

Introduction to Network Devices

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- As the current network develops rapidly, people can easily learn the latest information in the world and communicate with friends and colleagues at any time. This greatly enriches people's communication and life. The future will hold a fully connected and intelligent world.
- Network construction requires the support of network devices, physical connections, and multiple protocols between devices. Routers and switches are the most commonly used network devices for network interconnection.
- There are various types of low-end, mid-range, and high-end routers and switches on networks of different levels. How do these devices work and process data packets?
- This course uses routers and switches as examples to describe the working mechanism and packet forwarding process of network devices.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe the logical structure of network devices.
  - Describe each functional module of network devices.
  - Describe the forwarding process of network devices.

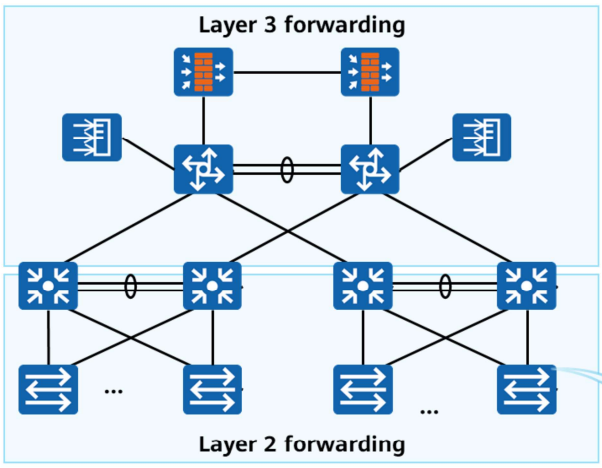
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Overview of Network Device Framework**
2. Packet Processing by Network Devices

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Network Devices

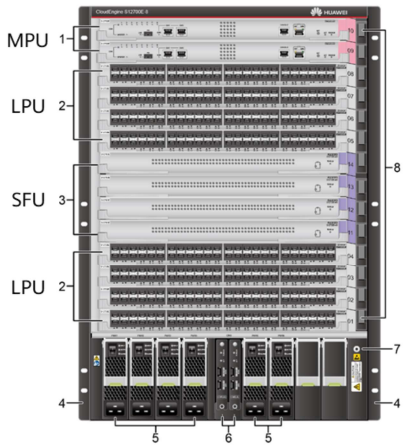


- The network infrastructure consists of switches, routers, and firewalls. These devices receive and send data. How does a device forward data received from an interface to another interface?
- What are components of a network device? How do these components work collaboratively?

How does a switch forward packets from one interface to another interface?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Hardware Modules of Modular Devices



Typical panel of a network device

The following uses the S12700E-8 as an example to describe the architecture of a typical network device.

- Main Processing Unit (MPU): is responsible for control plane and management plane of the entire system.
- Switch Fabric Unit (SFU): is responsible for the data plane of the entire system. The data plane provides high-speed non-blocking data channels for data switching between service modules.
- Line Processing Unit (LPU): provides data forwarding and optical and electrical interfaces with different rates.
- SFUs and LPUs have their own management chips, which work with MPUs to form the control and management plane of the entire device.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MPU



MPU

The MPU provides the control plane and management plane for the entire system.

- The control plane provides functions such as protocol processing, service processing, route calculation, forwarding control, service scheduling, traffic statistics collection, and system security.
- The management plane is responsible for system status monitoring, environment monitoring, log and alarm processing, system software loading, and system upgrade.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SFU



SFU

The SFU provides the data plane for the entire system. LPUs and MPUs communicate with each other through SFUs.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

LPU

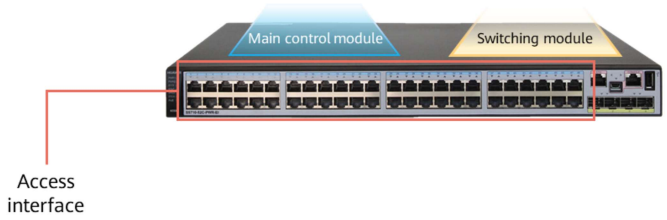


LPU

The LPU provides access interfaces of different types (optical and electrical interfaces) and different rates, and forwards data through the distributed data plane.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

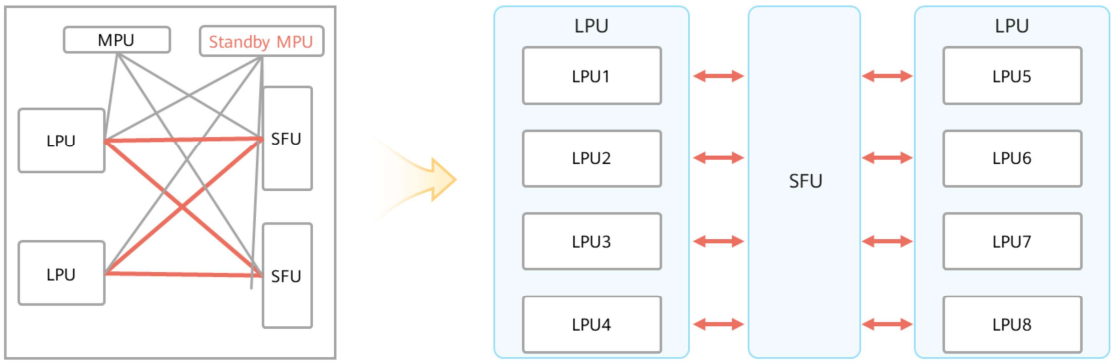
Fixed Device



Different from a modular device, service modules of fixed devices are not independent hardware modules. Instead, they are integrated in one chassis.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Logical Diagram of Module Connections

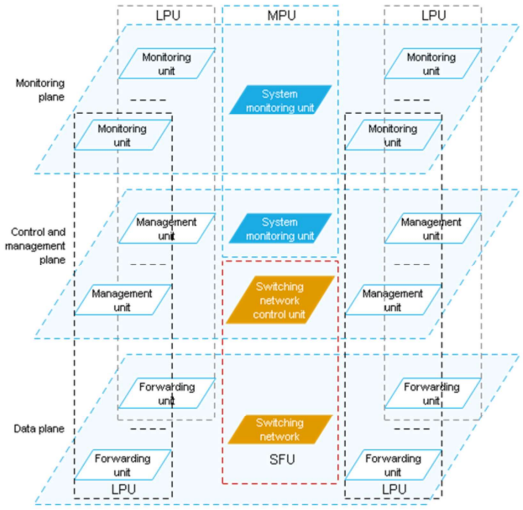


- Modules of a modular device are categorized as different boards. The cards communicate with each other through internal connections of the modular device.
- The fixed integrates the modules, and the modules communicate with each other through internal connections.
- LPUs are connected to each other through SFUs, and communication data between LPUs is forwarded through SFUs.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Logical Architecture of Network Devices

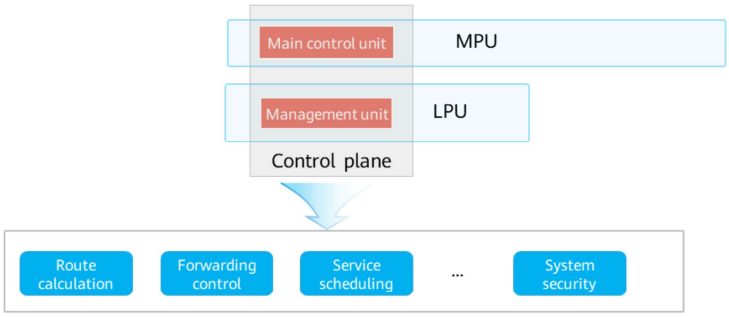
Network devices can be logically divided into three planes: data plane, control and management plane, and monitoring plane.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Control Plane

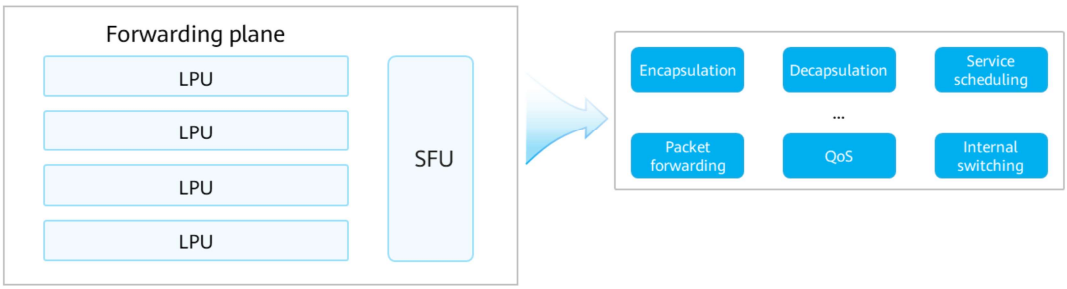
- The control plane of a device consists of the MPU and the management unit of the LPU.
- The control and management plane provides control and management functions for the system and is the core of the entire system. The control plane provides functions such as protocol processing, service processing, route calculation, forwarding control, service scheduling, traffic statistics collection, and system security. The control plane of a switch is used to control and manage the running of all network protocols. The control plane provides various network information and forwarding entries required by the data plane for data processing and forwarding.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Forwarding Plane

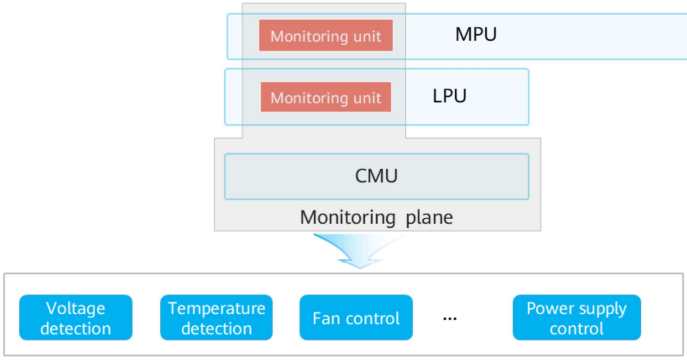
- The forwarding plane consists of SFUs and LPUs.
- An LPU has a forwarding plane engine (FPE), which is essentially a switching chip that implements switching between interfaces on the LPU.
- The data plane is responsible for high-speed processing and non-blocking switching of data packets. It encapsulates or decapsulates packets, forwards IPv4/IPv6/MPLS packets, performs QoS and scheduling, completes inner high-speed switching, and collects statistics.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Monitoring Plane

- The monitoring plane consists of the monitoring units of MPUs and LPUs. Some modular switches have independent centralized monitoring units (CMUs).
- The monitoring plane monitors the ambient environment to ensure the secure and stable operation of the system. It detects voltage levels, controls system power-on and power-off, monitors the temperature, and controls fan modules. If a unit fails, the monitoring plane isolates the faulty unit promptly so that the other units remain unaffected.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

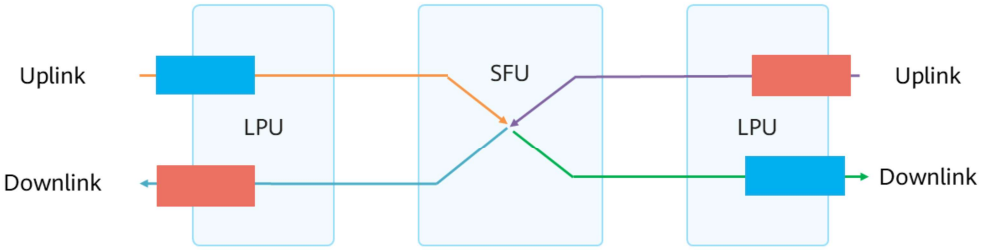
Contents

1. Overview of Network Device Framework
2. **Packet Processing by Network Devices**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Uplink and Downlink Packet Forwarding

With the SFU as the center, packet forwarding involves uplink and downlink processing.

