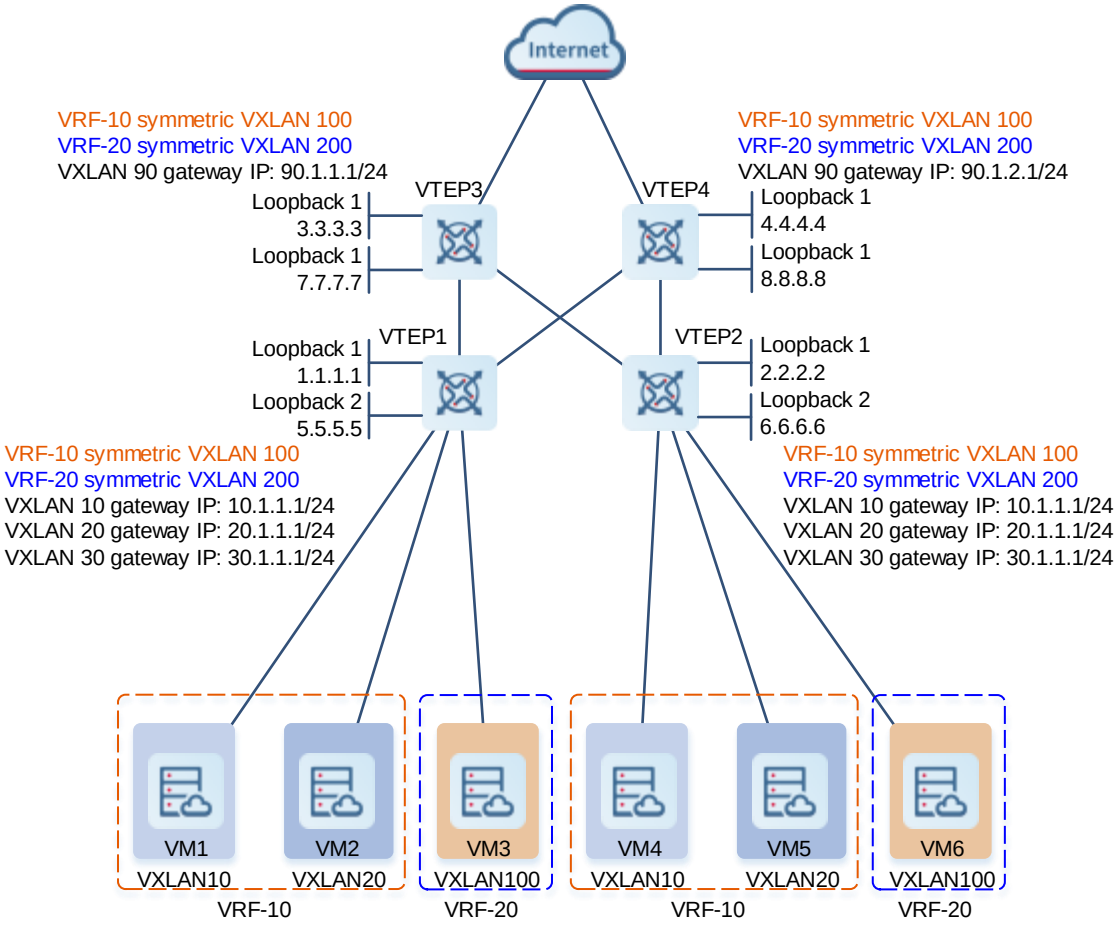
- Configure distributed VXLAN networking using dynamic tunnels on VTEP 1 and VTEP 2, and configure VRF instances to allocate the VXLANs to tenants.
- Configure the anycast gateway on VTEP 1 and VTEP 2.
- Configure gateways on VTEP 3 and VTEP 4 for Layer 3 forwarding between VXLAN networks and

non-VXLAN networks.

- Configure symmetric VXLAN networking on each VTEP, and configure VTEP 3 and VTEP 4 to import network routes to the VTEPs through the symmetric VXLAN.

5.7.2 Topology

Figure 5-7 Configuring EVPN-based Multi-Tenant VXLAN Routing



5.7.3 Notes

- On each VTEP, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that the unicast routes are reachable (omitted).
- On each VTEP, configure the BGP-EVPN routing protocol to ensure that the devices establish BGP neighbor relationships (not between VTEP 3 and VTEP 4) and exchange L2VPN EVPN information.
- On each VTEP, configure an EVI for BGP-EVPN.
- On the virtualized servers, configure VXLANs and specify the gateway address for the VMs.
- On each VTEP, configure and associate a loopback interface for the establishment of a tunnel.
- On VTEP 1 and VTEP 2, configure the anycast gateway MAC address.
- On VTEP 1 and VTEP 2, create VXLAN instances 10, 20, and 30 and associate them with VXLANs.
- On VTEP 3 and VTEP 4, create VXLAN instance 90 and associate it with a VLAN.

- On VTEP 1 and VTEP 2, create overlay router interfaces for VXLAN instances 10, 20, and 30 and configure the VXLAN gateway IP address. Configure different VRF instances for the overlay router interfaces and configure the anycast attribute.
- On each VTEP, create VXLAN instances100 and 200, configure the symmetric attribute, and configure the VXLAN instances as the Layer 3 routing VXLANs of VRF instances.
- On each VTEP, create overlay router interfaces for VXLAN instances 100 and 200, configure VRF instances for the overlay router interfaces, and configure the VXLAN instances as the symmetric VXLANs of the VRF instances.
- On each VTEP, associate a VXLAN instance with the overlay router interface to implement VXLAN routing.
- (Optional) Configure ARP suppression on VTEP 1 and VTEP 2 to reduce ARP packets flowing to VXLANs.

5.7.4 Procedure

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 1.1.1.1/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# interface loopback 2
VTEP1(cfg-if-Loopback 2)# ip address 5.5.5.5/32
VTEP1(cfg-if-Loopback 2)# exit
VTEP1(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP1(cfg)# vtep
VTEP1(cfg-vtep)# source loopback 2
VTEP1(cfg-vtep)# arp suppress enable
VTEP1(cfg-vtep)# exit
VTEP1(cfg)# ip vrf vrf-10
VTEP1(cfg-vrf)# rd 10:10
VTEP1(cfg-vrf)# route-target export 1000:1000
VTEP1(cfg-vrf)# exit
VTEP1(cfg)# interface overlayrouter 10
VTEP1(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP1(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 10)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# router-interface overlayrouter 10
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 20
VTEP1(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 20)# ip address 20.1.1.1/24
VTEP1(cfg-if-OverlayRouter 20)# anycast-gateway
```

```
VTEP1(cfg-if-OverlayRouter 20)# exit
VTEP1(cfg)# vxlan 20
VTEP1(cfg-vxlan)# extend-vlan 20
VTEP1(cfg-vxlan)# router-interface overlayrouter 20
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 100
VTEP1(cfg-if-OverlayRouter 100)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 100)# ip address 100.1.4.1/24
VTEP1(cfg-if-OverlayRouter 100)# exit
VTEP1(cfg)# vxlan 100
VTEP1(cfg-vxlan)# symmetric
VTEP1(cfg-vxlan)# router-interface overlayrouter 100
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# ip vrf vrf-20
VTEP1(cfg-vrf)# rd 20:20
VTEP1(cfg-vrf)# route-target export 2000:2000
VTEP1(cfg-vrf)# exit
VTEP1(cfg)# interface overlayrouter 30
VTEP1(cfg-if-OverlayRouter 30)# ip vrf binding vrf-20
VTEP1(cfg-if-OverlayRouter 30)# ip address 30.1.1.1/24
VTEP1(cfg-if-OverlayRouter 30)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 30)# exit
VTEP1(cfg)# vxlan 30
VTEP1(cfg-vxlan)# extend-vlan 30
VTEP1(cfg-vxlan)# router-interface overlayrouter 30
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 200
VTEP1(cfg-if-OverlayRouter 200)# ip vrf binding vrf-20
VTEP1(cfg-if-OverlayRouter 200)# ip address 200.1.4.1/24
VTEP1(cfg-if-OverlayRouter 200)# exit
VTEP1(cfg)# vxlan 200
VTEP1(cfg-vxlan)# symmetric
VTEP1(cfg-vxlan)# router-interface overlayrouter 200
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# router-protocol bgp 64512
VTEP1(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP1(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP1(cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP1(cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP1(cfg-router)# peer 4.4.4.4 remote-as 64512
VTEP1(cfg-router)# peer 4.4.4.4 update-source loopback 1
VTEP1(cfg-router)# address-family-mode l2vpn evpn
VTEP1(cfg-router-af)# peer 2.2.2.2 activate
VTEP1(cfg-router-af)# peer 3.3.3.3 activate
```

```
VTEP1(cfg-router-af)# peer 4.4.4.4 activate
VTEP1(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP1(cfg-router-af)# redistribute-route connected
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# address-family-mode ipv4 vrf vrf-20
VTEP1(cfg-router-af)# redistribute-route connected
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# exit
VTEP1(cfg)# evpn
VTEP1(cfg-evpn)# vni 10
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 20
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 30
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 100
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# route-target import 1000:1000
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 200
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# route-target import 2000:2000
VTEP1(cfg-evpn-vni)# exit
```

(2) Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 2.2.2.2/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# interface loopback 2
VTEP2(cfg-if-Loopback 2)# ip address 6.6.6.6/32
VTEP2(cfg-if-Loopback 2)# exit
VTEP2(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 2
VTEP2(cfg-vtep)# arp suppress enable
```

```
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# ip vrf vrf-10
VTEP2(cfg-vrf)# rd 10:10
VTEP2(cfg-vrf)# route-target export 1000:1000
VTEP2(cfg-vrf)# exit
VTEP2(cfg)# interface overlayrouter 10
VTEP2(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP2(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 10)# exit
VTEP2(cfg)# vxlan 10
VTEP2(cfg-vxlan)# extend-vlan 10
VTEP2(cfg-vxlan)# router-interface overlayrouter 10
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 20
VTEP2(cfg-if-OverlayRouter 20)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 20)# ip address 20.1.1.1/24
VTEP2(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 20)# exit
VTEP2(cfg)# vxlan 20
VTEP2(cfg-vxlan)# extend-vlan 20
VTEP2(cfg-vxlan)# router-interface overlayrouter 20
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 100
VTEP2(cfg-if-OverlayRouter 100)# ip vrf binding vrf-10
VTEP2(cfg-if-OverlayRouter 100)# ip address 100.1.3.1/24
VTEP2(cfg-if-OverlayRouter 100)# exit
VTEP2(cfg)# vxlan 100
VTEP2(cfg-vxlan)# symmetric
VTEP2(cfg-vxlan)# router-interface overlayrouter 100
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# ip vrf vrf-20
VTEP2(cfg-vrf)# rd 20:20
VTEP2(cfg-vrf)# route-target export 2000:2000
VTEP2(cfg-vrf)# exit
VTEP2(cfg)# interface overlayrouter 30
VTEP2(cfg-if-OverlayRouter 30)# ip vrf binding vrf-20
VTEP2(cfg-if-OverlayRouter 30)# ip address 30.1.1.1/24
VTEP2(cfg-if-OverlayRouter 30)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 30)# exit
VTEP2(cfg)# vxlan 30
VTEP2(cfg-vxlan)# extend-vlan 30
VTEP2(cfg-vxlan)# router-interface overlayrouter 30
VTEP2(cfg-vxlan)# arp suppress enable
```

```
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 200
VTEP2(cfg-if-OverlayRouter 200)# ip vrf binding vrf-20
VTEP2(cfg-if-OverlayRouter 200)# ip address 200.1.3.1/24
VTEP2(cfg-if-OverlayRouter 200)# exit
VTEP2(cfg)# vxlan 200
VTEP2(cfg-vxlan)# symmetric
VTEP2(cfg-vxlan)# router-interface overlayrouter 200
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 64512
VTEP2(cfg-router)# peer 1.1.1.1 remote-as 64512
VTEP2(cfg-router)# peer 1.1.1.1 update-source loopback 1
VTEP2(cfg-router)# peer 3.3.3.3 remote-as 64512
VTEP2(cfg-router)# peer 3.3.3.3 update-source loopback 1
VTEP2(cfg-router)# peer 4.4.4.4 remote-as 64512
VTEP2(cfg-router)# peer 4.4.4.4 update-source loopback 1
VTEP2(cfg-router)# address-family-mode l2vpn evpn
VTEP2(cfg-router-af)# peer 1.1.1.1 activate
VTEP2(cfg-router-af)# peer 3.3.3.3 activate
VTEP2(cfg-router-af)# peer 4.4.4.4 activate
VTEP2(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP2(cfg-router-af)# redistribute-route connected
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# address-family-mode ipv4 vrf vrf-20
VTEP2(cfg-router-af)# redistribute-route connected
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# exit
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 10
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 20
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 30
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 100
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# route-target import 1000:1000
```

```
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 200
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# route-target import 2000:2000
VTEP2(cfg-evpn-vni)# exit
```