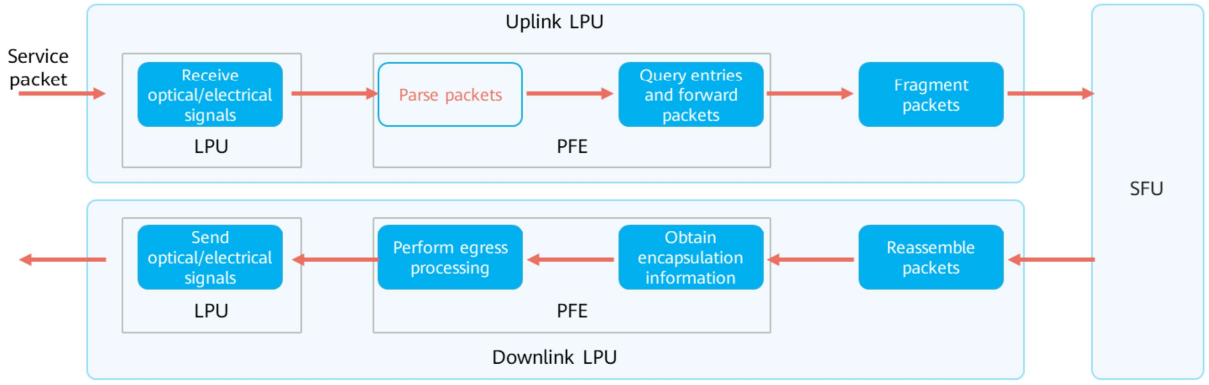


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Packets processed by a network device are classified into service packets and protocol packets.
- The device only forwards service packets from one interface to another interface based on forwarding entries.
- After receiving protocol packets (such as ARP, OSPF, and BGP packets), the device sends the packets to the control plane for processing. For example, the ARP packets are sent to the control plane for processing. After determining whether to respond to the ARP packets, the device determines whether to learn the source MAC address and source IP address in the ARP packets.

Processing of Service Packets

After service packets enter the uplink LPU from an interface, they are sent to the SFU through the internal bus of the modular switch. The SFU sends the service packets to the downlink LPU for processing and then sends them out from the interface.

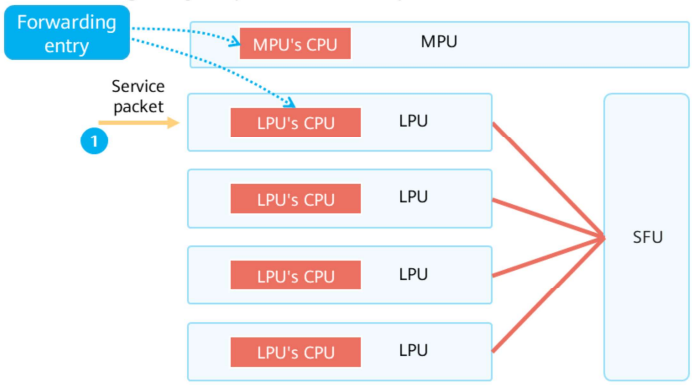


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- PFE: Packet Forwarding Engine
- Service packets: packets during interaction between services and applications
- Fragmentation: Before packets are sent to the SFU, they are sliced with a fixed length based on a certain granularity.
- Reassembly: Fragmented packets sent from the SFU are reassembled.

Determining the Egress of Packets (1)

- When a packet enters an LPU, the device determines the outbound interface of the packet based on the forwarding entry (such as the IP routing table and MAC address table). For a modular switch, the downlink LPU needs to be determined.
- When the packet reaches the SFU, the outbound interface and downlink LPU are specified. Therefore, the forwarding entry is queried on the uplink LPU.



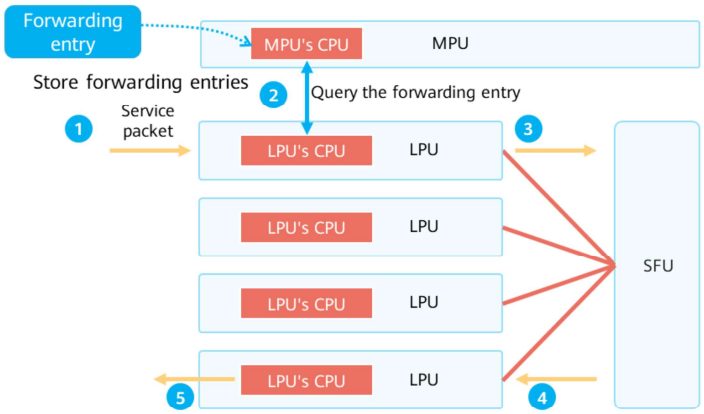
Both the MPU and LPU have CPUs and provide the control plane function. Are forwarding entries stored and queried on the MPU or LPU?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- For Layer 2 forwarding, the MAC address table is queried. For Layer 3 forwarding, the Layer 3 routing table is queried.

Determining the Egress of Packets (2)

- Forwarding entries are stored on the MPU. After packets enter the LPU, the LPU queries entries from the MPU.
- The LPU needs to communicate with the MPU when packets are forwarded each time. The forwarding efficiency is low and the packet delay increases. For a high-rate LPU, the forwarding rate decreases greatly.

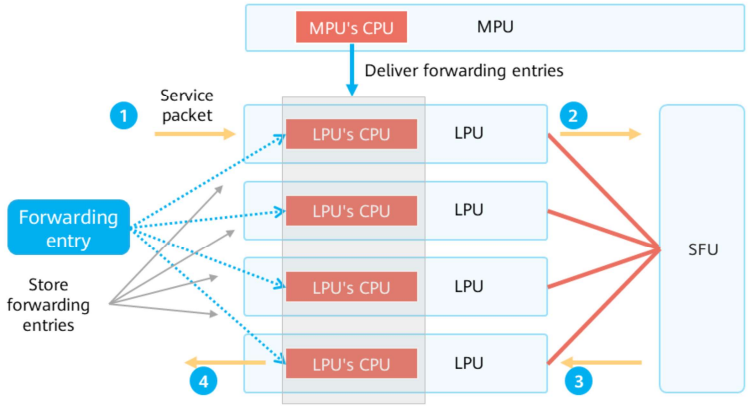


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- For Layer 2 forwarding, the MAC address table is queried. For Layer 3 forwarding, the Layer 3 routing table is queried.

Determining the Egress of Packets (3)

- Forwarding entries are stored on the LPU. After packets enter the LPU, the LPU directly queries the packets, improving the packet forwarding efficiency.
- Forwarding entries are stored on all LPUs, causing high resource usage on the control plane.

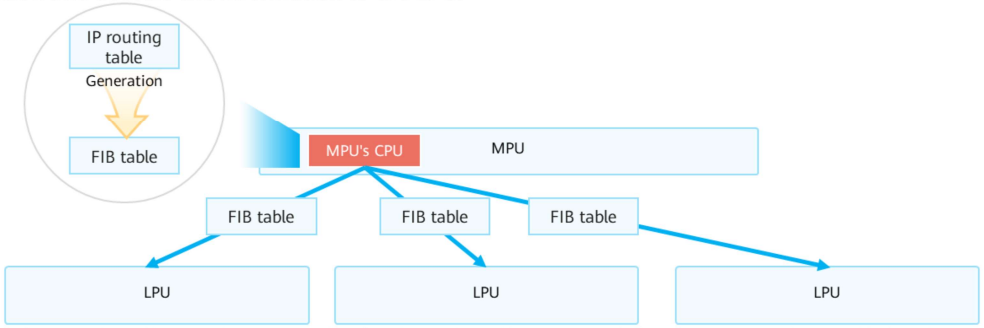


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- For Layer 2 forwarding, the MAC address table is queried. For Layer 3 forwarding, the Layer 3 routing table is queried.

Forwarding Information

- Service packets of high-end devices are not processed by the CPU of the MPU, and forwarding information is queried by the LPU.
- The forwarding information on the LPU does not exist in forwarding entries (such as the IP routing table and MAC address table) of the MPU. After generating forwarding entries, the MPU generates corresponding forwarding information and delivers the information to the LPU.

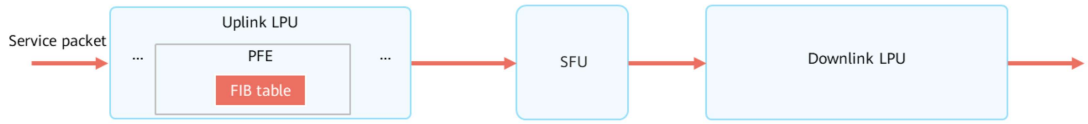


The IP routing table is used as an example. After the routing table is generated, the MPU generates the FIB table based on the routing table and delivers the FIB table to the LPU. The LPU then forwards packets based on the FIB table.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Hardware-based Forwarding

- The PFE, usually an NP or ASIC chip, is responsible for packet forwarding on an LPU. The LPU directly forwards packets without the involvement of the MPU.
- The forwarding plane and control plane of a high-end modular switch are distributed to different components. The MPU on the control plane runs forwarding-related protocols and maintains forwarding entries. The LPU on the forwarding plane can independently complete forwarding based on forwarding information delivered by the control plane. When the MPU on the control plane is heavily loaded, the forwarding plane is not affected. This working mechanism is called forwarding-control decoupling.

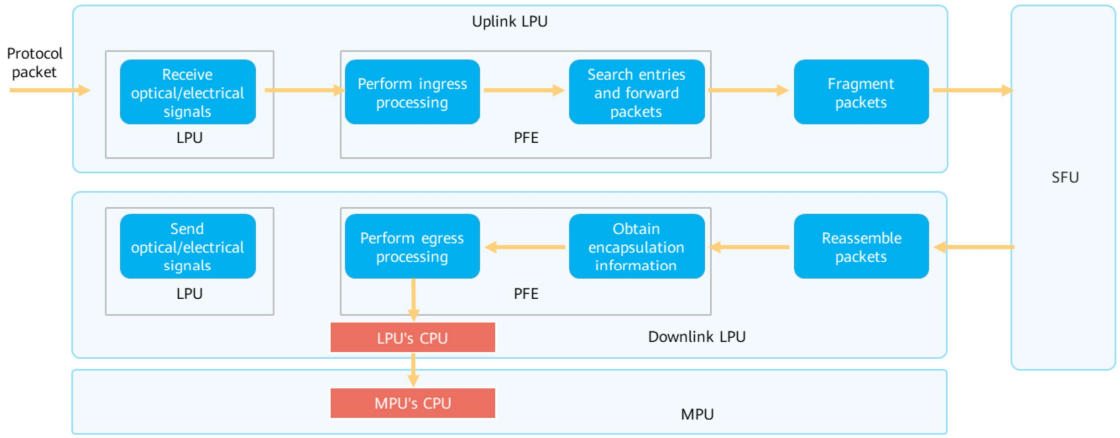


Service packets are forwarded by the LPU and SFU independently, without the participation of the MPU.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Processing of Protocol Packets

Protocol packets received by a device, such as OSPF packets, IS-IS packets, BGP packets, ARP packets, STP BPDUs, and ICMP Request packets, need to be processed by the control plane of the device. That is, these packets are sent to the CPU of the MPU for processing.



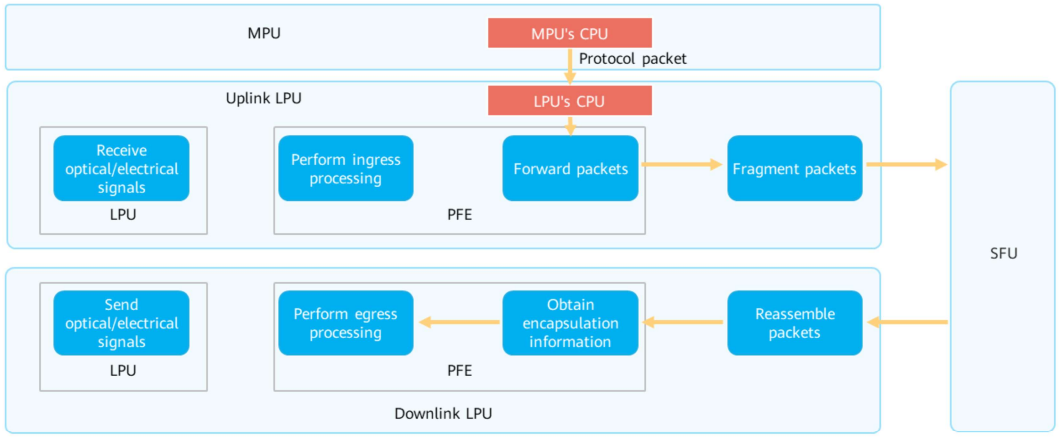
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- After receiving protocol packets, the CPU of the MPU processes the packets. If the CPU needs to respond to the packets, the control board constructs the protocol packets. For example, after receiving ARP Request and ICMP Echo Request packets sent to the CPU, the MPU constructs the ARP Reply and ICMP Echo Reply packets.
- The CPU processing capability of the MPU is limited. If too many protocol packets are sent to the CPU of the MPU, the CPU is busy and cannot respond to the protocol packets in a timely manner. Therefore, the rate at which various protocol packets are sent to the CPU of the MPU is limited by default.
- When the forwarding engine PFE (NP or ASIC chip) parses packets, the protocol field in the Layer 2 frame header can be used to determine whether the packets need to be sent to the local CPU for processing. (such as ARP, RARP, IS-IS, LLDP, LACP, and PPP control packets) or the destination address is a specified reserved multicast IP address. (As defined in the standard, multicast addresses 224.0. 0.1 to 224.0. 0.255 are used by routing protocols.) The protocol packets of the do not need to be forwarded in the upstream direction by looking up the table.

- After the packets are forwarded by querying the table in the upstream direction, the SFU obtains the information about the destination interface board and the outbound interface of the packets. In this way, the SFU can exchange data to the corresponding downstream board according to the information about the destination interface, and the downstream board sends the packets according to the outbound interface information. For the protocol packets (such as BGP) that can be identified by packet parsing, which are forwarded in the upstream direction without querying the table. What are the destination interface board and outbound interface information? The answer is that the board number is the same as that of the inbound interface board, and the outbound interface is the CPU. That is, the flow chart in the preceding figure applies to this type of protocol packets. The uplink LPU and the downlink LPU are actually the same LPU.
- For protocol packets that can be identified based on the Layer 2 frame header or destination IP address, the upstream LPU directly sends the packets to the CPU.

Processing of Protocol Packets Sent by the Device

Protocol packets sent by the device, such as OSPF packets, IS-IS packets, BGP packets, ARP packets, STP BPDUs, and ICMP packets, are constructed by the CPU of the MPU and then sent by the LPU.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Single) Which of the following modules on a modular switch is responsible for running routing protocols and generating and maintaining routing tables?
   A. LPU
   B. SFU
   C. MPU
   D. SPU
2. (Essay) When forwarding service packets, does a high-end modular switch query forwarding entries from the MPU?

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. C
2. No. The high-end modular switch delivers forwarding information to the LPU, and the LPU directly forwards packets without querying forwarding entries from the MPU.

Summary

- Each network device has an independent control plane, forwarding plane, and monitoring plane. The control plane is responsible for protocol processing, route calculation, and service scheduling, and the forwarding plane is responsible for data forwarding to implement packet exchange between service modules. The monitoring plane monitors the ambient environment to ensure the secure and stable operation of the system.
- On a high-end modular switch, each plane has different boards installed. LPUs and SFUs implement functions of the forwarding plane, MPUs implement functions of the control plane, and CMUs implement functions of the monitoring plane.
- High-end modular switches use hardware-based forwarding. Packets are directly forwarded by LPUs, without participation of the control plane. Therefore, the packet forwarding efficiency is high.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



IP Routing Basics

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- When a router receives an IP packet, the router searches the Forwarding Information Base (FIB) table according to the destination IP address of the packet. After finding the most matching routing entry, the router forwards the packet according to the outbound interface or next hop in the routing entry.
- A router can obtain three types of routes: dynamic routes such as the Open Shortest Path First (OSPF), static routes, and direct routes. A network may have the three types of routes. How are routers that use different types of routes are reachable?
- This course describes the basics of IP routing, including the routing forwarding, routing table, and route import.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Identify the IP routing table and FIB table.
  - Analyze the route-based forwarding process.
  - Describe the principles of route import.
  - Understand application scenarios of route import.

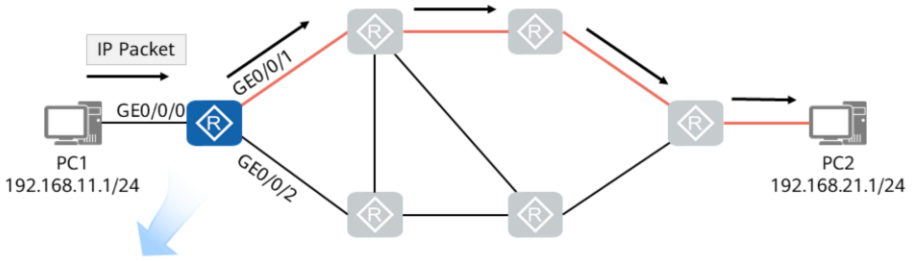
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **IP Routing Basics**
2. Advanced Applications of IP Routing

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Overview of IP Routing



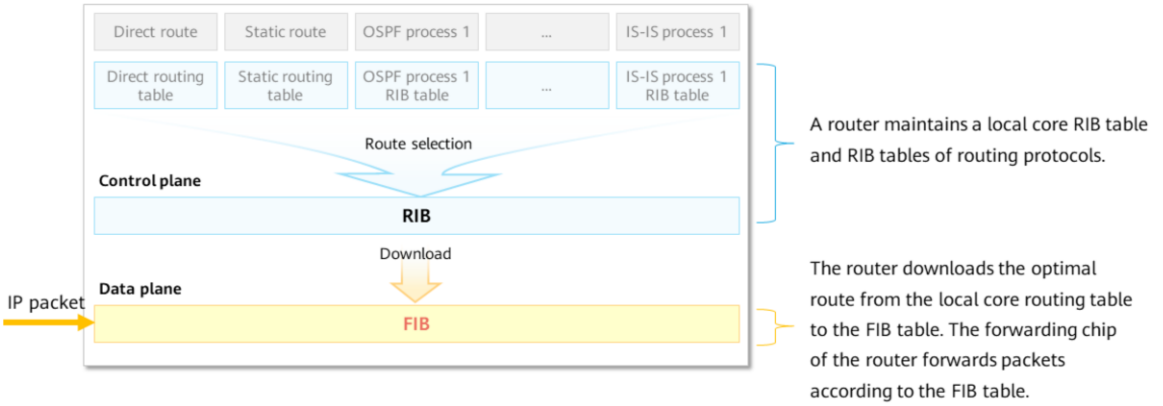
When receiving an IP packet, a router matches the destination address of the IP packet with a routing entry.

- If the IP packet matches a routing entry, the router forwards the packet according to the outbound interface or next hop in the routing entry.
- If the IP packet does not match any routing entry, the router does not have routing information to guide packet forwarding. In this case, the router discards the packet.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

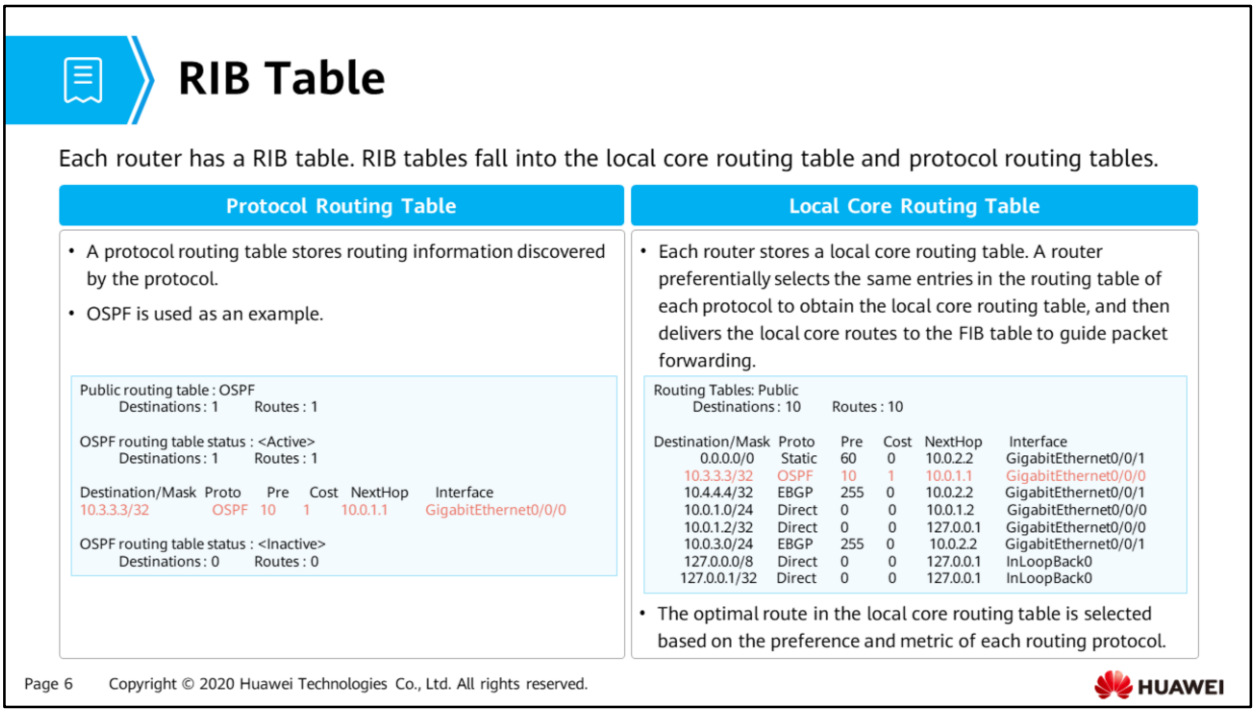
RIB and FIB

A network device that provides the routing function maintains two important data tables: Routing Information Base (RIB) and Forwarding Information Base (FIB).