(3) Configure VTEP 3.

```
VTEP3> privilege
VTEP3# config-system terminal
VTEP3(cfg)# interface loopback 1
VTEP3(cfg-if-Loopback 1)# ip address 3.3.3.3/32
VTEP3(cfg-if-Loopback 1)# exit
VTEP3(cfg)# interface loopback 2
VTEP3(cfg-if-Loopback 2)# ip address 7.7.7.7/32
VTEP3(cfg-if-Loopback 2)# exit
VTEP3(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP3(cfg)# vtep
VTEP3(cfg-vtep)# source loopback 2
VTEP3(cfg-vtep)# arp suppress enable
VTEP3(cfg-vtep)# exit
VTEP3(cfg)# ip vrf vrf-30
VTEP3(cfg-vrf)# rd 30:30
VTEP3(cfg-vrf)# route-target export 3000:3000
VTEP3(cfg-vrf)# route-target import 1000:1000
VTEP3(cfg-vrf)# route-target import 2000:2000
VTEP3(cfg-vrf)# exit
VTEP3(cfg)# interface overlayrouter 90
VTEP3(cfg-if-OverlayRouter 90)# ip vrf binding vrf-30
VTEP3(cfg-if-OverlayRouter 90)# ip address 90.1.1.1/24
VTEP3(cfg-if-OverlayRouter 90)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 90)# exit
VTEP3(cfg)# vxlan 90
VTEP3(cfg-vxlan)# extend-vlan 90
VTEP3(cfg-vxlan)# router-interface overlayrouter 90
VTEP3(cfg-vxlan)# arp suppress enable
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# ip vrf vrf-10
VTEP3(cfg-vrf)# rd 10:10
VTEP3(cfg-vrf)# route-target export 1000:1000
VTEP3(cfg-vrf)# route-target import 3000:3000
VTEP3(cfg-vrf)# exit
VTEP3(cfg)# interface overlayrouter 100
VTEP3(cfg-if-OverlayRouter 100)# ip vrf binding vrf-10
VTEP3(cfg-if-OverlayRouter 100)# ip address 100.1.1.1/24
VTEP3(cfg-if-OverlayRouter 100)# exit
VTEP3(cfg)# vxlan 100
```

```
VTEP3(cfg-vxlan)# symmetric
VTEP3(cfg-vxlan)# router-interface overlayrouter 100
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# ip vrf vrf-20
VTEP3(cfg-vrf)# rd 20:20
VTEP3(cfg-vrf)# route-target export 2000:2000
VTEP3(cfg-vrf)# route-target import 3000:3000
VTEP3(cfg-vrf)# exit
VTEP3(cfg)# interface overlayrouter 200
VTEP3(cfg-if-OverlayRouter 200)# ip vrf binding vrf-20
VTEP3(cfg-if-OverlayRouter 200)# ip address 200.1.1.1/24
VTEP3(cfg-if-OverlayRouter 200)# exit
VTEP3(cfg)# vxlan 200
VTEP3(cfg-vxlan)# symmetric
VTEP3(cfg-vxlan)# router-interface overlayrouter 200
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# router-protocol bgp 64512
VTEP3(cfg-router)# peer 1.1.1.1 remote-as 64512
VTEP3(cfg-router)# peer 1.1.1.1 update-source loopback 1
VTEP3(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP3(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP3(cfg-router)# address-family-mode l2vpn evpn
VTEP3(cfg-router-af)# peer 1.1.1.1 activate
VTEP3(cfg-router-af)# peer 2.2.2.2 activate
VTEP3(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# address-family-mode ipv4 vrf vrf-20
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# address-family-mode ipv4 vrf vrf-30
VTEP3(cfg-router-af)# redistribute-route static
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# exit
VTEP3(cfg)# evpn
VTEP3(cfg-evpn)# vni 100
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# route-target import 3000:3000
VTEP3(cfg-evpn-vni)# exit
VTEP3(cfg-evpn)# vni 200
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# route-target import 3000:3000
VTEP3(cfg-evpn-vni)# exit
```

(4) Configure VTEP 4.

```
VTEP4> privilege
VTEP4# config-system terminal
VTEP4(cfg)# interface loopback 1
VTEP4(cfg-if-Loopback 1)# ip address 4.4.4.4/32
VTEP4(cfg-if-Loopback 1)# exit
VTEP4(cfg)# interface loopback 2
VTEP4(cfg-if-Loopback 2)# ip address 8.8.8.8/32
VTEP4(cfg-if-Loopback 2)# exit
VTEP4(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP4(cfg)# vtep
VTEP4(cfg-vtep)# source loopback 2
VTEP4(cfg-vtep)# arp suppress enable
VTEP4(cfg-vtep)# exit
VTEP4(cfg)# ip vrf vrf-30
VTEP4(cfg-vrf)# rd 30:30
VTEP4(cfg-vrf)# route-target export 3000:3000
VTEP4(cfg-vrf)# route-target import 1000:1000
VTEP4(cfg-vrf)# route-target import 2000:2000
VTEP4(cfg-vrf)# exit
VTEP4(cfg)# interface overlayrouter 90
VTEP4(cfg-if-OverlayRouter 90)# ip vrf binding vrf-30
VTEP4(cfg-if-OverlayRouter 90)# ip address 90.1.2.1/24
VTEP4(cfg-if-OverlayRouter 90)# anycast-gateway
VTEP4(cfg-if-OverlayRouter 90)# exit
VTEP4(cfg)# vxlan 90
VTEP4(cfg-vxlan)# extend-vlan 90
VTEP4(cfg-vxlan)# router-interface OverlayRouter 90
VTEP4(cfg-vxlan)# arp suppress enable
VTEP4(cfg-vxlan)# exit
VTEP4(cfg)# ip vrf vrf-10
VTEP4(cfg-vrf)# rd 10:10
VTEP4(cfg-vrf)# route-target export 1000:1000
VTEP4(cfg-vrf)# route-target import 3000:3000
VTEP4(cfg-vrf)# exit
VTEP4(cfg)# interface overlayrouter 100
VTEP4(cfg-if-OverlayRouter 100)# ip vrf binding vrf-10
VTEP4(cfg-if-OverlayRouter 100)# ip address 100.1.2.1/24
VTEP4(cfg-if-OverlayRouter 100)# exit
VTEP4(cfg)# vxlan 100
VTEP4(cfg-vxlan)# symmetric
VTEP4(cfg-vxlan)# router-interface OverlayRouter 100
VTEP4(cfg-vxlan)# exit
VTEP4(cfg)# ip vrf vrf-20
VTEP4(cfg-vrf)# rd 20:20
VTEP4(cfg-vrf)# route-target export 2000:2000
VTEP4(cfg-vrf)# route-target import 3000:3000
```

```
VTEP4(cfg-vrf)# exit
VTEP4(cfg)# interface overlayrouter 200
VTEP4(cfg-if-OverlayRouter 200)# ip vrf binding vrf-20
VTEP4(cfg-if-OverlayRouter 200)# ip address 200.1.2.1/24
VTEP4(cfg-if-OverlayRouter 200)# exit
VTEP4(cfg)# vxlan 200
VTEP4(cfg-vxlan)# symmetric
VTEP4(cfg-vxlan)# router-interface OverlayRouter 200
VTEP4(cfg-vxlan)# exit
VTEP4(cfg)# router-protocol bgp 64512
VTEP4(cfg-router)# peer 1.1.1.1 remote-as 64512
VTEP4(cfg-router)# peer 1.1.1.1 update-source loopback 1
VTEP4(cfg-router)# peer 2.2.2.2 remote-as 64512
VTEP4(cfg-router)# peer 2.2.2.2 update-source loopback 1
VTEP4(cfg-router)# address-family-mode l2vpn evpn
VTEP4(cfg-router-af)# peer 1.1.1.1 activate
VTEP4(cfg-router-af)# peer 2.2.2.2 activate
VTEP4(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP4(cfg-router-af)# exit
VTEP4(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP4(cfg-router-af)# exit
VTEP4(cfg-router)# address-family-mode ipv4 vrf vrf-20
VTEP4(cfg-router-af)# exit
VTEP4(cfg-router)# address-family-mode ipv4 vrf vrf-30
VTEP4(cfg-router-af)# redistribute-route static
VTEP4(cfg-router-af)# exit
VTEP4(cfg-router)# exit
VTEP4(cfg)# evpn
VTEP4(cfg-evpn)# vni 100
VTEP4(cfg-evpn-vni)# rd auto
VTEP4(cfg-evpn-vni)# route-target both auto
VTEP4(cfg-evpn-vni)# route-target import 3000:3000
VTEP4(cfg-evpn-vni)# exit
VTEP4(cfg-evpn)# vni 200
VTEP4(cfg-evpn-vni)# rd auto
VTEP4(cfg-evpn-vni)# route-target both auto
VTEP4(cfg-evpn-vni)# route-target import 3000:3000
VTEP4(cfg-evpn-vni)# exit
```

5.7.5 Verification

- VM 1, VM 2, VM 4, and VM 5 can ping each other.
- VM 3 and VM 6 can ping each other.
- VMs belonging to different VRF instances cannot ping each other.
- VMs can be migrated in the same VXLAN, they can access the Internet service successfully after migration, and the configurations do not need to be modified.

- Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 3 as an example.

```
VTEP3# display vxlan
VXLAN Total Count: 3
VXLAN Capacity     : 8000

VXLAN 90
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 90 (anycast)
      Extend VLAN           : 90
      VTEP Adjacency Count: 0

VXLAN 100
      Symmetric property    : TRUE
      Routed interface      : overlayrouter 100 (non-anycast)
      Extend VLAN           : -
      VTEP Adjacency Count: 2
      VTEP Adjacency List :
      Interface                Source IP                                   Destination IP
Type
      ---------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145       7.7.7.7                                     5.5.5.5
dynamic
      OverlayTunnel 6146       7.7.7.7                                     6.6.6.6
dynamic

VXLAN 200
      Symmetric property    : TRUE
      Routed interface      : overlayrouter 200 (non-anycast)
      Extend VLAN           : -
      VTEP Adjacency Count: 2
      VTEP Adjacency List :
      Interface                Source IP                                   Destination IP
Type
      ---------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145       7.7.7.7                                     5.5.5.5
dynamic
      OverlayTunnel 6146       7.7.7.7                                     6.6.6.6
dynamic
```

5.7.6 Configuration Files

- VTEP 1 configuration file

```
!
ip vrf vrf-10
 rd 10:10
```

```
 route-target export 1000:1000
!
ip vrf vrf-20
 rd 20:20
 route-target export 2000:2000
!
vtep
 source loopback 2
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 1
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 2
 ip address 5.5.5.5 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 20.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 30
 ip vrf binding vrf-20
 ip address 30.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-10
 ip address 100.1.4.1 255.255.255.0
!
interface OverlayRouter 200
 ip vrf binding vrf-20
 ip address 200.1.4.1 255.255.255.0
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
```

```
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
 peer 3.3.3.3 remote-as 64512
 peer 3.3.3.3 update-source Loopback 1
 peer 4.4.4.4 remote-as 64512
 peer 4.4.4.4 update-source Loopback 1
 address-family-mode ipv4
  peer 2.2.2.2 activate
  peer 3.3.3.3 activate
  peer 4.4.4.4 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  advertise ipv4 unicast-af-modifier
  peer 2.2.2.2 activate
  peer 2.2.2.2 send-community extended
  peer 3.3.3.3 activate
  peer 3.3.3.3 send-community extended
  peer 4.4.4.4 activate
  peer 4.4.4.4 send-community extended
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-10
  redistribute-route connected
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-20
  redistribute-route connected
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
 vni 30
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
  route-target import 1000:1000
 vni 200
  rd auto
  route-target both auto
```

```
   route-target import 2000:2000
!
vxlan 10
 router-interface OverlayRouter 10
 extend-vlan 10
 arp suppress enable
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
 arp suppress enable
!
vxlan 30
 router-interface OverlayRouter 30
 extend-vlan 30
 arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 symmetric
!
vxlan 200
 router-interface OverlayRouter 200
 symmetric
!
```

- VTEP 2 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target export 1000:1000
!
ip vrf vrf-20
 rd 20:20
 route-target export 2000:2000
!
vtep
 source loopback 2
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 1
 ip address 2.2.2.2 255.255.255.255
!
interface Loopback 2
 ip address 6.6.6.6 255.255.255.255
```

```
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 20
 ip vrf binding vrf-10
 ip address 20.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 30
 ip vrf binding vrf-20
 ip address 30.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-10
 ip address 100.1.3.1 255.255.255.0
!
interface OverlayRouter 200
 ip vrf binding vrf-20
 ip address 200.1.3.1 255.255.255.0
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 peer 3.3.3.3 remote-as 64512
 peer 3.3.3.3 update-source Loopback 1
 peer 4.4.4.4 remote-as 64512
 peer 4.4.4.4 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.1 activate
  peer 3.3.3.3 activate
  peer 4.4.4.4 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  advertise ipv4 unicast-af-modifier
  peer 1.1.1.1 activate
  peer 1.1.1.1 send-community extended
  peer 3.3.3.3 activate
  peer 3.3.3.3 send-community extended
```

```
   peer 4.4.4.4 activate
   peer 4.4.4.4 send-community extended
   exit-address-family-mode
   !
  address-family-mode ipv4 vrf vrf-10
   redistribute-route connected
   exit-address-family-mode
   !
  address-family-mode ipv4 vrf vrf-20
   redistribute-route connected
   exit-address-family-mode
!
evpn
  vni 10
   rd auto
   route-target both auto
  vni 20
   rd auto
   route-target both auto
  vni 30
   rd auto
   route-target both auto
  vni 100
   rd auto
   route-target both auto
   route-target import 1000:1000
  vni 200
   rd auto
   route-target both auto
   route-target import 2000:2000
!
vxlan 10
  router-interface OverlayRouter 10
  extend-vlan 10
  arp suppress enable
!
vxlan 20
  router-interface OverlayRouter 20
  extend-vlan 20
  arp suppress enable
!
vxlan 30
  router-interface OverlayRouter 30
  extend-vlan 30
  arp suppress enable
!
```

```
vxlan 100
 router-interface OverlayRouter 100
 symmetric
!
vxlan 200
 router-interface OverlayRouter 200
 symmetric
!
```

- VTEP 3 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target export 1000:1000
 route-target import 3000:3000
!
ip vrf vrf-20
 rd 20:20
 route-target export 2000:2000
 route-target import 3000:3000
!
ip vrf vrf-30
 rd 30:30
 route-target export 3000:3000
 route-target import 1000:1000
 route-target import 2000:2000
!
vtep
 source loopback 2
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 1
 ip address 3.3.3.3 255.255.255.255
!
interface Loopback 2
 ip address 7.7.7.7 255.255.255.255
!
interface OverlayRouter 90
 ip vrf binding vrf-30
 ip address 90.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-10
 ip address 100.1.1.1 255.255.255.0
```

```
!
interface OverlayRouter 200
 ip vrf binding vrf-20
 ip address 200.1.1.1 255.255.255.0
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.1 activate
  peer 2.2.2.2 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  advertise ipv4 unicast-af-modifier
  peer 1.1.1.1 activate
  peer 1.1.1.1 send-community extended
  peer 2.2.2.2 activate
  peer 2.2.2.2 send-community extended
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-10
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-20
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-30
  redistribute-route static
  exit-address-family-mode
!
evpn
 vni 100
  rd auto
  route-target both auto
  route-target import 3000:3000
 vni 200
  rd auto
  route-target both auto
  route-target import 3000:3000
!
```

```
vxlan 90
 router-interface OverlayRouter 90
 extend-vlan 90
 arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 symmetric
!
vxlan 200
 router-interface OverlayRouter 200
 symmetric
!
```