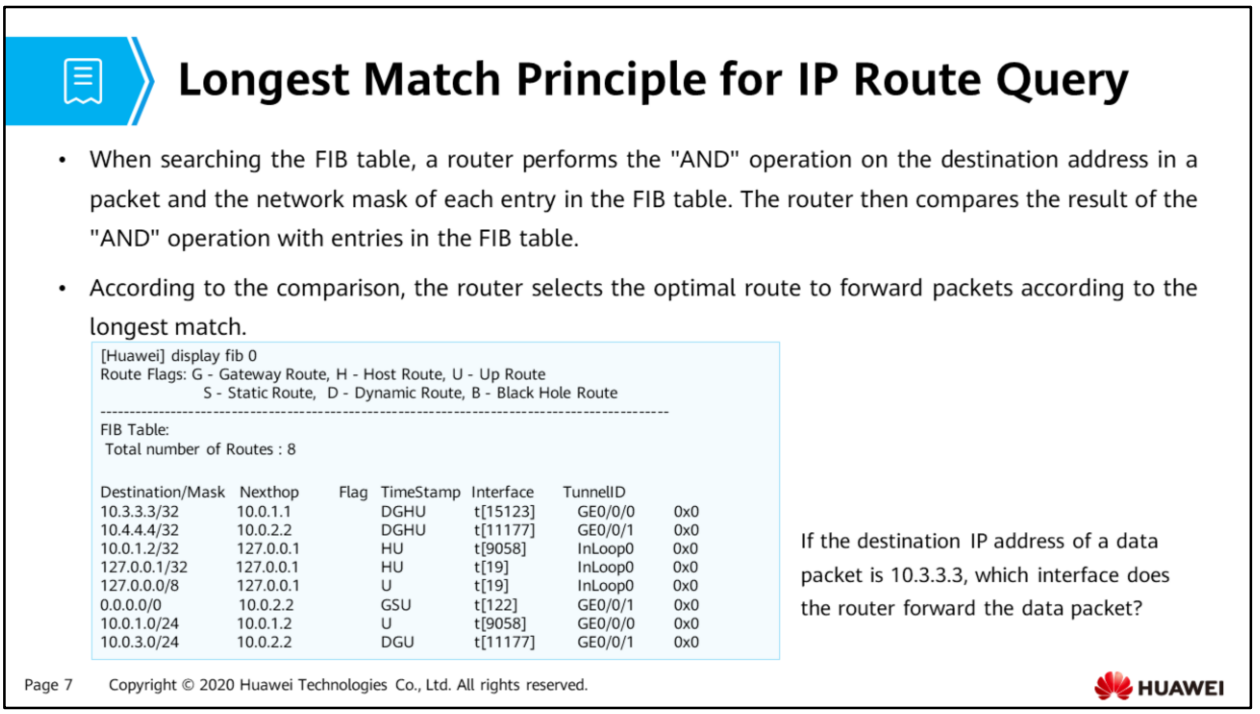


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- RIB table:
  - A RIB table can be considered to be located on the control plane of a router. Actually, a RIB table does not directly guide data forwarding. When a router queries routes, it does not query the destination address of a packet in the RIB table. Instead, it queries the FIB table to guide data forwarding. The router downloads the optimal route from the RIB table to the FIB table. If related entries in the RIB table change, the FIB table is synchronized immediately.
  - Because the two tables are consistent and the RIB table is easy to read, the RIB table (routing table) is used in most cases to describe the data forwarding process of a router. Actually, the router queries the FIB table, and the RIB table at the control layer provides only routing information.
- FIB table:
  - The FIB table is located on the data plane of a router and is also called the forwarding table. Each forwarding entry specifies the outbound interface and next-hop IP address for reaching a destination.
- Note:
  - Huawei routers and Layer 3 switches provide the routing function. This course uses routers as an example.
  - Both OSPF and Intermediate System to Intermediate System (IS-IS) use the Shortest Path First (SPF) algorithm to calculate routes based on link state information. For details about OSPF and IS-IS, see the following courses.
  - Routing process: A router supports multiple OSPF and IS-IS processes. Different processes can be assigned based on service types, and they are independent of each other. An OSPF process ID takes effect on the local device, and does not affect packet exchange between the local route and other routers. Packets can be exchanged between routers with different process IDs.



- Key fields in a routing table:
  - Destination: indicates the destination address of a route. It identifies the destination IP address or destination network segment of IP packets.
  - Mask: indicates the subnet mask of the destination IP address. It is used with the destination address to identify the address of the network segment where the destination host or router is located.
  - Proto (protocol): indicates the protocol through which routes are learned.
  - Pre (Preference): indicates the routing protocol preference of the route.
    - Routers define external and internal preferences. The external preference can be manually configured for each routing protocol, while the internal preference cannot be manually modified.
    - During route selection, a router first compares the external preferences of routes. When the same external preference is set for different routing protocols, the router selects the optimal route based on the internal preference.
  - Cost: indicates the cost of a route.
  - NextHop: indicates the next hop to the destination network. It specifies the next-hop device to which packets are forwarded.
  - Interface: indicates the outbound interface that forwards packets to the destination network. It specifies the local router interface from which packets are forwarded.
- The Preference value is used to compare the preferences of different routing protocols, while the Cost value is used to compare the preferences of different routes of the same routing protocol.
- Note: The routing table in the body is truncated.

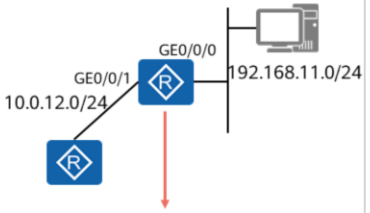


- Each entry in the FIB table contains the physical or logical interface through which a packet is sent to a network segment or host to reach the next-hop router. An entry also indicates whether the packet can be sent to a destination host on a directly connected network.
- The **display fib** [*slot-id*] command is used to check information about the FIB table.
  - slot-id: displays information about the FIB table with a specified slot ID. The value is an integer, and the value range depends on the device configuration.
- Fields in the FIB table:
  - Total number of Routes: indicates the total number of routes in the routing table.
  - Destination/Mask: indicates the destination address or mask length.
  - Nexthop: indicates the next hop.
  - Flag: indicates the current flag, which is the combination of G, H, U, S, D, and B.
    - G (Gateway): indicates that the next hop is a gateway.
    - H (Host): indicates that the next hop is a host.
    - U (Up): indicates that the route status is Up.
    - S (Static): indicates the static route.
    - D (Dynamic): indicates the dynamic route.
    - B (Blackhole): indicates the blackhole route, with the next hop as a null interface.

Route Types

Direct Route

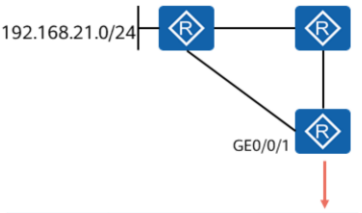
Direct routes are automatically generated by devices and point to local directly connected networks.



Route Type	Destination/Mask	Outbound Interface
Direct route	192.168.11.0/24	GE0/0/0
Direct route	10.0.12.0/24	GE0/0/1

Static Route

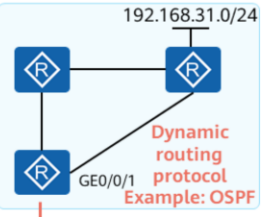
Static routes are manually configured by network administrators.



Route Type	Destination/Mask	Outbound Interface
Static route	192.168.21.0/24	GE0/0/1

Dynamic Route

Dynamic routes are learned by dynamic routing protocols running on routers.



Route Type	Destination/Mask	Outbound Interface
OSPF	192.168.31.0/24	GE0/0/1

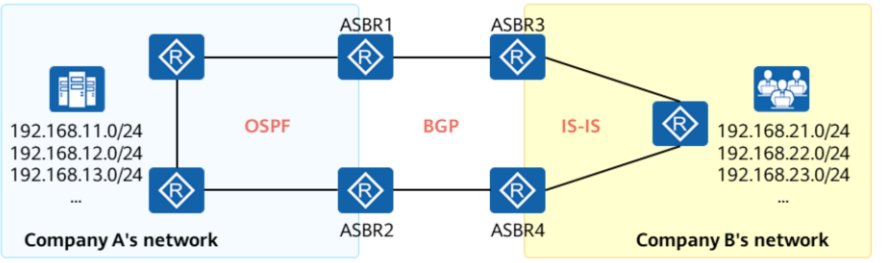
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Direct routes are destined for the subnets to which directly connected interfaces belong. They are automatically generated by devices.
- Static routes are manually configured by network administrators.
- Dynamic routes are learned by dynamic routing protocols, such as OSPF, IS-IS, and Border Gateway Protocol (BGP).
  - The Border Gateway Protocol (BGP) is a distance vector routing protocol that allows devices in different ASs to communicate and select optimal routes.
  - An AS is a group of IP networks that are controlled by one entity, typically an Internet service provider (ISP), and have the same routing policy.

Dynamic Routing Protocols

Dynamic routing protocols are classified into the following types based on application scopes:

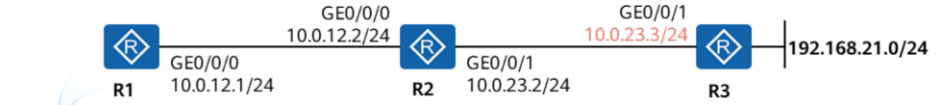
- Interior Gateway Protocols (IGPs): run inside an autonomous system (AS), including the OSPF and IS-IS.
- Exterior Gateway Protocols (EGPs): run between ASs, including the Border Gateway Protocol (BGP).



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Recursion

Routes can be used to forward traffic only when they have directly connected next hops. However, a static or BGP route may contain an indirect next hop. Therefore, in a process referred to as route recursion, a device needs to search for a directly connected next hop for the route.



ip route-static 192.168.21.0 24 10.0.23.3

Recursion

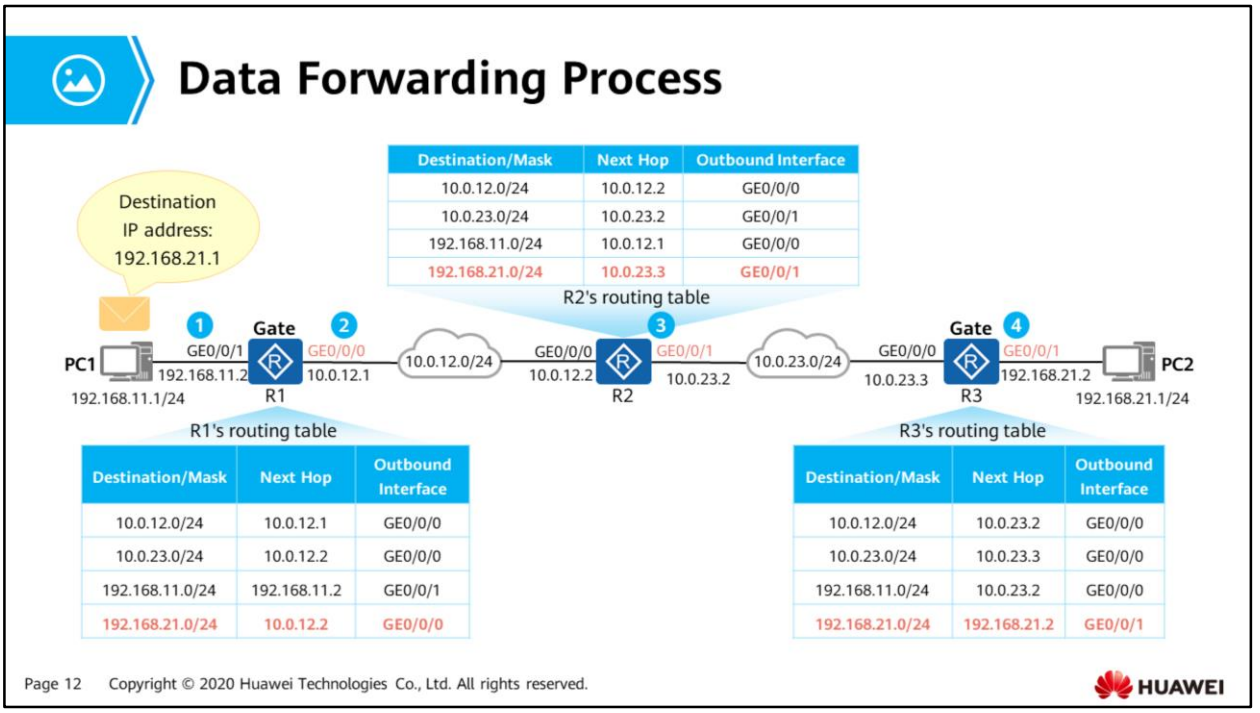
ip route-static 10.0.23.0 24 10.0.12.2

Destination Network	Next Hop	Outbound Interface
192.168.21.0/24	10.0.23.3	GE0/0/0
10.0.23.0/24	10.0.12.2	GE0/0/0

Obtain the directly connected next hop through recursion

In this example, a route to 10.0.23.3 is added to R1 so that the route to 192.168.21.0/24 can be recursive.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- The process for PC1 to send a data packet to PC2 is as follows:
  1. PC1 sends the packet to the gateway R1.
  2. R1 searches the routing table for the next hop and outbound interface, and forwards the packet to R2.
  3. R2 forwards the packet to R3 based on the routing table.
  4. After receiving the packet, R3 looks up the routing table and finds that the destination IP address of the packet belongs to the network segment where the local interface resides. R3 then forwards the packet locally and finally sends the packet to the destination PC2.

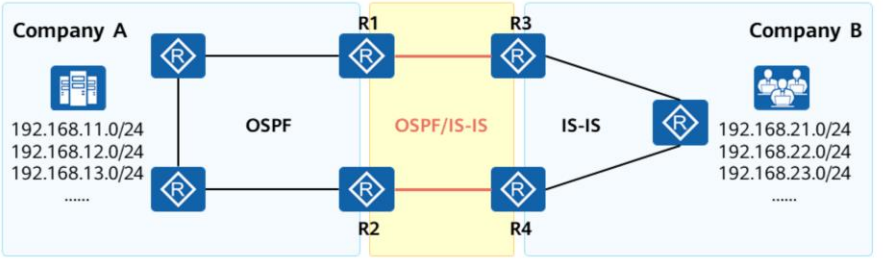
Contents

1. IP Routing Basics

2. Advanced Applications of IP Routing

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Scenario Analysis of Advanced Applications of IP Routing (1)



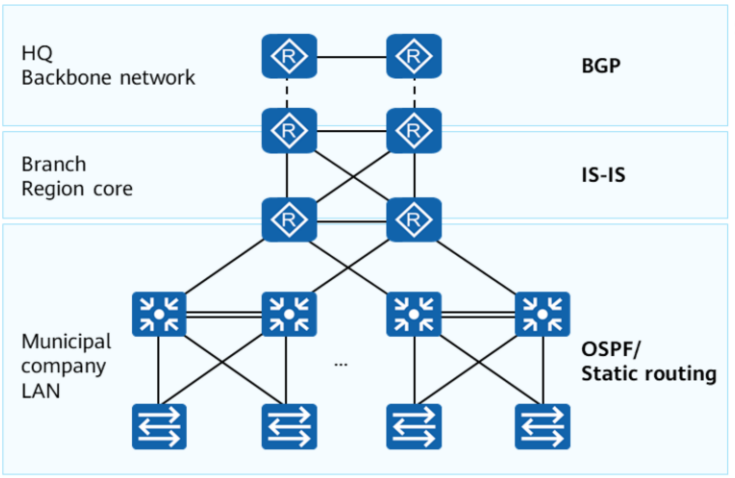
Scenario:

- Assume that company A and company B have their own networks, which are managed and maintained independently. The networks of company A and company B run OSPF and IS-IS, respectively.
- After the two companies are merged into one company, the original two networks must be integrated. To ensure that service traffic of the new company can be normally exchanged on the integrated network, the company requires interworking based on routing.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- OSPF and IS-IS are two different dynamic routing protocols, so they cannot directly exchange routing information.
- In the figure, OSPF is deployed on the network of company A, and R1 and R2 are edge devices. IS-IS is deployed on the network of company B, and R3 and R4 are edge devices. OSPF or IS-IS can be deployed on the connected network segments of borders. For example, OSPF can be deployed on network segments between R1 and R3 and between R2 and R4. In this case, only R3 and R4 are border devices.

Scenario Analysis of Advanced Applications of IP Routing (2)

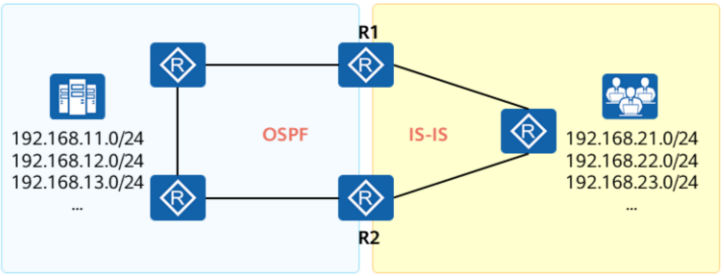


Scenario:

- On a large-scale enterprise network, a single routing protocol cannot meet network requirements. In most cases, multiple routing protocols coexist.
- In addition, different routing protocols are designed and deployed in different network topologies considering the service logic or administrative management, making the routing hierarchy clear and controllable.
- In such a network environment, interworking based on routing also needs to be implemented.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Concept of Route Import



Routing information of two routing protocols is isolated from each other.

To implement interworking based on routing, perform the following operations:

1. Re-plan and modify network-wide routing protocols, which is complex.
2. Perform operations on the edge devices in OSPF and IS-IS routing domains so that routing information can be transmitted between OSPF and IS-IS. This solution does not change the original topology and is easy to deploy. However, loops may occur.

Route import refers to the process of advertising routing information from one routing protocol to another.

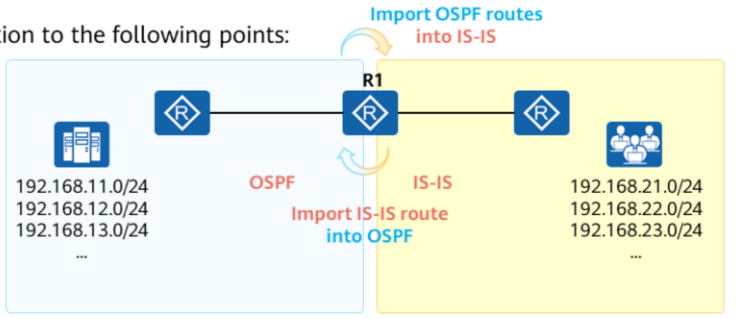
- Route import allows routing information to be transmitted between different routing protocols.
- When importing routes, you can deploy route control to flexibly control service traffic.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In the figure, OSPF and IS-IS networks have different network segments. Only R1 and R2 know all routing entries.
- Question: How do all devices obtain all routes?

Direction of Route Import

- Route import is directional. If routing information is imported from routing protocol A to routing protocol B, routing protocol B can learn routing information about routing protocol A. However, routing protocol A does not know routing information about routing protocol B, unless route import from routing protocol B to routing protocol A is configured.
- When importing routes, pay attention to the following points:
  - Route preference
  - Route injection
  - Route metric

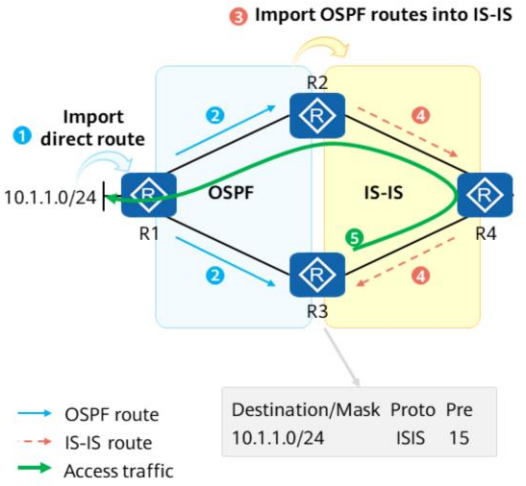


Question: Can two networks communicate with each other if only OSPF routes are imported to IS-IS?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- During route import, focus on the route convergence time. This course does not describe the route convergence time.
- The implementation and configuration of route import will be described in other HCIP-Datacom certification courses.

Route Import: Route Preference

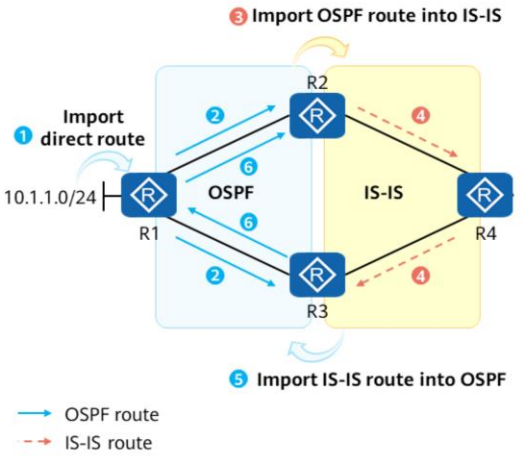


- Scenario:
  1. R1 imports the direct route 10.1.1.0/24 into OSPF.
  2. R3 learns the route (OSPF external route with the preference of 150) to the network segment 10.1.1.0/24 through OSPF.
  3. R2 imports OSPF routes to the IS-IS process.
  4. R3 also learns the route (the preference is 15) to the network segment 10.1.1.0/24 through IS-IS.
  5. For R3, the IS-IS route has a higher priority than the OSPF external route, so the IS-IS route from R4 is preferred.
- R3 accesses the network segment 10.1.1.0/24 through the path R3->R4->R2->R1, which is the second optimal path.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Route preferences defined by Huawei:
  - Direct: 0
  - OSPF: 10
  - IS-IS: 5
  - Static: 60
  - OSPF ASE: 150
  - OSPF NSSA: 150
  - IBGP: 255
  - EBGP: 255
- Note: The route preferences may vary with vendors.