- VTEP 4 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target export 1000:1000
 route-target import 3000:3000
!
ip vrf vrf-20
 rd 20:20
 route-target export 2000:2000
 route-target import 3000:3000
!
ip vrf vrf-30
 rd 30:30
 route-target export 3000:3000
 route-target import 1000:1000
 route-target import 2000:2000
!
vtep
 source loopback 2
 arp suppress enable
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 1
 ip address 4.4.4.4 255.255.255.255
!
interface Loopback 2
 ip address 8.8.8.8 255.255.255.255
!
interface OverlayRouter 90
 ip vrf binding vrf-30
 ip address 90.1.2.1 255.255.255.0
```

```
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-10
 ip address 100.1.2.1 255.255.255.0
!
interface OverlayRouter 200
 ip vrf binding vrf-20
 ip address 200.1.2.1 255.255.255.0
!
router-protocol bgp 64512
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 1.1.1.1 remote-as 64512
 peer 1.1.1.1 update-source Loopback 1
 peer 2.2.2.2 remote-as 64512
 peer 2.2.2.2 update-source Loopback 1
 address-family-mode ipv4
  peer 1.1.1.1 activate
  peer 2.2.2.2 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  advertise ipv4 unicast-af-modifier
  peer 1.1.1.1 activate
  peer 1.1.1.1 send-community extended
  peer 2.2.2.2 activate
  peer 2.2.2.2 send-community extended
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-10
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-20
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-30
  redistribute-route static
  exit-address-family-mode
!
evpn
 vni 100
  rd auto
  route-target both auto
  route-target import 3000:3000
```

```
 vni 200
  rd auto
  route-target both auto
  route-target import 3000:3000
!
vxlan 90
 router-interface OverlayRouter 90
 extend-vlan 90
 arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 symmetric
!
vxlan 200
 router-interface OverlayRouter 200
 symmetric
!
```

5.8 Configuring an EVPN Distributed Network to Support Non-EVPN VTEPs

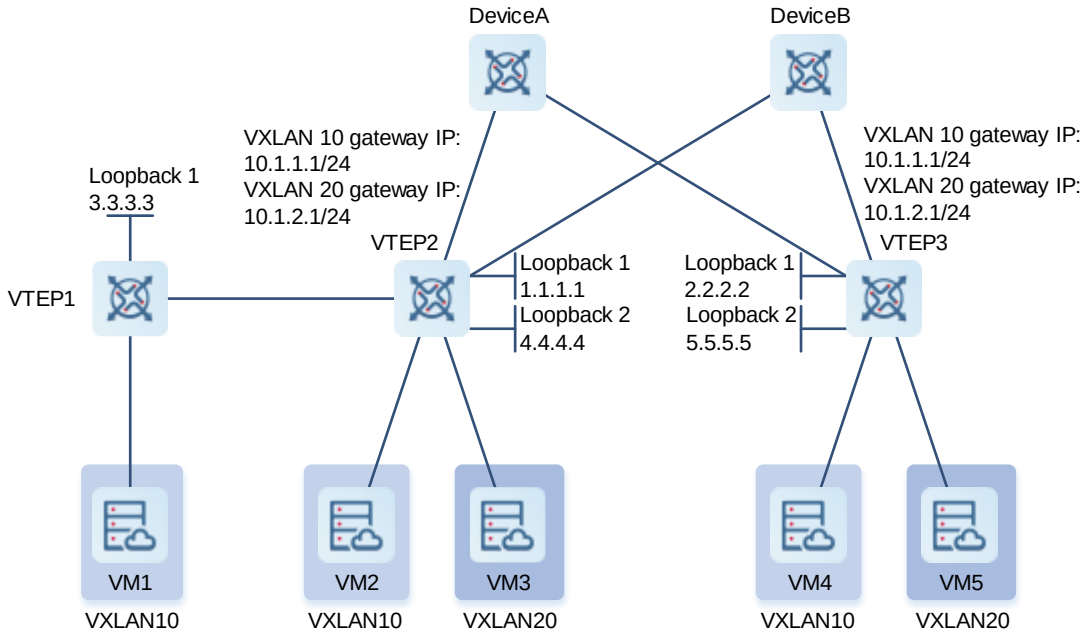
5.8.1 Requirements

In the topology as shown in Figure 5-8, VMs, access devices, and edge switches form a simple data center network. VTEP 1 does not support the EVPN function but VTEP 2 and VTEP 3 support the EVPN function. VTEP 2 and VTEP 3 form an EVPN distributed network. The following configurations are needed to connect VTEP 1 to the EVPN distributed network:

- On VTEP 2 and VTEP 3, configure distributed VXLAN networking using dynamic tunnels.
- On VTEP 2 and VTEP 3, configure the anycast gateway.
- On VTEP 2 and VTEP 3, configure the gateway to perform Layer 3 forwarding on traffic from the VXLAN network and non-VXLAN network.
- On VTEP 1 and VTEP 2, configure VXLAN networking using static tunnels and specify the tunnel.

5.8.2 Topology

Figure 5-8 Configuring an EVPN Distributed Network to Support Non-EVPN VTEPs



5.8.3 Notes

- On each device, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that the unicast routes are reachable (omitted).
- On VTEP 2, VTEP 3, device A, and device B, configure the BGP-EVPN routing protocol to ensure that the devices establish BGP neighbor relationships and exchange L2VPN EVPN information.
- On VTEP 2 and VTEP 3, configure an EVI for BGP-EVPN.
- On the virtualized servers, configure VXLANs and specify the gateway address for the VMs.
- On VTEP 2 and VTEP 3, configure and associate a loopback interface for the establishment of a tunnel.
- On VTEP 2 and VTEP 3, configure the anycast gateway MAC address.
- On each VTEP, create a VXLAN instance.
- On VTEP 2 and VTEP 3, create an overlay router interface for the VXLAN instance and configure the VXLAN gateway IP address. Configure the anycast attribute for the overlay router interface.
- On VTEP 2 and VTEP 3, associate the VXLAN instance with the overlay router interface to implement VXLAN routing.
- On VTEP 1 and VTEP 2, create a VXLAN tunnel and configure the source and destination IP addresses.
- On VTEP 1 and VTEP 2, associate the VXLAN instance with the VXLAN tunnel, and configure the tunnel as the data center interconnection tunnel.
- On VTEP 2, configure a route map and apply it to the EVPN protocol family so that the next hop of all EVPN type-2 routes advertised by VTEP 2 is changed to the IP address of VTEP 2.
- (Optional) Configure ARP suppression on VTEP 2 and VTEP 3 to reduce ARP packets flowing to the

VXLANs.

5.8.4 Procedure

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 3.3.3.3/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# interface overlaytunnel 1
VTEP1(cfg-if-OverlayTunnel 1)# overlaytunnel source 3.3.3.3
VTEP1(cfg-if-OverlayTunnel 1)# overlaytunnel destination 1.1.1.1
VTEP1(cfg-if-OverlayTunnel 1)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# tunnel-interface overlaytunnel 1
```

(2) Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP2(cfg)# route-map dc
VTEP2(cfg-routemap)# match-rule route-type evpn-type-2
VTEP2(cfg-routemap)# set-rule next-hop 4.4.4.4
VTEP2(cfg-routemap)# exit
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 1.1.1.1/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# interface loopback 2
VTEP2(cfg-if-Loopback 2)# ip address 4.4.4.4/32
VTEP2(cfg-if-Loopback 2)# exit
VTEP2(cfg)# fabric anycast-gateway-mac 0000.1234.5678
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 1
VTEP2(cfg-vtep)# arp suppress enable
VTEP2(cfg-vtep)# vxlan outside center vtep-ip 3.3.3.3
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# interface overlayrouter 10
VTEP2(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP2(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 10)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP2(cfg-if-OverlayRouter 10)# exit
VTEP2(cfg)# interface overlaytunnel 1
VTEP2(cfg-if-OverlayTunnel 1)# overlaytunnel source 1.1.1.1
VTEP2(cfg-if-OverlayTunnel 1)# overlaytunnel destination 3.3.3.3
VTEP2(cfg-if-OverlayTunnel 1)# evpn mac advertise enable
```

```
VTEP2(cfg-if-OverlayTunnel 1)# evpn mac-ip advertise enable
VTEP2(cfg-if-OverlayTunnel 1)# exit
VTEP2(cfg)# vxlan 10
VTEP2(cfg-vxlan)# extend-vlan 10
VTEP2(cfg-vxlan)# router-interface overlayrouter 10
VTEP2(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 20
VTEP2(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP2(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 20)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP2(cfg-if-OverlayRouter 20)# exit
VTEP2(cfg)# vxlan 20
VTEP2(cfg-vxlan)# extend-vlan 20
VTEP2(cfg-vxlan)# router-interface overlayrouter 20
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 100
VTEP2(cfg-router)# peer 5.5.5.5 remote-as 100
VTEP2(cfg-router)# peer 5.5.5.5 update-source loopback 2
VTEP2(cfg-router)# address-family-mode l2vpn evpn
VTEP2(cfg-router-af)# peer 5.5.5.5 activate
VTEP2(cfg-router-af)# peer 5.5.5.5 route-map dc out
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# exit
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 10
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 20
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
```

(3) Configure VTEP 3.

```
VTEP3# config-system terminal
VTEP3(cfg)# interface loopback 1
VTEP3(cfg-if-Loopback 1)# ip address 2.2.2.2/32
VTEP3(cfg-if-Loopback 1)# exit
VTEP3(cfg)# interface loopback 2
VTEP3(cfg-if-Loopback 2)# ip address 5.5.5.5/32
VTEP3(cfg-if-Loopback 2)# exit
VTEP3(cfg)# fabric anycast-gateway-mac 0000.1234.5678
VTEP3(cfg)# vtep
```

```
VTEP3(cfg-vtep)# source loopback 1
VTEP3(cfg-vtep)# arp suppress enable
VTEP3(cfg-vtep)# exit
VTEP3(cfg)# interface overlayrouter 10
VTEP3(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP3(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 10)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP3(cfg-if-OverlayRouter 10)# exit
VTEP3(cfg)# vxlan 10
VTEP3(cfg-vxlan)# extend-vlan 10
VTEP3(cfg-vxlan)# router-interface overlayrouter 10
VTEP3(cfg-vxlan)# arp suppress enable
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# interface overlayrouter 20
VTEP3(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP3(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 20)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP3(cfg-if-OverlayRouter 20)# exit
VTEP3(cfg)# vxlan 20
VTEP3(cfg-vxlan)# extend-vlan 20
VTEP3(cfg-vxlan)# router-interface overlayrouter 20
VTEP3(cfg-vxlan)# arp suppress enable
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# router-protocol bgp 100
VTEP3(cfg-router)# peer 4.4.4.4 remote-as 100
VTEP3(cfg-router)# peer 4.4.4.4 update-source loopback 2
VTEP3(cfg-router)# address-family-mode l2vpn evpn
VTEP3(cfg-router-af)# peer 4.4.4.4 activate
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# exit
VTEP3(cfg)# evpn
VTEP3(cfg-evpn)# vni 10
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# exit
VTEP3(cfg-evpn)# vni 20
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# exit
```

5.8.5 Verification

- VMs can ping each other.
- VMs can be migrated in the same VXLAN, they can access the Internet service successfully after migration, and the configurations do not need to be modified.

- Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 2 as an example.

```
VTEP2# display vxlan
VXLAN Total Count: 2
VXLAN Capacity     : 8000

VXLAN 10
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 10 (anycast)
      Extend VLAN           : 10
      VTEP Adjacency Count: 2
      VTEP Adjacency List :
      Interface                Source IP                                    Destination IP
Type
      ---------------------- ---------------------------------------------- ---------------------------------------------- -------
      OverlayTunnel 1          1.1.1.1                                      3.3.3.3
static
      OverlayTunnel 6145       1.1.1.1                                      2.2.2.2
dynamic

VXLAN 20
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 20 (anycast)
      Extend VLAN           : 20
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                Source IP                                    Destination IP
Type
      ---------------------- ---------------------------------------------- ---------------------------------------------- -------
      OverlayTunnel 6145       1.1.1.1                                      2.2.2.2
dynamic
```

5.8.6 Configuration Files

- VTEP 1 configuration file

```
!
interface Loopback 1
 ip address 3.3.3.3 255.255.255.255
!
interface OverlayTunnel 1
 overlaytunnel source 3.3.3.3
 overlaytunnel destination 1.1.1.1
!
vxlan 10
 tunnel-interface OverlayTunnel 1
 extend-vlan 10
```

```
!
```

- VTEP 2 configuration file

```
!
route-map dc permit 10
 match-rule route-type evpn-type-2
 set-rule next-hop 4.4.4.4
!
vtep
 source loopback 1
 arp suppress enable
 vxlan outside center vtep-ip 3.3.3.3
!
fabric anycast-gateway-mac 0000.1234.5678
!
interface Loopback 1
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 2
 ip address 4.4.4.4 255.255.255.255
!
interface OverlayTunnel 1
 overlaytunnel source 1.1.1.1
 overlaytunnel destination 3.3.3.3
 evpn mac advertise enable
 evpn mac-ip advertise enable
!
interface OverlayRouter 10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
interface OverlayRouter 20
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 5.5.5.5 remote-as 100
 peer 5.5.5.5 update-source Loopback 2
 address-family-mode ipv4
  peer 5.5.5.5 activate
  exit-address-family-mode
```