Route Import: Route Injection

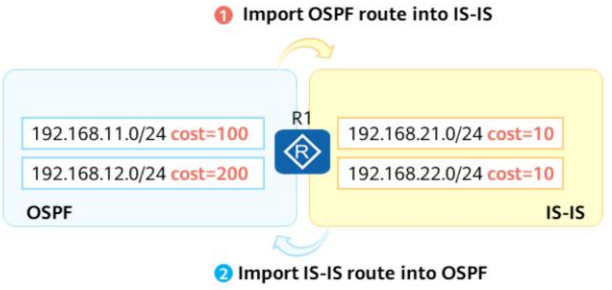


Scenario:

1. R1 imports the direct route 10.1.1.0/24 to OSPF.
2. The direct route 10.1.1.0/24 is advertised in the entire OSPF domain.
3. R2 imports OSPF routes to the IS-IS process.
4. The direct route 10.1.1.0/24 is advertised in the entire IS-IS domain.
5. R3 imports IS-IS routes to the OSPF process.
6. The direct route 10.1.1.0/24 is advertised in the entire OSPF domain again, resulting in route injection.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Import: Route Metric



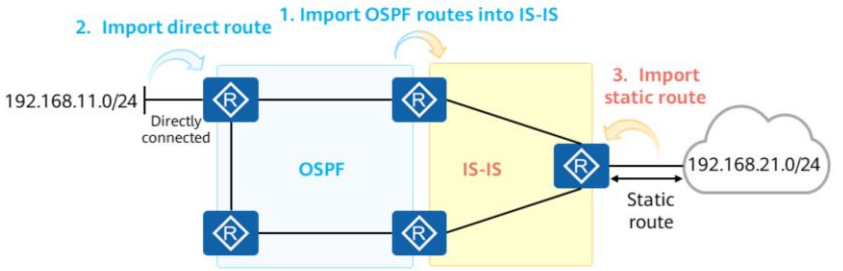
- Scenario:
  1. OSPF routes are imported into IS-IS.
  2. IS-IS routes are imported into OSPF.
- Different routing protocols define different route metrics. How are metrics of imported routes defined when routes are imported between routing protocols? What are the metrics?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Import Scenario

Route import involves the following scenarios:

1. Route import between dynamic routing protocols
2. Importing direct routes to a dynamic routing protocol
3. Importing static routes to a dynamic routing protocol



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Configuration Commands for Route Import

1. Configure OSPF to import external routes.

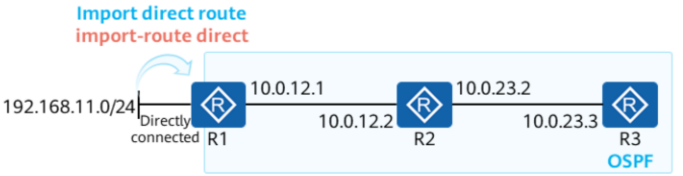
   [Huawei-ospf-100] **import-route** { **bgp** | **direct** | **static** | **isis** [*process-id-isis*] | **ospf** [*process-id-ospf*]}

   In the OSPF view, import BGP routes, direct routes, static routes, IS-IS routes, and routes of other OSPF processes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- If a device on an OSPF network needs to access a device on the network running a non-OSPF protocol, the OSPF device needs to import routes from the non-OSPF protocol into the OSPF network.

Case 1: Importing Direct Routes to OSPF



R1's routing table

Destination/Mask	Protocol	Next Hop
192.168.11.0/24	Direct	192.168.11.1
10.0.12.0/24	Direct	10.0.12.1
10.0.23.0/24	OSPF	10.0.12.2

R3's routing table

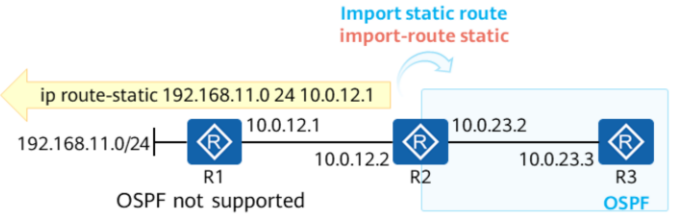
Destination/Mask	Protocol	Next Hop
192.168.11.0/24	O_ASE	10.0.23.2
10.0.12.0/24	OSPF	10.0.23.2
10.0.23.0/24	Direct	10.0.23.3

- You can run the **import-route direct** command to import all direct routes in the routing table to a dynamic routing protocol.
- The imported routes are advertised as OSPF external routes on the entire OSPF network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- To enable a device configured with a dynamic routing protocol to advertise the routes of its directly connected interface to a dynamic routing protocol, enable the dynamic routing protocol on the interface. In addition, direct routes can be imported to dynamic routing protocols.
- In the figure:
  - OSPF is deployed on R1, R2, and R3. R1 has a direct network segment 192.168.11.0/24. To enable R2 and R3 to generate a route to 192.168.11.0/24, import the direct route to OSPF on R1.
- Note: On an OSPF network, if the protocol field in the routing table is displayed as O_ASE, the route is an OSPF external route.

Case 2: Importing Static Routes to OSPF



R2's routing table

Destination/Mask	Protocol	Next Hop
192.168.11.0/24	Static	10.0.12.1
10.0.12.0/24	Direct	10.0.12.2
10.0.23.0/24	Direct	10.0.23.2

R3's routing table

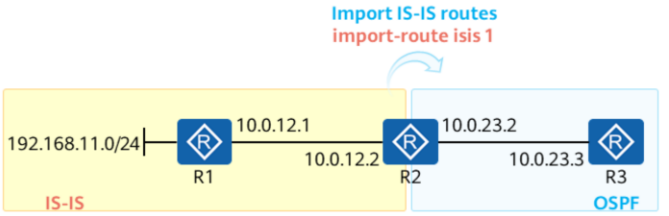
Destination/Mask	Protocol	Next Hop
192.168.11.0/24	O_ASE	10.0.23.2
10.0.23.0/24	Direct	10.0.23.3

- You can run the **import-route static** command to import all static routes in the routing table to a dynamic routing protocol.
- The imported routes are advertised as OSPF external routes on the entire OSPF network.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- For dynamic routing protocols, static routes are considered as external routes and are not detected by dynamic routing protocols. To enable all devices in a dynamic routing protocol domain to learn a static route, import the static route to the dynamic routing protocol.
- In the figure:
  - R2 and R3 run OSPF, but R1 does not support OSPF. Add a static route pointing to network segment 192.168.11.0/24 and import the static route to OSPF on R2 so that both R2 and R3 can generate a route to 192.168.11.0/24.

Case 3: Importing IS-IS Routes to OSPF



R2's routing table

Destination/Mask	Protocol	Next Hop
192.168.11.0/24	IS-IS	10.0.12.1
10.0.12.0/24	Direct	10.0.12.2
10.0.23.0/24	Direct	10.0.23.2

R3's routing table

Destination/Mask	Protocol	Next Hop
192.168.11.0/24	O_ASE	10.0.23.2
10.0.12.0/24	O_ASE	10.0.23.2
10.0.23.0/24	Direct	10.0.23.3

- You can run the **import-route isis 1** command to import all IS-IS routes in the routing table to a dynamic routing protocol.
- The imported routes are advertised as OSPF external routes on the entire OSPF network.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The typical scenario is to import routes from one dynamic routing protocol to another.
- In the figure:
  - IS-IS runs on R1 and R2, and OSPF runs on R2 and R3. The routes maintained by the two protocols are isolated. Therefore, R1 has all routes on the IS-IS network but cannot access the OSPF network. R3 has all routes on the OSPF network but cannot access the IS-IS network. You can configure R2 to import IS-IS routes to OSPF.

Quiz

1. (Single) If a router has the following four FIB entries, which route is used by the router to forward the IP packet with the destination IP address 10.0.1.1? ( )

   A. 0.0.0.0/0

   B. 10.0.0.0.0/16

   C. 10.0.1.0/24

   D. 10.0.2.0/24

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. C

Quiz

2. (Multiple) Which of the following statements about route import are false? ( )

A. In the OSPF process of a router, you can run the import-route command to import routes from other OSPF processes.

B. In the OSPF process of a router, you can run the import-route command to import static routes of the routing table.

C. On a router, to import routes from routing protocol X to routing protocol Y, you need to run the import-route command in the view of routing protocol X.

D. After the import-route command is run on a router and the direct route of GE0/0/1 is imported to OSPF, OSPF is activated on the interface and the interface periodically sends Hello packets.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

2. CD

Summary

- Different routing protocols have different working mechanisms, so multiple routes may be generated to the same destination network segment. A router selects the optimal route based on the priority of the routing protocol and the route cost, and adds the optimal route to the FIB table. The router forwards data according to the FIB table.
- On a large-scale network running multiple routing protocols, routes are advertised between routing protocols by importing routes. A large number of routes may be imported, and some low-performance devices cannot support the imported routes. Therefore, route control is required to implement on-demand route distribution.
- OSPF, IS-IS, and route control listed in this course will be illustrated in subsequent courses.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



OSPF Basics

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- Routers forward data packets based on routing tables. Routing entries can be manually configured or generated using dynamic routing protocols.
- Compared with dynamic routes, static routes use less bandwidth and do not utilize CPU resources for route calculation and update analysis. Static routes alone can implement interworking for simple networks. If a network fault occurs or the topology changes, static routes cannot be automatically updated and must be manually reconfigured to adapt to the network change.
- Compared with static routes, dynamic routing protocols have higher scalability and better adaptability.
- The Open Shortest Path First (OSPF), as an Interior Gateway Protocol (IGP), is widely used because it features high scalability and fast convergence.
- This course describes basic OSPF concepts, OSPF adjacency establishment, and basic OSPF configurations.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- On completion of this course, you will be able to:
  - Describe the overall process of OSPF route calculation.
  - Clarify functions of the DR and BDR.
  - Describe OSPF packets and their functions.
  - Configure basic OSPF functions.
  - Distinguish the OSPF neighbor relationship and adjacency.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