```
 address-family-mode l2vpn evpn
  peer 5.5.5.5 activate
  peer 5.5.5.5 send-community extended
  peer 5.5.5.5 route-map dc out
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 tunnel-interface OverlayTunnel 1
 extend-vlan 10
 arp suppress enable
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
 arp suppress enable
!
```

- VTEP 3 configuration file

```
!
vtep
 source loopback 1
 arp suppress enable
!
fabric anycast-gateway-mac 0000.1234.5678
!
interface Loopback 1
 ip address 2.2.2.2 255.255.255.255
!
interface Loopback 2
 ip address 5.5.5.5 255.255.255.255
!
interface OverlayRouter 10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
interface OverlayRouter 20
 ip address 10.1.2.1 255.255.255.0
```

```
 anycast-gateway
 route-in-vni
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 4.4.4.4 remote-as 100
 peer 4.4.4.4 update-source Loopback 2
 address-family-mode ipv4
  peer 4.4.4.4 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 4.4.4.4 activate
  peer 4.4.4.4 send-community extended
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 extend-vlan 10
 arp suppress enable
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
 arp suppress enable
!
```

5.9 Configuring Non-EVPN VTEPs to Connect to an EVPN Distributed Network in Full-Mesh Pattern

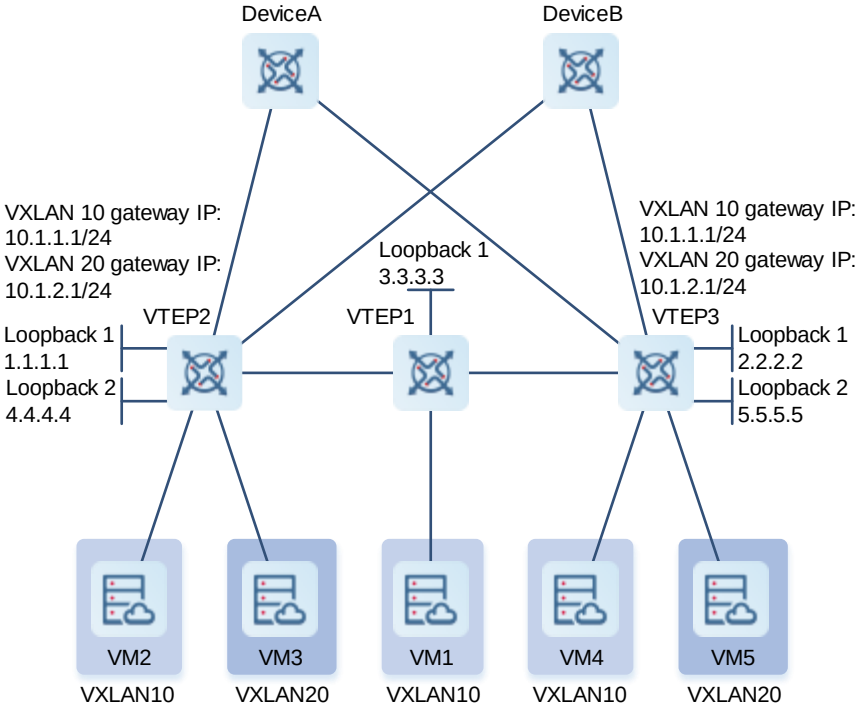
5.9.1 Requirements

In the topology as shown in Figure 5-9, VMs, access devices, and edge switches form a simple data center network. VTEP 1 does not support the EVPN function but VTEP 2 and VTEP 3 support the EVPN function. VTEP 2 and VTEP 3 form an EVPN distributed network. The following configurations are needed to connect VTEP 1 to the EVPN distributed network in full-mesh pattern:

- On VTEP 2 and VTEP 3, configure distributed VXLAN networking using dynamic tunnels.
- On VTEP 2 and VTEP 3, configure the anycast gateway.
- On VTEP 2 and VTEP 3, configure the gateway to perform Layer 3 forwarding on traffic from the VXLAN network and non-VXLAN network.
- On VTEP 1 and VTEP 2, configure VXLAN networking using static tunnels and specify the tunnel.

5.9.2 Topology

Figure 5-9 Configuring Non-EVPN VTEPs to Connect to an EVPN Distributed Network in Full-Mesh Pattern



5.9.3 Notes

- On each device, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that the unicast routes are reachable (omitted).
- On VTEP 2, VTEP 3, device A, and device B, configure the BGP-EVPN routing protocol to ensure that the devices establish BGP neighbor relationships and exchange L2VPN EVPN information.
- On VTEP 2 and VTEP 3, configure an EVI for BGP-EVPN.
- On the virtualized servers, configure VXLANs and specify the gateway address for the VMs.
- On VTEP 2 and VTEP 3, configure and associate a loopback interface for the establishment of a tunnel.
- On VTEP 2 and VTEP 3, configure the anycast gateway MAC address.
- On each VTEP, create a VXLAN instance.
- On VTEP 2 and VTEP 3, create an overlay router interface for the VXLAN instance and configure the VXLAN gateway IP address. Configure the anycast attribute for the overlay router interface.

- On VTEP 2 and VTEP 3, associate the VXLAN instance with the overlay router interface to implement VXLAN routing.
- On VTEP 1, VTEP 2, and VTEP 3, create a VXLAN tunnel and configure the source and destination IP addresses for the tunnel.
- On VTEP 1, VTEP 2, and VTEP 3, configure a VXLAN instance to associate with the VXLAN tunnel.
- (Optional) Configure ARP suppression on VTEP 2 and VTEP 3 to reduce ARP packets flowing to the VXLANs.

5.9.4 Procedure

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 3.3.3.3/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# interface overlaytunnel 1
VTEP1(cfg-if-OverlayTunnel 1)# overlaytunnel source 3.3.3.3
VTEP1(cfg-if-OverlayTunnel 1)# overlaytunnel destination 1.1.1.1
VTEP1(cfg-if-OverlayTunnel 1)# exit
VTEP1(cfg)# interface overlaytunnel 2
VTEP1(cfg-if-OverlayTunnel 2)# overlaytunnel source 3.3.3.3
VTEP1(cfg-if-OverlayTunnel 2)# overlaytunnel destination 2.2.2.2
VTEP1(cfg-if-OverlayTunnel 2)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP1(cfg-vxlan)# tunnel-interface overlaytunnel 2
```

(2) Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 1.1.1.1/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# interface loopback 2
VTEP2(cfg-if-Loopback 2)# ip address 4.4.4.4/32
VTEP2(cfg-if-Loopback 2)# exit
VTEP2(cfg)# fabric anycast-gateway-mac 0000.1234.5678
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 1
VTEP2(cfg-vtep)# arp suppress enable
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# interface overlayrouter 10
VTEP2(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP2(cfg-if-OverlayRouter 10)# anycast-gateway
```

```
VTEP2(cfg-if-OverlayRouter 10)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP2(cfg-if-OverlayRouter 10)# exit
VTEP2(cfg)# interface overlaytunnel 1
VTEP2(cfg-if-OverlayTunnel 1)# overlaytunnel source 1.1.1.1
VTEP2(cfg-if-OverlayTunnel 1)# overlaytunnel destination 3.3.3.3
VTEP2(cfg-if-OverlayTunnel 1)# exit
VTEP2(cfg)# vxlan 10
VTEP2(cfg-vxlan)# extend-vlan 10
VTEP2(cfg-vxlan)# router-interface overlayrouter 10
VTEP2(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 20
VTEP2(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP2(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 20)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP2(cfg-if-OverlayRouter 20)# exit
VTEP2(cfg)# vxlan 20
VTEP2(cfg-vxlan)# extend-vlan 20
VTEP2(cfg-vxlan)# router-interface OverlayRouter 20
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 100
VTEP2(cfg-router)# peer 5.5.5.5 remote-as 100
VTEP2(cfg-router)# peer 5.5.5.5 update-source loopback 2
VTEP2(cfg-router)# address-family-mode l2vpn evpn
VTEP2(cfg-router-af)# peer 5.5.5.5 activate
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# exit
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 10
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 20
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
```

(3) Configure VTEP 3.

```
VTEP3> privilege
VTEP3# config-system terminal
VTEP3(cfg)# interface loopback 1
VTEP3(cfg-if-Loopback 1)# ip address 2.2.2.2/32
VTEP3(cfg-if-Loopback 1)# exit
```

```
VTEP3(cfg)# interface loopback 2
VTEP3(cfg-if-Loopback 2)# ip address 5.5.5.5/32
VTEP3(cfg-if-Loopback 2)# exit
VTEP3(cfg)# fabric anycast-gateway-mac 0000.1234.5678
VTEP3(cfg)# vtep
VTEP3(cfg-vtep)# source loopback 1
VTEP3(cfg-vtep)# arp suppress enable
VTEP3(cfg-vtep)# exit
VTEP3(cfg)# interface overlayrouter 10
VTEP3(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP3(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 10)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP3(cfg-if-OverlayRouter 10)# exit
VTEP3(cfg)# interface overlaytunnel 1
VTEP3(cfg-if-OverlayTunnel 1)# overlaytunnel source 2.2.2.2
VTEP3(cfg-if-OverlayTunnel 1)# overlaytunnel destination 3.3.3.3
VTEP3(cfg-if-OverlayTunnel 1)# exit
VTEP3(cfg)# vxlan 10
VTEP3(cfg-vxlan)# extend-vlan 10
VTEP3(cfg-vxlan)# router-interface overlayrouter 10
VTEP3(cfg-vxlan)# tunnel-interface overlaytunnel 1
VTEP3(cfg-vxlan)# arp suppress enable
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# interface overlayrouter 20
VTEP3(cfg-if-OverlayRouter 20)# ip address 10.1.2.1/24
VTEP3(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP3(cfg-if-OverlayRouter 20)# route-in-vni    //Optional. The arp suppress enable command needs to be used together.
VTEP3(cfg-if-OverlayRouter 20)# exit
VTEP3(cfg)# vxlan 20
VTEP3(cfg-vxlan)# extend-vlan 20
VTEP3(cfg-vxlan)# router-interface overlayrouter 20
VTEP3(cfg-vxlan)# arp suppress enable
VTEP3(cfg-vxlan)# exit
VTEP3(cfg)# router-protocol bgp 100
VTEP3(cfg-router)# peer 4.4.4.4 remote-as 100
VTEP3(cfg-router)# peer 4.4.4.4 update-source loopback 2
VTEP3(cfg-router)# address-family-mode l2vpn evpn
VTEP3(cfg-router-af)# peer 4.4.4.4 activate
VTEP3(cfg-router-af)# exit
VTEP3(cfg-router)# exit
VTEP3(cfg)# evpn
VTEP3(cfg-evpn)# vni 10
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
```

```
VTEP3(cfg-evpn-vni)# exit
VTEP3(cfg-evpn)# vni 20
VTEP3(cfg-evpn-vni)# rd auto
VTEP3(cfg-evpn-vni)# route-target both auto
VTEP3(cfg-evpn-vni)# exit
```

5.9.5 Verification

- VMs can ping each other.
- VMs can be migrated in the same VXLAN, they can access the Internet service successfully after migration, and the configurations do not need to be modified.
- Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 2 as an example.

```
VTEP2# display vxlan
VXLAN Total Count: 2
VXLAN Capacity     : 8000

VXLAN 10
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 10 (anycast)
      Extend VLAN           : 10
      VTEP Adjacency Count: 2
      VTEP Adjacency List :
Interface                   Source IP                                       Destination IP
Type
      ---------------------- ---------------------------------------------- ---------------------------------------------- -------
OverlayTunnel 1             1.1.1.1                                         3.3.3.3
static
OverlayTunnel 6145          1.1.1.1                                         2.2.2.2
dynamic

VXLAN 20
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 20 (anycast)
      Extend VLAN           : 20
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                   Source IP                                       Destination IP
Type
      ---------------------- ---------------------------------------------- ---------------------------------------------- -------
      OverlayTunnel 6145          1.1.1.1                                         2.2.2.2
dynamic
```

5.9.6 Configuration Files

- VTEP 1 configuration file

```
!
```

```
interface Loopback 1
 ip address 3.3.3.3 255.255.255.255
!
interface OverlayTunnel 1
 overlaytunnel source 3.3.3.3
 overlaytunnel destination 1.1.1.1
!
interface OverlayTunnel 2
 overlaytunnel source 3.3.3.3
 overlaytunnel destination 2.2.2.2
!
vxlan 10
 tunnel-interface OverlayTunnel 1
 tunnel-interface OverlayTunnel 2
 extend-vlan 10
!
```

- VTEP 2 configuration file

```
!
vtep
 source loopback 1
 arp suppress enable
!
fabric anycast-gateway-mac 0000.1234.5678
!
interface Loopback 1
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 2
 ip address 4.4.4.4 255.255.255.255
!
interface OverlayTunnel 1
 overlaytunnel source 1.1.1.1
 overlaytunnel destination 3.3.3.3
!
interface OverlayRouter 10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
interface OverlayRouter 20
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
router-protocol bgp 100
 bgp-instance log-peer-changes
```

```
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 5.5.5.5 remote-as 100
 peer 5.5.5.5 update-source Loopback 2
 address-family-mode ipv4
  peer 5.5.5.5 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 5.5.5.5 activate
  peer 5.5.5.5 send-community extended
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 tunnel-interface OverlayTunnel 1
 extend-vlan 10
 arp suppress enable
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
 arp suppress enable
!
```

- VTEP 3 configuration file

```
!
vtep
 source loopback 1
 arp suppress enable
!
fabric anycast-gateway-mac 0000.1234.5678
!
interface Loopback 1
 ip address 2.2.2.2 255.255.255.255
!
interface Loopback 2
 ip address 5.5.5.5 255.255.255.255
!
```

```
interface OverlayTunnel 1
 overlaytunnel source 2.2.2.2
 overlaytunnel destination 3.3.3.3
!
interface OverlayRouter 10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
interface OverlayRouter 20
 ip address 10.1.2.1 255.255.255.0
 anycast-gateway
 route-in-vni
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 4.4.4.4 remote-as 100
 peer 4.4.4.4 update-source Loopback 2
 address-family-mode ipv4
  peer 4.4.4.4 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 4.4.4.4 activate
  peer 4.4.4.4 send-community extended
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 tunnel-interface OverlayTunnel 1
 extend-vlan 10
 arp suppress enable
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
```

```
 arp suppress enable
!
```

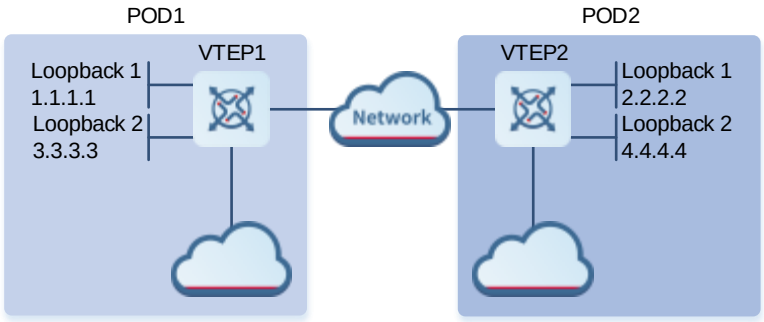
5.10 Configuring Data Center Interconnection Based on VNI Mappings

5.10.1 Requirements

In the topology as shown in Figure 5-10, PODs (data centers) connected through an IP network. VNI mappings need to be configured on Layer 3 routing devices in PODs so that an independent VNI is maintained inside each POD and a public VNI is maintained for the interaction of PODs.

5.10.2 Topology

Figure 5-10 Configuring Data Center Interconnection Based on VNI Mappings



5.10.3 Notes

- On each VTEP, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that the unicast routes are reachable (omitted).
- On each VTEP, configure the BGP-EVPN routing protocol to ensure that the devices establish BGP neighbor relationships and exchange L2VPN EVPN information.
- On each VTEP, configure an EVI for BGP-EVPN.
- On each VTEP, configure and associate a loopback interface for the establishment of a tunnel.
- On each VTEP, create a VXLAN instance.
- On each VTEP, create an overlay router interface for the VXLAN instance and configure the VXLAN gateway IP address.
- On each VTEP, associate the VXLAN instance with the overlay router interface to implement VXLAN routing.
- On VTEP 1 and VTEP 2, configure a data center interconnection tunnel.
- On each VTEP, configure a VNI mapping.
- On each VTEP, create a symmetric instance and configure the VXLAN gateway. No IP address needs to be configured for the gateway associated with the symmetric instance.
- On each VTEP, configure Layer 3 VNI mappings.

5.10.4 Procedure

For the network topology and device configuration of each POD, see 5.4 Configuring EVPN-based Multi-Tenant Distributed VXLAN Networking (Asymmetric Deployment). This section focuses on VXLAN configuration on core devices in PODs.

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 1.1.1.1/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# interface loopback 2
VTEP1(cfg-if-Loopback 2)# ip address 3.3.3.3/32
VTEP1(cfg-if-Loopback 2)# exit
VTEP1(cfg)# fabric anycast-gateway-mac 0011.2233.2016
VTEP1(cfg)# vtep
VTEP1(cfg-vtep)# source loopback 1
VTEP1(cfg-vtep)# vxlan outside center vtep-ip 2.2.2.2
VTEP1(cfg-vtep)# exit
VTEP1(cfg)# ip vrf vrf-10
VTEP1(cfg-vrf)# rd 10:10
VTEP1(cfg-vrf)# route-target export 1000:1000
VTEP1(cfg-vrf)# exit
VTEP1(cfg)# interface overlayrouter 10
VTEP1(cfg-if-OverlayRouter 10)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP1(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 10)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# router-interface overlayrouter 10
VTEP1(cfg-vxlan)# mapping vni 30
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 30
VTEP1(cfg-if-OverlayRouter 30)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 30)# ip address 30.1.1.1/24
VTEP1(cfg-if-OverlayRouter 30)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 30)# exit
VTEP1(cfg)# vxlan 30
VTEP1(cfg-vxlan)# router-interface overlayrouter 30
VTEP1(cfg-vxlan)# arp suppress enable
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 100
VTEP1(cfg-if-OverlayRouter 100)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 100)# exit
VTEP1(cfg)# vxlan 100
```

```
VTEP1(cfg-vxlan)# symmetric
VTEP1(cfg-vxlan)# router-interface overlayrouter 100
VTEP1(cfg-vxlan)# l3mapping vni 300
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 300
VTEP1(cfg-if-OverlayRouter 300)# ip vrf binding vrf-10
VTEP1(cfg-if-OverlayRouter 300)# exit
VTEP1(cfg)# vxlan 300
VTEP1(cfg-vxlan)# router-interface overlayrouter 300
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# router-protocol bgp 100
VTEP1(cfg-router)# peer 4.4.4.4 remote-as 100
VTEP1(cfg-router)# peer 4.4.4.4 update-source loopback 2
VTEP1(cfg-router)# address-family-mode l2vpn evpn
VTEP1(cfg-router-af)# peer 4.4.4.4 activate
VTEP1(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# address-family-mode ipv4 vrf vrf-10
VTEP1(cfg-router-af)# redistribute-route connected
VTEP1(cfg-router-af)# exit
VTEP1(cfg)# evpn
VTEP1(cfg-evpn)# vni 10
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 30
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 100
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 300
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
```

(2) Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 2.2.2.2/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# interface loopback 2
VTEP2(cfg-if-Loopback 2)# ip address 4.4.4.4/32
VTEP2(cfg-if-Loopback 2)# exit
```

```
VTEP2(cfg)# fabric anycast-gateway-mac 0000.1111.2222
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 1
VTEP2(cfg-vtep)# vxlan outside center vtep-ip 1.1.1.1
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# ip vrf vrf-20
VTEP2(cfg-vrf)# rd 20:20
VTEP2(cfg-vrf)# route-target export 2000:2000
VTEP2(cfg-vrf)# exit
VTEP2(cfg)# interface overlayrouter 20
VTEP2(cfg-if-OverlayRouter 20)# ip vrf binding vrf-20
VTEP2(cfg-if-OverlayRouter 20)# ip address 20.1.1.1/24
VTEP2(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 20)# exit
VTEP2(cfg)# vxlan 20
VTEP2(cfg-vxlan)# router-interface overlayrouter 20
VTEP2(cfg-vxlan)# mapping vni 30
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 30
VTEP2(cfg-if-OverlayRouter 30)# ip vrf binding vrf-20
VTEP2(cfg-if-OverlayRouter 30)# ip address 30.1.1.1/24
VTEP2(cfg-if-OverlayRouter 30)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 30)# exit
VTEP2(cfg)# vxlan 30
VTEP2(cfg-vxlan)# router-interface overlayrouter 30
VTEP2(cfg-vxlan)# arp suppress enable
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 200
VTEP2(cfg-if-OverlayRouter 200)# ip vrf binding vrf-20
VTEP2(cfg-if-OverlayRouter 200)# exit
VTEP2(cfg)# vxlan 200
VTEP2(cfg-vxlan)# symmetric
VTEP2(cfg-vxlan)# router-interface overlayrouter 200
VTEP2(cfg-vxlan)# l3mapping vni 300
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 300
VTEP2(cfg-if-OverlayRouter 300)# ip vrf binding vrf-20
VTEP2(cfg-if-OverlayRouter 300)# exit
VTEP2(cfg)# vxlan 300
VTEP2(cfg-vxlan)# router-interface overlayrouter 300
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 100
VTEP2(cfg-router)# peer 3.3.3.3 remote-as 100
VTEP2(cfg-router)# peer 3.3.3.3 update-source loopback 2
VTEP2(cfg-router)# address-family-mode l2vpn evpn
```

```
VTEP2(cfg-router-af)# peer 3.3.3.3 activate
VTEP2(cfg-router-af)# advertise ipv4 unicast-af-modifier
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# address-family-mode ipv4 vrf vrf-20
VTEP2(cfg-router-af)# redistribute-route connected
VTEP2(cfg-router-af)# exit
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 20
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 30
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 200
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 300
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
```

5.10.5 Verification

- Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 1 as an example.

```
VTEP1# display vxlan
VXLAN Total Count: 4
VXLAN Capacity     : 8000

VXLAN 10
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 10 (anycast)
      Extend VLAN           : -
      VTEP Adjacency Count: 2
      VTEP Adjacency List :
      Interface                 Source IP                                   Destination IP
Type
      ---------------------- ------------------------------------------- ------------------------------------------- -------
      OverlayTunnel 6145        1.1.1.1                                     5.5.5.5
dynamic
      OverlayTunnel 6146        1.1.1.1                                     6.6.6.6
dynamic

VXLAN 30
```

```
    Symmetric property   : FALSE
    Routed interface     : overlayrouter 30 (anycast)
    Extend VLAN          : -
    VTEP Adjacency Count: 1
    VTEP Adjacency List :
    Interface                Source IP                                   Destination IP
Type
    ---------------------- ------------------------------------------- ------------------------------------------- -------
    OverlayTunnel 6147       1.1.1.1                                     2.2.2.2
dynamic

VXLAN 100
    Symmetric property   : TRUE
    Routed interface     : overlayrouter 100 (non-anycast)
    Extend VLAN          : -
    VTEP Adjacency Count: 2
    VTEP Adjacency List :
    Interface                Source IP                                   Destination IP
Type
    ---------------------- ------------------------------------------- ------------------------------------------- -------
    OverlayTunnel 6145       1.1.1.1                                     5.5.5.5
dynamic
    OverlayTunnel 6146       1.1.1.1                                     6.6.6.6
dynamic

VXLAN 300
    Symmetric property   : FALSE
    Routed interface     : overlayrouter 300 (non-anycast)
    Extend VLAN          : -
    VTEP Adjacency Count: 1
    VTEP Adjacency List :
    Interface                Source IP                                   Destination IP
Type
    ---------------------- ------------------------------------------- ------------------------------------------- -------
    OverlayTunnel 6147       1.1.1.1                                     2.2.2.2
dynamic
```

5.10.6 Configuration Files

- VTEP 1 configuration file

```
!
ip vrf vrf-10
 rd 10:10
 route-target export 1000:1000
!
```

```
vtep
 source loopback 1
 vxlan outside center vtep-ip 2.2.2.2
!
fabric anycast-gateway-mac 0011.2233.2016
!
interface Loopback 1
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 2
 ip address 3.3.3.3 255.255.255.255
!
interface OverlayRouter 10
 ip vrf binding vrf-10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 30
 ip vrf binding vrf-10
 ip address 30.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 ip vrf binding vrf-10
!
interface OverlayRouter 300
 ip vrf binding vrf-10
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 4.4.4.4 remote-as 100
 peer 4.4.4.4 update-source Loopback 2
 address-family-mode ipv4
  peer 4.4.4.4 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  advertise ipv4 unicast-af-modifier
  peer 4.4.4.4 activate
  peer 4.4.4.4 send-community extended
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-10
  redistribute-route connected
```

```
   exit-address-family-mode
!
evpn
 vni 10
   rd auto
   route-target both auto
 vni 30
   rd auto
   route-target both auto
 vni 100
   rd auto
   route-target both auto
 vni 300
   rd auto
   route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 mapping vni 30
arp suppress enable
!
vxlan 30
 router-interface OverlayRouter 30
 arp suppress enable
!
vxlan 100
 router-interface OverlayRouter 100
 l3mapping vni 300
 symmetric
!
vxlan 300
 router-interface OverlayRouter 300
!
```

- VTEP 2 configuration file

```
!
ip vrf vrf-20
 rd 20:20
 route-target export 2000:2000
!
vtep
 source loopback 1
 vxlan outside center vtep-ip 1.1.1.1
!
fabric anycast-gateway-mac 0011.2233.2016
```

```
!
interface Loopback 1
 ip address 2.2.2.2 255.255.255.255
!
interface Loopback 2
 ip address 4.4.4.4 255.255.255.255
!
interface OverlayRouter 20
 ip vrf binding vrf-20
 ip address 20.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 30
 ip vrf binding vrf-20
 ip address 30.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 200
 ip vrf binding vrf-20
!
interface OverlayRouter 300
 ip vrf binding vrf-20
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 3.3.3.3 remote-as 100
 peer 3.3.3.3 update-source Loopback 2
 address-family-mode ipv4
  peer 3.3.3.3 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  advertise ipv4 unicast-af-modifier
  peer 3.3.3.3 activate
  peer 3.3.3.3 send-community extended
  exit-address-family-mode
  !
 address-family-mode ipv4 vrf vrf-20
  redistribute-route connected
  exit-address-family-mode
!
evpn
 vni 20
  rd auto
```

```
   route-target both auto
  vni 30
   rd auto
   route-target both auto
  vni 200
   rd auto
   route-target both auto
  vni 300
   rd auto
   route-target both auto
!
vxlan 20
 router-interface OverlayRouter 20
 mapping vni 30
arp suppress enable
!
vxlan 30
 router-interface OverlayRouter 30
 arp suppress enable
!
vxlan 200
 router-interface OverlayRouter 200
 l3mapping vni 300
 symmetric
!
vxlan 300
 router-interface OverlayRouter 300
!
```

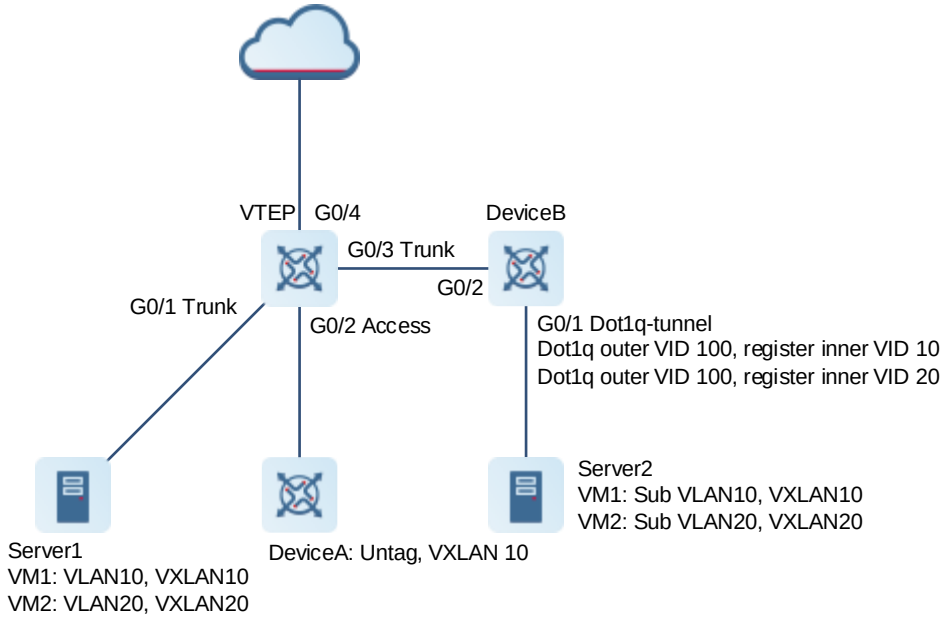
5.11 Configuring a Layer 2 Ethernet Sub-interface to Access a VXLAN

5.11.1 Requirements

In the topology as shown in Figure 5-11, different Layer 2 Ethernet sub-interface access modes can be configured to enable VMs to connect to a VXLAN.

5.11.2 Topology

Figure 5-11 Configuring a Layer 2 Ethernet Sub-interface to Access a VXLAN



5.11.3 Notes

- Configure virtual servers, VMs, and physical servers (omitted).
- On the VTEP, create a Layer 2 Ethernet sub-interface, configure a VXLAN, and the untagged, VLAN, or QinQ encapsulation rule.
- On the VTEP, create an overlay router interface and configure the VXLAN gateway IP address.
- On the VTEP, associate the VXLAN instance with the overlay router interface to implement VXLAN routing.

5.11.4 Procedure

This section focuses only on VTEP-related configuration.

Configure the VTEP.