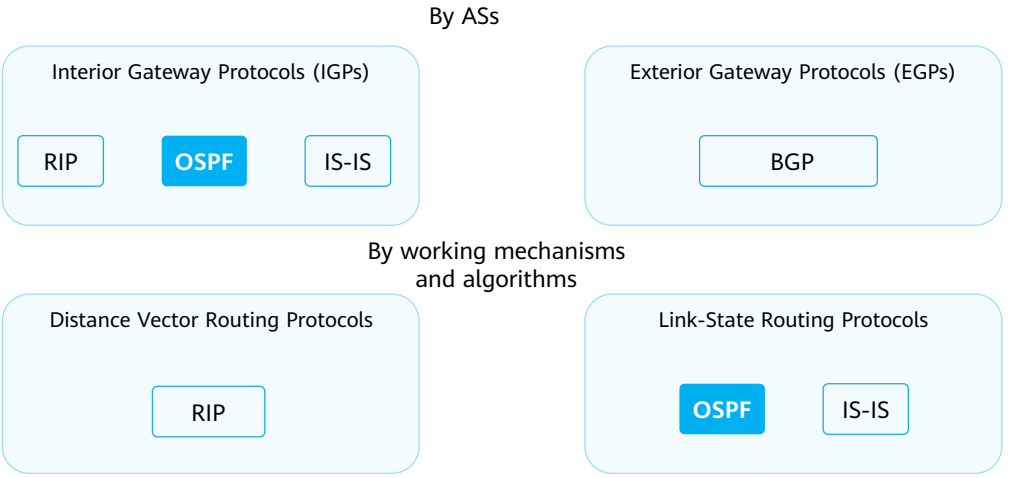
Contents

1. **Introduction to Dynamic Routing Protocols**
2. Overview of OSPF
3. OSPF Working Mechanism
4. Basic OSPF Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Classification of Dynamic Routing Protocols

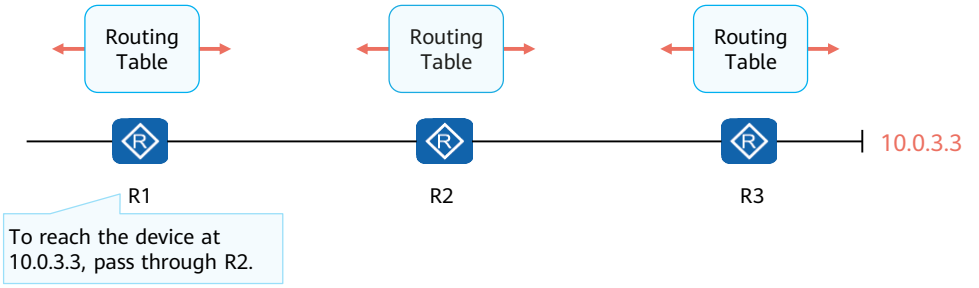


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- BGP uses the path-vector algorithm, which is a modified version of the distance-vector algorithm.

Distance-Vector Routing Protocol

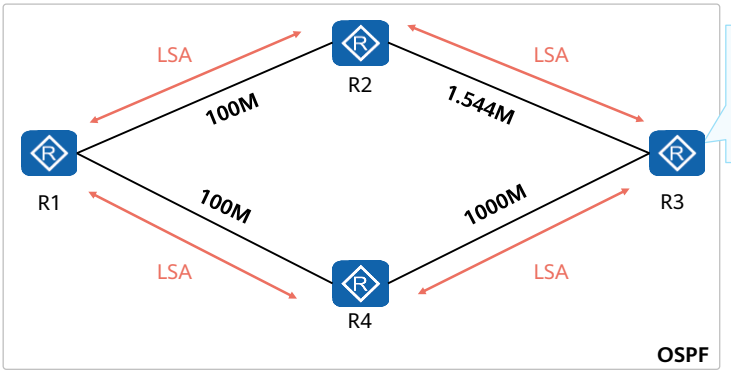
- A router running a distance-vector routing protocol periodically floods its routing table. Through route exchange, each router learns routes from neighboring routers, loads the routes to its routing table, and then advertises the routes to other neighboring routers.
- All routers on a network do not know the network topology. They only know the direction to a destination network segment and the cost.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Link State Routing Protocol: LSA Flooding

- A link-state routing protocol advertises the link state but not routing information.
- Routers running link-state routing protocols establish neighbor relationships and then exchange Link State Advertisements (LSAs).



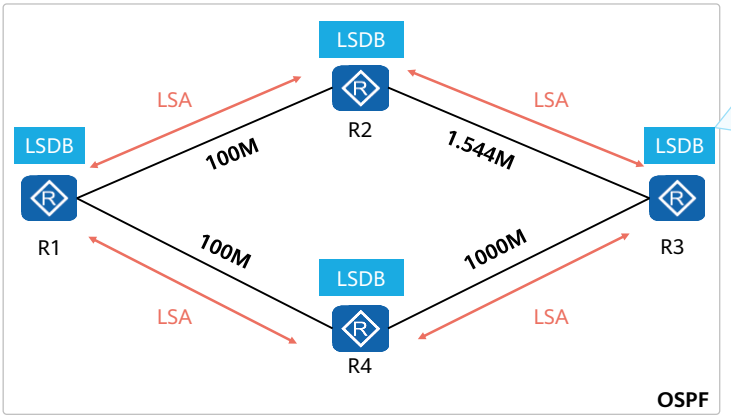
- Advertise LSAs to describe link status information.
- An LSA describes the status of a router interface, such as the cost of the interface and the connected object.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Each router generates a link state advertisement (LSA) that describes the status of its directly connected interface, including the interface cost and the relationship between the router and its neighboring router.

Link State Routing Protocol: LSDB Maintenance

Each router generates LSAs and adds the received LSAs to its own link state database (LSDB). Routers parse the LSAs stored in their LSDBs to obtain the network topology.

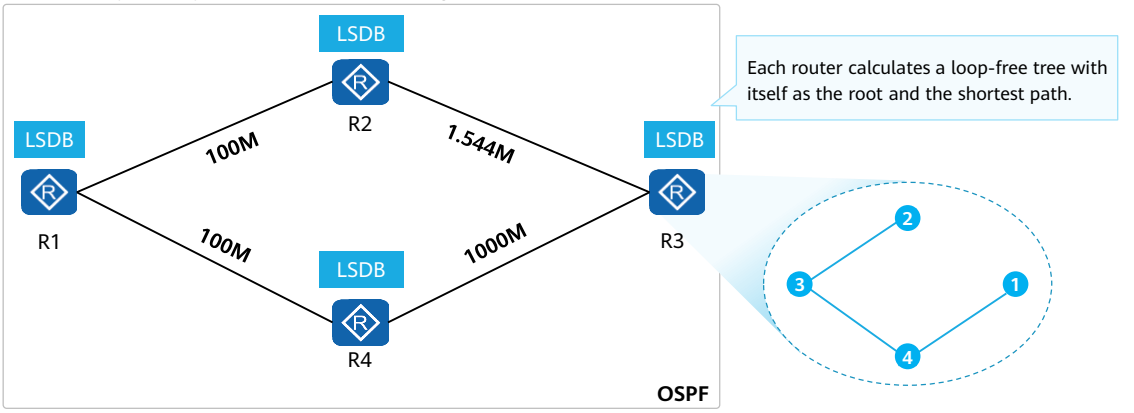


- Routers use LSDBs to store LSAs.
- An LSDB usually stores various types of LSAs, and each type of LSA describes different information.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Link State Routing Protocol: SPF Calculation

Each router uses the Shortest Path First (SPF) algorithm to calculate routes based on the LSDB. Each router calculates a loop-free tree with itself as the root and the shortest path. With the tree, the router knows the optimal paths to all network segments.

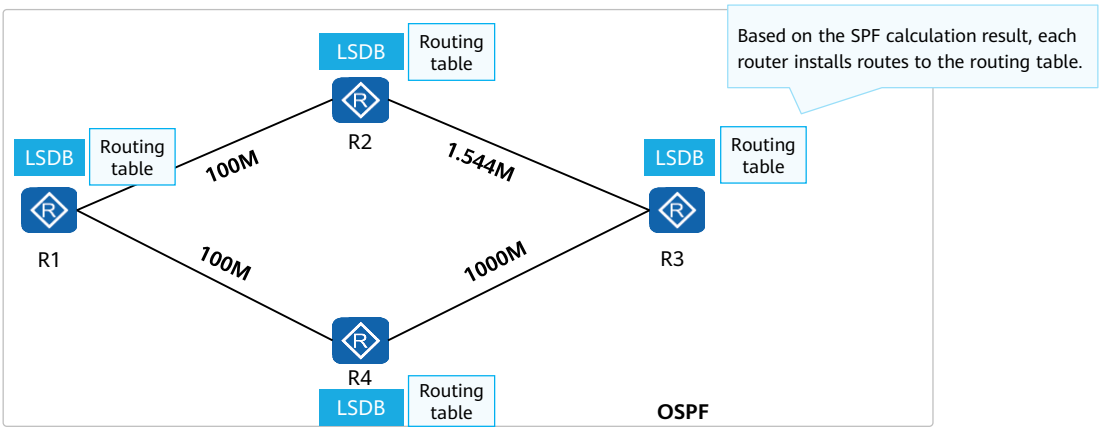


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- SPF is a core algorithm of OSPF. It is used to select optimal routes on a complex network.

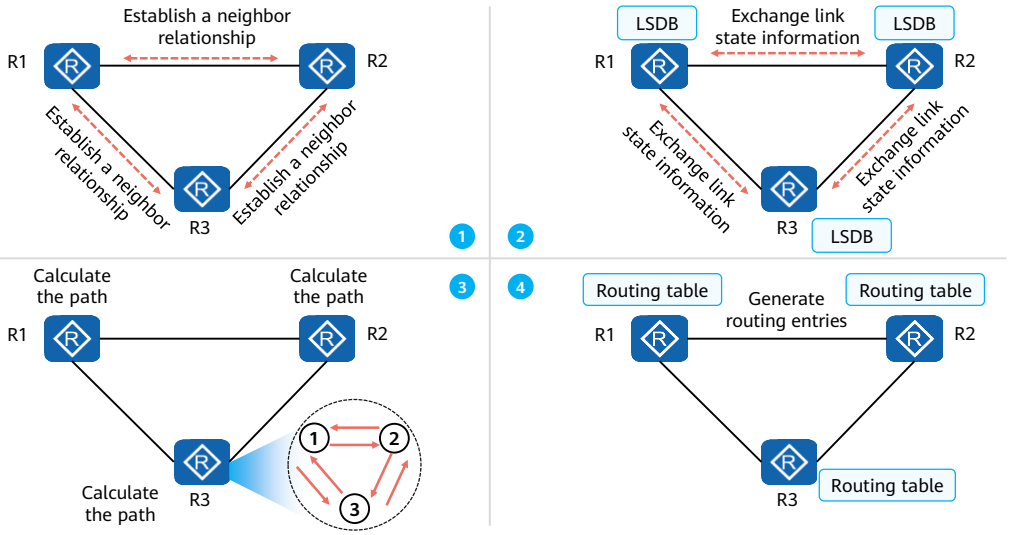
Link State Routing Protocol: Routing Table Generation

A router installs the calculated optimal path to its routing table.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Summary of Link State Routing Protocols



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- A link state routing protocol involves four steps:
  - Step 1: Establish a neighbor relationship between neighboring routers.
  - Step 2: Exchange link state information and synchronize LSDBs between neighbors.
  - Step 3: Calculate the optimal path.
  - Step 4: Generate routing entries according to the shortest path tree and load the routing entries to the routing table.