```
VTEP> privilege
VTEP# config-system terminal
VTEP(cfg)# vlan 10
VTEP(cfg-vlan)# exit
VTEP(cfg)# vlan 20
VTEP(cfg-vlan)# exit
VTEP(cfg)# interface gigabitethernet 0/1
VTEP(cfg-if-GigabitEthernet 0/1)# switch-port mode trunk
VTEP(cfg-if-GigabitEthernet 0/1)# exit
VTEP(cfg)# interface gigabitethernet 0/1.1
VTEP(cfg-subif-GigabitEthernet 0/1.1)# subinf-traffic-encapsulation dot1q s-vid 10
VTEP(cfg-subif-GigabitEthernet 0/1.1)# subinf-traffic-encapsulation vxlan 10
VTEP(cfg-subif-GigabitEthernet 0/1.1)# exit
```

```
VTEP(cfg)#interface gigabitethernet 0/1.2
VTEP(cfg-subif-GigabitEthernet 0/1.2)# subinf-traffic-encapsulation dot1q s-vid 20
VTEP(cfg-subif-GigabitEthernet 0/1.2)# subinf-traffic-encapsulation vxlan 20
VTEP(cfg-subif-GigabitEthernet 0/1.2)# exit
VTEP(cfg)#interface gigabitethernet 0/2
VTEP(cfg-if-GigabitEthernet 0/2)# switch-port mode access
VTEP(cfg-if-GigabitEthernet 0/2)# switch-port access vlan 10
VTEP(cfg-if-GigabitEthernet 0/2)# exit
VTEP(cfg)#interface gigabitethernet 0/2.1
VTEP(cfg-subif-GigabitEthernet 0/2.1)# subinf-traffic-encapsulation dot1q untag
VTEP(cfg-subif-GigabitEthernet 0/2.1)# subinf-traffic-encapsulation vxlan 10
VTEP(cfg-subif-GigabitEthernet 0/2.1)# exit
VTEP(cfg)#interface gigabitethernet 0/3
VTEP(cfg-if-GigabitEthernet 0/3)# switch-port mode trunk
VTEP(cfg-if-GigabitEthernet 0/3)# exit
VTEP(cfg)#interface gigabitethernet 0/3.1
VTEP(cfg-subif-GigabitEthernet 0/3.1)# subinf-traffic-encapsulation qinq s-vid 100 c-vid 10
VTEP(cfg-subif-GigabitEthernet 0/3.1)# subinf-traffic-encapsulation vxlan 10
VTEP(cfg-subif-GigabitEthernet 0/3.1)# exit
VTEP(cfg)#interface gigabitethernet 0/3.2
VTEP(cfg-subif-GigabitEthernet 0/3.2)# subinf-traffic-encapsulation s-vid 100 c-vid 20
VTEP(cfg-subif-GigabitEthernet 0/3.2)# subinf-traffic-encapsulation vxlan 20
VTEP(cfg-subif-GigabitEthernet 0/3.2)# exit
VTEP(cfg)# interface overlayrouter 10
VTEP(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP(cfg-if-OverlayRouter 10)# exit
VTEP(cfg)# vxlan 10
VTEP(cfg-vxlan)# router-interface overlayrouter 10
VTEP(cfg-vxlan)# exit
VTEP(cfg)# interface overlayrouter 20
VTEP(cfg-if-OverlayRouter 20)# ip address 20.1.1.1/24
VTEP(cfg-if-OverlayRouter 20)# exit
VTEP(cfg)# vxlan 20
VTEP(cfg-vxlan)# router-interface overlayrouter 20
VTEP1(cfg-vxlan)# exit
```

5.11.5 Verification

Run the **display vxlan** command to display VXLAN instances on the devices.

```
VTEP# display vxlan
VXLAN Total Count: 2
VXLAN Capacity     : 8000

VXLAN 10
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 10 (non-anycast)
      Extend VLAN             : -
```

```
    VTEP Adjacency Count: 0

VXLAN 20
    Symmetric property   : FALSE
    Routed interface     : overlayrouter 20 (non-anycast)
    Extend VLAN          : -
    VTEP Adjacency Count: 0
```

5.11.6 Configuration Files

VTEP configuration file

```
!
interface GigabitEthernet 0/1
 switch-port mode trunk
 switch-port trunk allow-pass vlan only 1-9,11-19,21-4094
!
interface GigabitEthernet 0/1.1
 encapsulation dot1q s-vid 10
 encapsulation vxlan 10
!
interface GigabitEthernet 0/1.2
 encapsulation dot1q s-vid 20
 encapsulation vxlan 20
!
interface GigabitEthernet 0/2
 switch-port access vlan 10
!
interface GigabitEthernet 0/2.1
 encapsulation dot1q untag
 encapsulation vxlan 10
!
interface GigabitEthernet 0/3
 shutdown
 switch-port mode trunk
 switch-port trunk allowed vlan only 1-99,101-4094
!
interface GigabitEthernet 0/3.1
 encapsulation qinq s-vid 100 c-vid 10
 encapsulation vxlan 10
!
interface GigabitEthernet 0/3.2
 encapsulation qinq s-vid 100 c-vid 20
 encapsulation vxlan 20
!
interface OverlayRouter 10
 ip address 10.1.1.1 255.255.255.0
!
```

```
interface OverlayRouter 20
 ip address 20.1.1.1 255.255.255.0
!
vxlan 10
 router-interface OverlayRouter 10
!
vxlan 20
 router-interface OverlayRouter 20
!
```

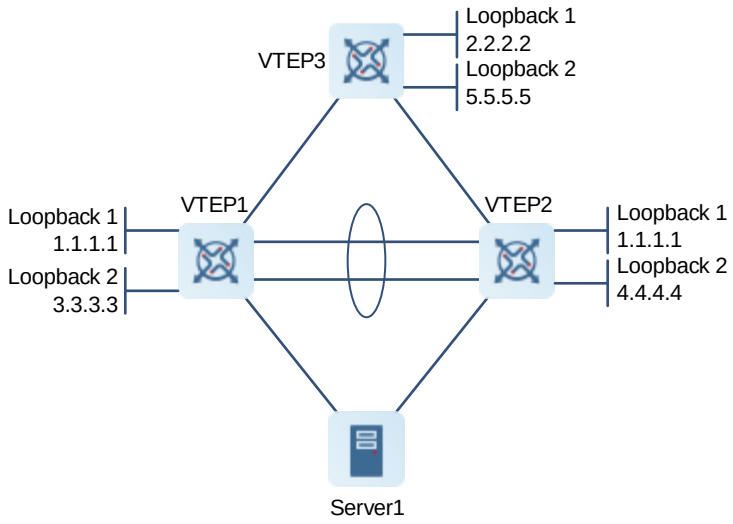
5.12 Configuring a VXLAN in an M-LAG Scenario

5.12.1 Requirements

In the topology as shown in Figure 5-12, all VTEPs form an EVPN-based VXLAN network. VTEP 1 and VTEP 2 compose an M-LAG system to enhance the network reliability. VMs connect to the VXLAN network through VTEP 1 and VTEP 2.

5.12.2 Topology

Figure 5-12 Configuring a VXLAN in an M-LAG Scenario



5.12.3 Notes

- Configure VTEP 1 and VTEP 2 to form an M-LAG system and configure the server to connect to the M-LAG system at Layer 2 through a VAP. For details about the M-LAG configuration, see *Configuring M-LAG*.
- On each VTEP, configure an IPv4 unicast routing protocol (such as OSPF) to ensure that the unicast routes are reachable (omitted).
- On each VTEP, configure the BGP-EVPN routing protocol to ensure that three VTEPs establish BGP neighbor relationships and exchange L2VPN EVPN information.

- On each VTEP, configure an EVI for BGP-EVPN.
- On the virtualized server, configure VMs and VXLAN and specify the VLAN, IP address, and gateway address for the VMs (omitted).
- On each VTEP, configure and associate a loopback interface for the establishment of a VXLAN tunnel.
- On each VTEP, create an overlay router interface and configure the VXLAN gateway IP address.
- On each VTEP, create a VXLAN instance and associate it with a VLAN.
- On each VTEP, create an overlay router interface and configure the VXLAN gateway IP address. Configure different VRF instances for the overlay router interfaces and identify tenants by VRF instance. Associate the VXLAN instance with the overlay router interface to implement VXLAN routing.
- On each VTEP, configure the same anycast gateway MAC address and configure the anycast attribute for all overlay router interfaces.

5.12.4 Procedure

This section focuses on VXLAN configuration on VTEP 1 and VTEP 2 that compose an M-LAG system.

(1) Configure VTEP 1.

```
VTEP1> privilege
VTEP1# config-system terminal
VTEP1(cfg)# interface loopback 1
VTEP1(cfg-if-Loopback 1)# ip address 1.1.1.1/32
VTEP1(cfg-if-Loopback 1)# exit
VTEP1(cfg)# interface loopback 2
VTEP1(cfg-if-Loopback 2)# ip address 3.3.3.3/32
VTEP1(cfg-if-Loopback 2)# exit
VTEP1(cfg)# fabric anycast-gateway-mac 0000.1122.3344
VTEP1(cfg)# vtep
VTEP1(cfg-vtep)# source loopback 1
VTEP1(cfg-vtep)# exit
VTEP1(cfg)# interface overlayrouter 10
VTEP1(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP1(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 10)# exit
VTEP1(cfg)# vxlan 10
VTEP1(cfg-vxlan)# extend-vlan 10
VTEP1(cfg-vxlan)# router-interface overlayrouter 10
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 20
VTEP1(cfg-if-OverlayRouter 20)# ip address 20.1.1.1/24
VTEP1(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 20)# exit
VTEP1(cfg)# vxlan 20
VTEP1(cfg-vxlan)# extend-vlan 20
VTEP1(cfg-vxlan)# router-interface overlayrouter 20
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# interface overlayrouter 100
```

```
VTEP1(cfg-if-OverlayRouter 100)# anycast-gateway
VTEP1(cfg-if-OverlayRouter 100)# exit
VTEP1(cfg)# vxlan 100
VTEP1(cfg-vxlan)# symmetric
VTEP1(cfg-vxlan)# router-interface overlayrouter 100
VTEP1(cfg-vxlan)# exit
VTEP1(cfg)# router-protocol bgp 100
VTEP1 (cfg-router)# peer 4.4.4.4 remote-as 100
VTEP1 (cfg-router)# peer 4.4.4.4 update-source loopback 2
VTEP1 (cfg-router)# peer 5.5.5.5 remote-as 100
VTEP1 (cfg-router)# peer 5.5.5.5 update-source loopback 2
VTEP1(cfg-router)# address-family-mode l2vpn evpn
VTEP1(cfg-router-af)# peer 4.4.4.4 activate
VTEP1(cfg-router-af)# peer 5.5.5.5 activate
VTEP1(cfg-router-af)# exit
VTEP1(cfg-router)# exit
VTEP1(cfg)# evpn
VTEP1(cfg-evpn)# vni 10
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 20
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
VTEP1(cfg-evpn)# vni 100
VTEP1(cfg-evpn-vni)# rd auto
VTEP1(cfg-evpn-vni)# route-target both auto
VTEP1(cfg-evpn-vni)# exit
```

(2) Configure VTEP 2.

```
VTEP2> privilege
VTEP2# config-system terminal
VTEP2(cfg)# interface loopback 1
VTEP2(cfg-if-Loopback 1)# ip address 1.1.1.1/32
VTEP2(cfg-if-Loopback 1)# exit
VTEP2(cfg)# interface loopback 2
VTEP2(cfg-if-Loopback 2)# ip address 4.4.4.4/32
VTEP2(cfg-if-Loopback 2)# exit
VTEP2(cfg)# fabric anycast-gateway-mac 0000.1122.3344
VTEP2(cfg)# vtep
VTEP2(cfg-vtep)# source loopback 1
VTEP2(cfg-vtep)# exit
VTEP2(cfg)# interface overlayrouter 10
VTEP2(cfg-if-OverlayRouter 10)# ip address 10.1.1.1/24
VTEP2(cfg-if-OverlayRouter 10)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 10)# exit
```

```
VTEP2(cfg)# vxlan 10
VTEP2(cfg-vxlan)# extend-vlan 10
VTEP2(cfg-vxlan)# router-interface overlayrouter 10
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 20
VTEP2(cfg-if-OverlayRouter 20)# ip address 20.1.1.1/24
VTEP2(cfg-if-OverlayRouter 20)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 20)# exit
VTEP2(cfg)# vxlan 20
VTEP2(cfg-vxlan)# extend-vlan 20
VTEP2(cfg-vxlan)# router-interface overlayrouter 20
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# interface overlayrouter 100
VTEP2(cfg-if-OverlayRouter 100)# anycast-gateway
VTEP2(cfg-if-OverlayRouter 100)# exit
VTEP2(cfg)# vxlan 100
VTEP2(cfg-vxlan)# symmetric
VTEP2(cfg-vxlan)# router-interface overlayrouter 100
VTEP2(cfg-vxlan)# exit
VTEP2(cfg)# router-protocol bgp 100
VTEP2 (cfg-router)# peer 3.3.3.3 remote-as 100
VTEP2 (cfg-router)# peer 3.3.3.3 update-source loopback 2
VTEP2 (cfg-router)# peer 5.5.5.5 remote-as 100
VTEP2 (cfg-router)# peer 5.5.5.5 update-source loopback 2
VTEP2(cfg-router)# address-family-mode l2vpn evpn
VTEP2(cfg-router-af)# peer 3.3.3.3 activate
VTEP2(cfg-router-af)# peer 5.5.5.5 activate
VTEP2(cfg-router-af)# exit
VTEP2(cfg-router)# exit
VTEP2(cfg)# evpn
VTEP2(cfg-evpn)# vni 10
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 20
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
VTEP2(cfg-evpn)# vni 100
VTEP2(cfg-evpn-vni)# rd auto
VTEP2(cfg-evpn-vni)# route-target both auto
VTEP2(cfg-evpn-vni)# exit
```

5.12.5 Verification

Run the **display vxlan** command to display VXLAN instances on the devices. The following uses VTEP 1 as an example.

```
VTEP1# display vxlan
VXLAN Total Count: 3
VXLAN Capacity     : 8000

VXLAN 10
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 10 (anycast)
      Extend VLAN           : 10
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                 Source IP                                       Destination IP
Type
      ----------------------- ----------------------------------------------- ----------------------------------------------- -------
      OverlayTunnel 6145        1.1.1.1                                         2.2.2.2
dynamic

VXLAN 20
      Symmetric property    : FALSE
      Routed interface      : overlayrouter 20 (anycast)
      Extend VLAN           : 20
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                 Source IP                                       Destination IP
Type
      ----------------------- ----------------------------------------------- ----------------------------------------------- -------
      OverlayTunnel 6145        1.1.1.1                                         2.2.2.2
dynamic

VXLAN 100
      Symmetric property    : TRUE
      Routed interface      : overlayrouter 100 (anycast)
      Extend VLAN           : -
      VTEP Adjacency Count: 1
      VTEP Adjacency List :
      Interface                 Source IP                                       Destination IP
Type
      ----------------------- ----------------------------------------------- ----------------------------------------------- -------
      OverlayTunnel 6145        1.1.1.1                                         2.2.2.2
dynamic
```

5.12.6 Configuration Files

- VTEP 1 configuration file

```
!
vtep
 source loopback 1
!
```

```
fabric anycast-gateway-mac 0000.1122.3344
!
interface Loopback 1
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 2
 ip address 3.3.3.3 255.255.255.255
!
interface OverlayRouter 10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 20
 ip address 20.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 anycast-gateway
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 4.4.4.4 remote-as 100
 peer 4.4.4.4 update-source Loopback 2
 peer 5.5.5.5 remote-as 100
 peer 5.5.5.5 update-source Loopback 2
 address-family-mode ipv4
  peer 4.4.4.4 activate
  peer 5.5.5.5 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 4.4.4.4 activate
  peer 4.4.4.4 send-community extended
  peer 5.5.5.5 activate
  peer 5.5.5.5 send-community extended
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
```

```
 vni 100
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 extend-vlan 10
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
!
vxlan 100
 router-interface OverlayRouter 100
 symmetric
!
```

- VTEP 2 configuration file

```
!
vtep
 source loopback 1
!
fabric anycast-gateway-mac 0000.1122.3344
!
interface Loopback 1
 ip address 1.1.1.1 255.255.255.255
!
interface Loopback 2
 ip address 4.4.4.4 255.255.255.255
!
interface OverlayRouter 10
 ip address 10.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 20
 ip address 20.1.1.1 255.255.255.0
 anycast-gateway
!
interface OverlayRouter 100
 anycast-gateway
!
router-protocol bgp 100
 bgp-instance log-peer-changes
 bgp-instance graceful-restart restart-time 120
 bgp-instance graceful-restart staleroute-time 360
 bgp-instance graceful-restart
 peer 3.3.3.3 remote-as 100
```

```
 peer 3.3.3.3 update-source Loopback 2
 peer 5.5.5.5 remote-as 100
 peer 5.5.5.5 update-source Loopback 2
 address-family-mode ipv4
  peer 3.3.3.3 activate
  peer 5.5.5.5 activate
  exit-address-family-mode
 address-family-mode l2vpn evpn
  peer 3.3.3.3 activate
  peer 3.3.3.3 send-community extended
  peer 5.5.5.5 activate
  peer 5.5.5.5 send-community extended
  exit-address-family-mode
!
evpn
 vni 10
  rd auto
  route-target both auto
 vni 20
  rd auto
  route-target both auto
 vni 100
  rd auto
  route-target both auto
!
vxlan 10
 router-interface OverlayRouter 10
 extend-vlan 10
!
vxlan 20
 router-interface OverlayRouter 20
 extend-vlan 20
!
vxlan 100
 router-interface OverlayRouter 100
 symmetric
!
```

5.12.7 Common Errors

- Two devices composing an M-LAG system serve as VTEPs but the VXLAN configurations on the two devices are different, resulting in traffic interruption during forwarding.
- Two devices composing an M-LAG system serve as VTEPs but no anycast attribute is configured for the overlay router interfaces on the devices, resulting in traffic interruption during forwarding.

Contents

1 Configuring PFC 1
1.1 Overview 1
1.2 Introduction 1
1.2.1 Basic Concepts 1
1.2.2 Packet Format 1
1.2.3 Principles 3
1.3 Restrictions and Guidelines 4
1.4 Configuration Task Summary 5
1.5 Configuring Basic Features 5
1.6 Configuring the PFC Deadlock Detection and Threshold Functions 6
1.7 Monitoring 7
1.8 Configuration Examples 8
1.8.1 Configuring PFC 8

1 Configuring PFC

1.1 Overview

In a converged data center, SAN traffic needs to be transmitted on the Ethernet network with no packet loss.

The Ethernet Pause mechanism can ensure no packet loss. For the Ethernet Pause mechanism, when the transmit rate of a sender is higher than the receive rate of a receiver, the receiver sends a Pause frame to the sender, requesting that the traffic be paused for a given period of time. The Ethernet Pause mechanism operates on the entire link. In other words, all traffic on an interface is paused. Actually, link sharing is critical. Link sharing has the following requirements:

- One type of burst traffic does not affect forwarding of other types of traffic.
- A large backlog of one type of traffic in a queue does not preempt cache resources of other types of traffic.

PFC is introduced to resolve conflicts between the Ethernet Pause mechanism and link sharing.

1.2 Introduction

When SAN traffic, Inter Process Communication (IPC) traffic, and LAN traffic coexist on a physical link, PFC ensures that no packet loss occurs for SAN traffic and that SAN traffic does not affect the other types of traffic. (One 802.1p priority level is assigned to SAN traffic, and PFC is enabled for this priority level. Different 802.1p priority levels are assigned to IPC traffic and LAN traffic, and PFC is disabled for these priority levels).

1.2.1 Basic Concepts

- 802.1p priority /Class of Service (CoS)

  The 802.1p standard specifies eight priority levels (from 0 to 7) in the Layer 2 Virtual Local Area Network (VLAN) Tag field.

16 bits	3 bits	1 bit	12 bits
Tag Protocol Identifier (TPID) (Value: 0x8100)	Priority	Canonical Format Indicator (CFI)	VLAN ID

1.2.2 Packet Format

A backpressure signal is an Ethernet frame. The PFC frame format is as follows.

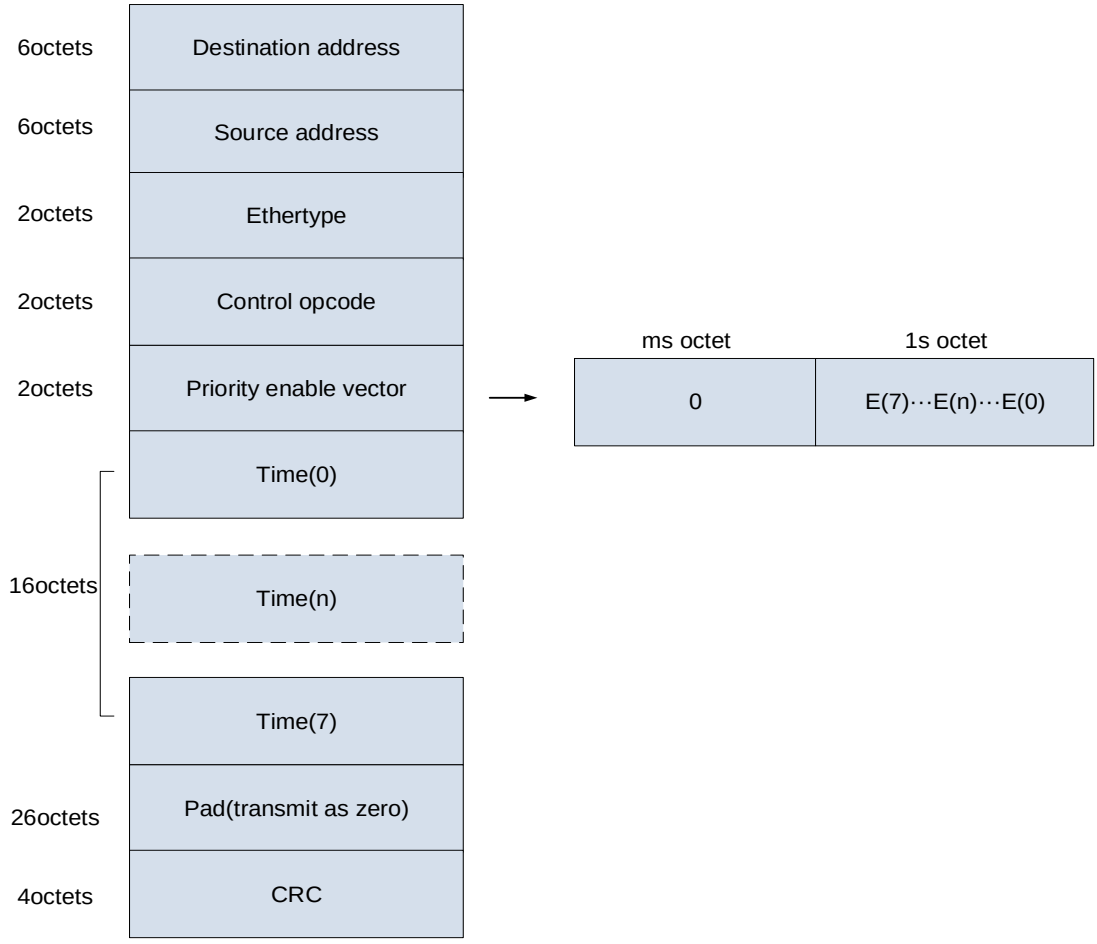


Table 1-1 Fields in a PFC Frame

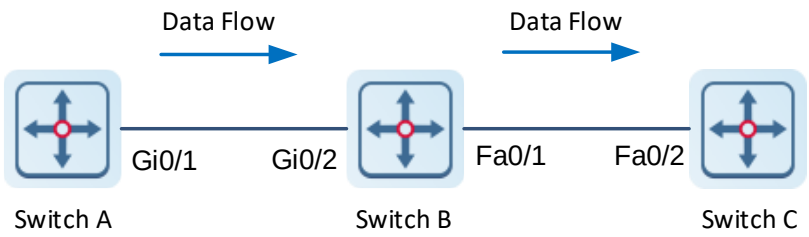
Field	Description
Destination address	Destination MAC address, which has a fixed value of 01-80-c2-00-00-01.
Source address	Source MAC address.
Ethertype	Ethernet frame type. The value is 88-08.
Control opcode	Control code. The value is 01-01.
Priority enable vector	E(*n*) corresponds to queue *n* and determines whether backpressure is enabled for queue *n*. If E(*n*) is 1, backpressure is enabled for queue *n*, and the backpressure time is Time(*n*). If E(*n*) is 0, backpressure is disabled for queue *n*.
Time(0) to Time(7)	The backpressure timer. If Time(*n*) is 0, backpressure is canceled.
Pad (transmit as zero)	Reserved. The value is 0 during PFC frame transmission.
CRC	Cyclic Redundancy Check (CRC).

1.2.3 Principles

1. Principles of PFC

PFC is an enhancement to flow control. Flow control is implemented as follows: When the receiver cannot process the received packets, it notifies the sender to temporarily stop sending packets. This prevents the packets from being dropped. Normal flow control is implemented by sending and receiving Pause frames on an interface.

Figure 1-1 Flow Control

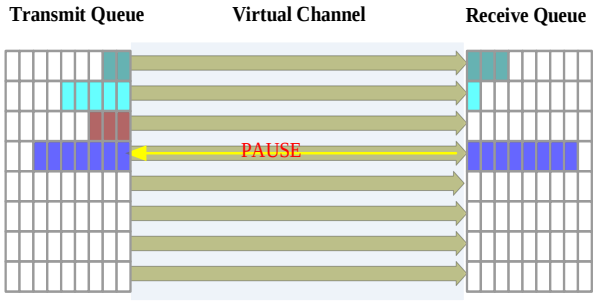


As shown in Figure 1-1, Fa0/1 is congested when Gi0/1 and Gi0/2 forward packets at 1 Gbps. To avoid packet loss, enable flow control for Gi0/1 and Gi0/2.

- When port Fa0/1 is congested during packet forwarding, Switch B handles the packets in the buffer and discards packets when the traffic exceeds the forwarding capability. In this case, Switch B sends a normal Pause frame through Gi0/2 to port Gi0/1 of Switch A, requesting Switch A to temporarily stop sending packets.
- Upon receiving the Pause frame, Switch A stops sending packets to Gi0/2 of Switch B in a specified period. The pause duration is carried by the Pause frame. When congestion persists, repeat the preceding steps until congestion is resolved.

Flow control can prevent packets from being dropped. However, all traffic is blocked on the entire link. With PFC, eight virtual channels can be created on an Ethernet link, with an 802.1p priority level specified for each virtual channel. Any of the virtual channels can be separately suspended or restarted without interrupting traffic on the other virtual channels.

Figure 1-2 PFC



In the preceding figure, PFC is enabled for packets with the priority of 3 for ports Gi0/1 and Gi0/2.

- When port Fa0/1 is congested during packet forwarding, Switch B handles the packets in the buffer and discards packets if the traffic exceeds the forwarding capability. To prevent packet loss, Switch B sends a PFC Pause frame through Gi0/2 to Gi0/1, requesting Switch A stop sending packets with the priority of 3.
- Upon receiving the PFC Pause frame, Switch A stops sending packets to Gi0/2 of Switch B in a specified

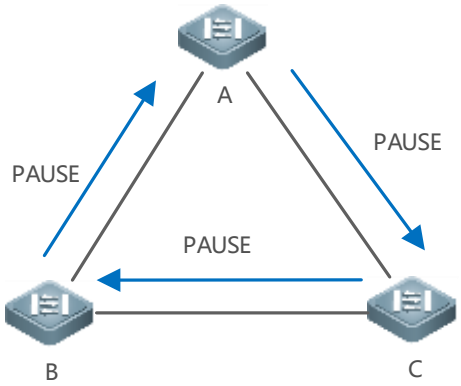
period. The pause duration is carried by the PFC Pause frame. When congestion persists, repeat the preceding steps until congestion is resolved. As shown in Figure 1-2, the sender stops sending packets with the priority level of 3, and other packets sent through ports Gi0/1 and Gi0/2 are not affected.

2. Principles of PFC Deadlock

After you configure the PFC function for switches, a PFC deadlock may occur when a link is congested.

The following figure shows a typical scenario where a cyclic buffer dependency exists among a group of devices in a data center. In a ring topology, any device affects the buffer of the upstream device while waiting for the downstream device to release part of the buffer to resume packet transmission. When the PFC-XOFF thresholds of A, B, and C switches are reached and the three switches send PFC Pause frames to the peer device simultaneously, all switches in the topology stop sending traffic. In this case, a PFC deadlock occurs. In addition, due to the backpressure effect of PFC Pause frames, the throughput of the entire or part of the network becomes zero. Deadlocks may easily occur on such networks. Deadlocks may also occur when a temporary loop is formed on a closed-loop network. In a data center, a temporary loop may occur when Border Gateway Protocol (BGP) route reselection is performed in the case of link failures. During upgrade of a data center based on Software-defined Networking (SDN), a temporary loop may also occur. Although temporary loops disappear quickly, the resulting deadlocks persist. Deadlocks cannot be eliminated automatically even after the error causing the issue (misconfiguration, link failure, or update) is fixed.

Figure 1-3 Figure 1-1 Principles of Deadlock



- Switch A, Switch B, and Switch C are on a network.
- To eliminate the deadlock, set a buffer period for the priority queues of an interface. This way, when the queue is congested and a PFC Pause frame is received, a few time slices are provided to process packets. This helps remove the deadlock.
- When deadlocks occur for a specific PFC priority frequently, disable the PFC function for this priority to prevent normal packet forwarding from being affected.

1.3 Restrictions and Guidelines

- You can adopt UTF-8/GBK to perform configuration in Chinese. When you use the cursor to process words or delete words by using the **Backspace** key in the CLI, the interface may display errors or garbled characters. You are advised to delete all current characters and then configure a unified character set

encoding format. Upon word processing or Backspace deletion, you can run the **display current-config** command to check whether the configurations are correct. (To delete a Chinese character, you must press the **Backspace** key twice in the case of GBK, and press the **Backspace** key three times in the case of UTF-8.)

1.4 Configuration Task Summary

Configuration in the CLI includes the following tasks: All the configuration tasks below are optional. Select the configuration tasks as required.

- Configuring Basic Features
- Configuring the PFC Deadlock Detection and Threshold Functions

1.5 Configuring Basic Features

1. Overview

You can control the traffic of an interface based on the 802.1p priority levels.

2. Restrictions and Guidelines

- PFC only applies to point-to-point full-duplex links, not to half-duplex or shared links.
- PFC and the normal flow control function are mutually exclusive.
- If a PFC-enabled priority level and a PFC-disabled priority level map to the same queue, the packets with a PFC-disabled priority level are blocked. Therefore, it is not recommended to map a PFC-enabled priority level and a PFC-disabled priority level to the same queue.
- PFC is not valid for packets sent by the CPU.
- PFC takes effect only if CoS is trusted on the interface.
- The devices that support globally enabled ETS can be used. PFC-enabled priority cannot be mapped to the same TC or queue as any other PFC-enabled priority; PFC feature cannot be enabled with any other priority mapping of the same TC or queue priority as any other priority.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Enable the PFC function.

priority-flow-control nodrop *cos-value-list* { **off** | **on** }

The PFC function is disabled by default.

(4) Disable PFC in both the send (Tx) and receive (Rx) directions.

priority-flow-control { rx | tx } disable

PFC is enabled in both the send (Tx) and receive (Rx) directions by default.

(5) Configure the pause quanta of the PFC Pause frame. A quanta is the time that is required for transmitting 512 bits at the speed of the chip.

priority-flow-control xoff-time *number*

The pause quanta is 65535 by default.

(6) Configuring the global PFC counting interval.

priority-flow-control counting-interval *number*

The global PFC counting interval is 5s by default.

(7) Configure the packet per second (PPS) rate limit for an interface (interface configuration mode).

priority-flow-control early-warning *cos-value-list* **{ output pps** *rate-num*] **| input pps** *rate-num* **}**

The PPS rate limit function is disabled by default.

1.6 Configuring the PFC Deadlock Detection and Threshold Functions

1. Overview

You can control the traffic of an interface based on the 802.1p priority levels.

2. Restrictions and Guidelines

- The PFC deadlock detection function is valid only if PFC is enabled for the priority level.
- The PFC deadlock threshold function takes effect only if the PFC deadlock detection function is enabled for the priority level.
- If you modify the time precision of the PFC deadlock functions, the configured PFC deadlock detection period and the packet processing time are affected.
- When PFC deadlock detection is disabled for a priority level, the configuration of the PFC deadlock threshold function is also removed for this priority level.
- If you disable the globally configured PFC deadlock detection function or the PFC function of a priority level, the command used to enable the PFC deadlock detection function for this priority level is removed from all interfaces.

3. Procedure

(1) Enter the privileged EXEC mode.

privilege

(2) Enter the global configuration mode.

config-system terminal

(3) Configure the PFC deadlock detection function for the priority levels.

priority-flow-control deadlock cos-value *cos-value-list* [**detect** *detect-time* **recover** *recover-time* [**continue** | **backoff** *most-times*]]

The deadlock detection period is 15 time slices and the deadlock recovery time is 150 ms by default.

(4) Configure the deadlock detection function for the priority levels of an interface (interface configuration mode).

priority-flow-control deadlock cos-value *cos-value-list* **enable**

The deadlock detection function is disabled for all priority levels of an interface by default.

(5) Configure the action to drop packets when the PFC deadlock occurs.

priority-flow-control deadlock drop

Packets are forwarded by the interface by default.

(6) Configure a low time precision for the PFC deadlock detection.

priority-flow-control deadlock precision low

The time precision for the deadlock detection is 10ms by default, which is of a high precision.

(7) Configure the PFC deadlock threshold function for the priority levels.

priority-flow-control deadlock limit cos-value *cos-value-list* **enable** [**rx**]

The deadlock threshold function is disabled for all priority levels by default.

(8) Configure the PFC deadlock threshold and the deadlock detection duration.

priority-flow-control deadlock limit frequency *number* { **period | period-second** } *time*

The deadlock detection duration is 60 seconds, and the deadlock threshold is 10 by default.

1.7 Monitoring

Run the **display** command to display the configurations.

Run the **cleanup** command to clear information.

Run the **debugging** commands to output debugging information.

Table 1-2 Monitoring of the CLI

Command	Purpose
cleanup priority-flow-control statistics [**interface** *type interface-number*]	Clears PFC statistics.
cleanup priority-flow-control deadlock statistics [interface *type interface-number* **]**	Clears the deadlock statistics of a specified interface.
display priority-flow-control statistics [**interface** *type interface-number*]	Displays PFC statistics.
display priority-flow-control status	Displays the global configuration of PFC.
display priority-flow-control status [**interface** *type interface-number*]	Displays the PFC configuration and status of a specified interface.
display priority-flow-control deadlock statistics [**interface** *type interface-number*]	Displays PFC deadlock statistics of priority levels of an interface.

Command	Purpose
display priority-flow-control deadlock status	Displays the configuration of the PFC deadlock detection and threshold functions and the enabling status.
debugging pfc cap	Debugs the PFC capability of a device.
debugging pfc deadlock	Debugs the PFC deadlock detection and threshold functions.
debugging pfc detail	Debugs the PFC details.
debugging pfc error	Debugs the PFC errors.
debugging pfc event	Debugs the PFC events.
debugging pfc mom	Debugs the message-oriented middleware of PFC.
debugging pfc statistics	Debugs the PFC statistics.

1.8 Configuration Examples

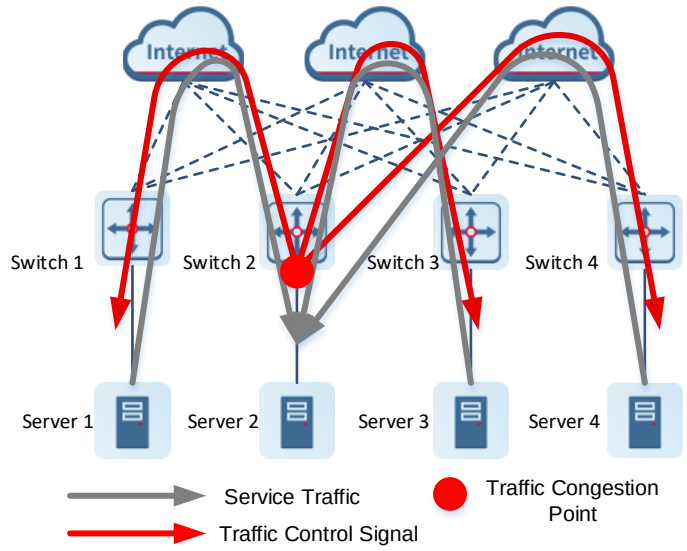
1.8.1 Configuring PFC

1. Requirements

A typical data center topology is shown in Figure 1-4. A distributed storage service is deployed on the servers. When the application on Server 2 needs to read data from Server 1, Server 3, and Server 4 at the same time, the traffic converges on Switch 2. At this point, the cache space in the egress direction of Switch 2 toward Server 2 becomes the bottleneck. To ensure that packets are not dropped due to buffer overflows, the PFC function is deployed on Switch 2. This way, once the cache space usage of the traffic reaches the PFC Pause frame threshold, Switch 2 sends the PFC Pause frame to the upstream switch, requesting the sender to stop transmitting packets of the specified priority level temporarily. This reduces congestion and ensures no packet loss.

2. Topology

Figure 1-4 Topology of a PFC-configured Data Center



3. Note

- Configure PFC priority levels on Switch 1/2/3/4 to ensure that PFC Pause frames are sent and received during traffic congestion.
- (Optional) Disable the PFC Pause frame in the egress or ingress direction of Switch 1/2/3/4.
- (Optional) Configure the PFC pause quanta on Switch 1/2/3/4.
- (Optional) Configure the PFC traffic warning threshold on Switch 1/2/3/4.
- (Optional) Configure the interval of PFC traffic statistics on Switch 1/2/3/4.
- (Optional) Configure the PFC deadlock detection and threshold functions on Switch 1/2/3/4.

4. Procedure

Configure the PFC function on Switch 2.

```
Switch2> privilege
Switch2# config-system terminal
Switch2(cfg)# priority-flow-control nodrop 3 on
Switch2(cfg)# priority-flow-control counting-interval 10
```

Configure the PFC function on Switch 1, Switch 3, and Switch 4.

```
Switch1> privilege
Switch1# config-system terminal
Switch1(cfg)# priority-flow-control nodrop 3 on
```

5. Verification

Display the PFC configuration of Switch 2.

```
Switch2# display priority-flow-control status
```

```
PFC Global PriorityEnablePFC: 3
Switch2# display priority-flow-control status interface FastEthernet 0/1
Interface                 PriorityEnabledPFC PriorityEnalbedPFCByUser
----------------      -----------------   --------------------------
FastEthernet 0/1      3
Switch2# display priority-flow-control statistics interface FastEthernet 0/1
interface   Priority PauseSend     PauseSend-Rate     PauseReceived PauseReceived-Rate
                                        (packets/sec)                  (packets/sec)
---------- -------- ----------- ---------------- ------------- ------------------
Fa 0/1        0          0              0                0              0
Fa 0/1        1          0              0                0              0
Fa 0/1        2          0              0                0              0
Fa 0/1        3          0              0                0              0
Fa 0/1        4          0              0                0              0
Fa 0/1        5          0              0                0              0
Fa 0/1        6          0              0                0              0
Fa 0/1        7          0              0                0              0
```

6. Configuration Files

Display the configuration file of Switch 2.

```
Building configuration...
!
version S6500_RGOS 12.5(4)B1004, Release(09221309)
hostname 6510
priority-flow-control counting-interval 10


!
end
```

Display the configuration files of Switch 1, Switch 3, and Switch 4.

```
Building configuration...
!
version S6500_RGOS 12.5(4)B1004, Release(09221309)
hostname 6510
priority-flow-control counting-interval 10


!
end
```