Contents

1. Introduction to Dynamic Routing Protocols
2. **Overview of OSPF**
3. OSPF Working Mechanism
4. Basic OSPF Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

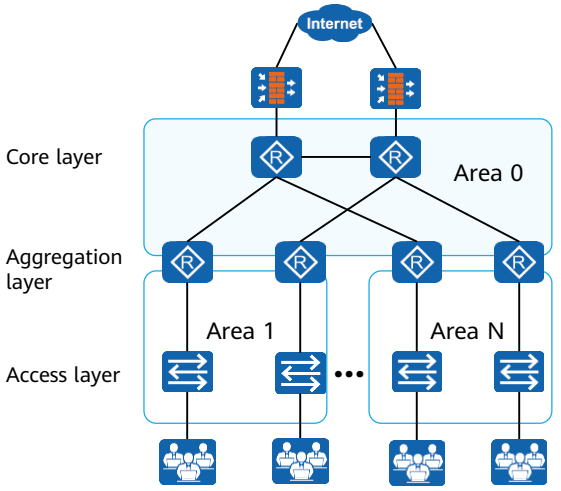
HUAWEI

Overview of OSPF

- OSPF, defined by the Internet Engineering Task Force (IETF), is an IGP based on the link state. OSPF version 2 (OSPFv2), defined in RFC 2328, is intended for IPv4, and OSPF version 3 (OSPFv3)), defined in RFC 2740, is intended for IPv6.
- OSPF has the following advantages:
  - Uses the accumulated link cost as the reference value for route selection based on the SPF algorithm.
  - Transmits and receives some protocol packets in multicast mode.
  - Supports area partition
  - Supports load balancing among equal-cost routes.
  - Supports packet authentication.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

OSPF Application Scenarios



OSPF is usually deployed on large-scale enterprise networks to ensure reachable routes between buildings.

- The core and aggregation layers are deployed in the OSPF backbone area.
- The access and aggregation layers are deployed in the OSPF non-backbone area.

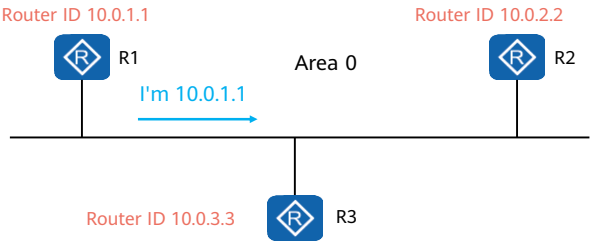
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Access layer: uses transmission media such as optical fibers, twisted pairs, coaxial cables, and wireless access technologies to connect to users and allocate services and bandwidth. The access layer allows terminal users to connect to the network. Therefore, access switches have low costs and high port density.
- Aggregation layer: provides policy-based connections for the access layer, such as address combination, protocol filtering, routing service, and authentication management. Network segments are divided to implement isolation, preventing network faults from spreading and affecting the core layer. The aggregation layer also provides interconnection between virtual networks at the access layer, controls and restricts access from the access layer to the core layer, and ensures security and stability of the core layer.
- Core layer: implements optimized transmission between backbone networks. The core layer focuses on redundancy, reliability, and high-speed transmission.

Basic OSPF Concepts: Router ID

- A router ID is a 32-bit integer that uniquely identifies an OSPF router in an AS.
- The rules for selecting a router ID are as follows:
  - The router ID of an OSPF router is manually configured (recommended).
  - If the router ID is not manually configured, a router uses the largest IP address of a loopback interface as the router ID.
  - If no loopback interface is configured, the router uses the largest IP address of a physical interface as the router ID.

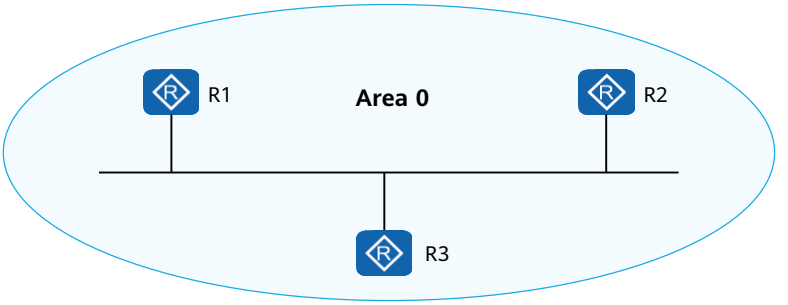


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- To change a specified router ID, you must restart the OSPF process.
- In practice, it is recommended that router IDs of OSPF routers be specified manually. First, plan a private network segment such as 192.168.1.0/24 for OSPF router ID selection. Before starting the OSPF process, create a loopback interface on each OSPF router, and use a private IP address with a 32-bit mask as the IP address of the loopback interface. This private IP address is then used as the router's router ID. If there is no special requirement, this loopback interface address does not need to be advertised to the OSPF network.

Basic OSPF Concepts: Area

- Each OSPF area is regarded as a logical group and identified by an area ID.
- An OSPF area ID is a 32-bit non-negative integer in dotted decimal notation (the format is the same as that of an IPv4 address), for example, area 0.0.0.1. For simplicity, an OSPF area ID is also expressed in decimal notation.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

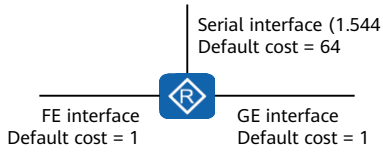
HUAWEI

- For example, area 0.0.0.1 is equivalent to area 1, area 0.0.0.255 is equivalent to area 255, and area 0.0.1.0 is equivalent to area 256. Devices of many network vendors support the two area ID configuration and representation modes.

Basic OSPF Concepts: Metric

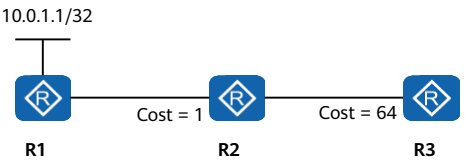
- OSPF uses the cost as the route metric. Each OSPF-enabled interface maintains an interface cost. The default interface cost is 100 Mbit/s divided by interface bandwidth. The value 100 Mbit/s is the default reference value specified by OSPF and is configurable.
- OSPF uses the accumulated cost, that is, the total cost of the outbound interfaces of all routers that the traffic passes from the source network to the destination network.

Cost of the OSPF Interface



Different OSPF interfaces have different costs because of their different bandwidths.

Accumulated Cost of the OSPF Path

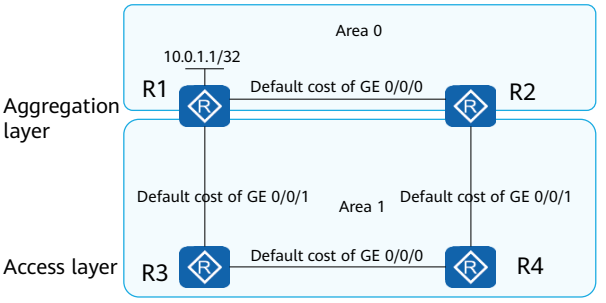


In the routing table of R3, the cost of the OSPF route to network segment 10.0.1.1/32 is 1 plus 64, that is, 65.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In practice, you are advised to manually set the cost based on the interface bandwidth instead of changing the OSPF reference bandwidth.

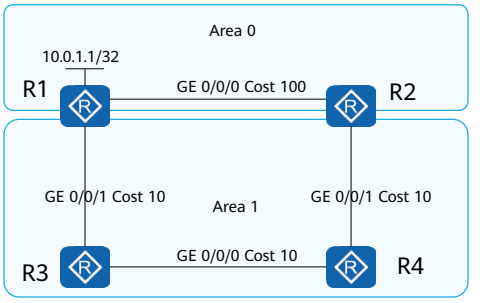
Basic OSPF Concepts: Example for Changing the Metric



[R4]display ip routing-table 10.0.1.1
Summary Count : 2

Destination/Mask	Proto	Cost	NextHop	Interface
10.0.1.1/32	OSPF	2	10.0.34.3	GigabitEthernet0/0/1
	OSPF	2	10.0.24.2	GigabitEthernet0/0/0

By default, there are two paths from R4 to network segment 10.0.1.1/32, and the data forwarding path is uncontrollable.



[R4]display ip routing-table 10.0.1.1
Summary Count : 1

Destination/Mask	Proto	Cost	NextHop	Interface
10.0.1.1/32	OSPF	20	10.0.34.3	GigabitEthernet0/0/0

In the figure, the cost of the device interface is changed to ensure that traffic does not need to pass through R2 when the access router accesses R1.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

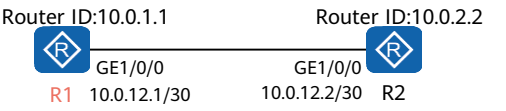
- During traffic path planning, it is recommended that the cost of the direct link at the aggregation layer be greater than the sum of costs of all links on the access ring. In this way, traffic can be directly transmitted to R1 or R2 through the access router.
- The preceding figure is used as an example. R1 and R2 are located at the aggregation layer of the enterprise network. The direct link between R1 and R2 belongs to area 0. R1 has a directly connected network segment 10.0.1.1/32 in area 0.
  - By default, the route from R4 to 10.0.1.1/32 has two next hops.
  - After the cost is changed, the route from R4 to 10.0.1.1 has only one next hop.

Three OSPF Tables: OSPF Neighbor Table

OSPF has three important tables: OSPF neighbor table, LSDB, and OSPF routing table. Pay attention to the following information about the OSPF neighbor table:

- Before OSPF transmits link state information, OSPF neighbor relationships must be established.
- OSPF neighbor relationships are established by exchanging Hello packets.
- The OSPF neighbor table displays the status of the neighbor relationship between OSPF routers. You can run the **display ospf peer** command to view the status.

[R1]display ospf peer



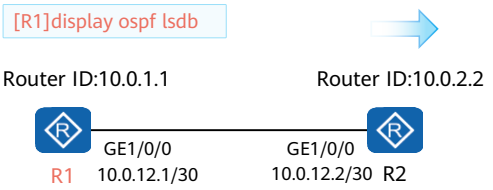

```
<R1> display ospf peer
    OSPF Process 1 with Router ID 10.0.1.1
                Neighbors
Area 0.0.0.0 interface 10.0.12.1(GigabitEthernet1/0/0)'s neighbors
Router ID: 10.0.2.2     Address: 10.0.12.2     GR State: Normal
    State: Full  Mode:Nbr is  Master  Priority: 1
    DR: 10.0.12.1  BDR: 10.0.12.2   MTU: 0
    Dead timer due in 35  sec
    Retrans timer interval: 5
    Neighbor is up for 00:00:05
    Authentication Sequence: [ 0 ]
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Three OSPF Tables: LSDB

Pay attention to the following information about the LSDB:

- The LSDB stores the LSAs generated by R1 and received from its neighbors. In this example, the LSDB of R1 contains three LSAs.
- Type indicates the LSA type, and AdvRouter indicates the router that sends the LSA.
- You can run the **display ospf lsdb** command to check the LSDB.



<R1> display ospf lsdb

OSPF Process 1 with Router ID 10.0.1.1

Link State Database

Area: 0.0.0.0

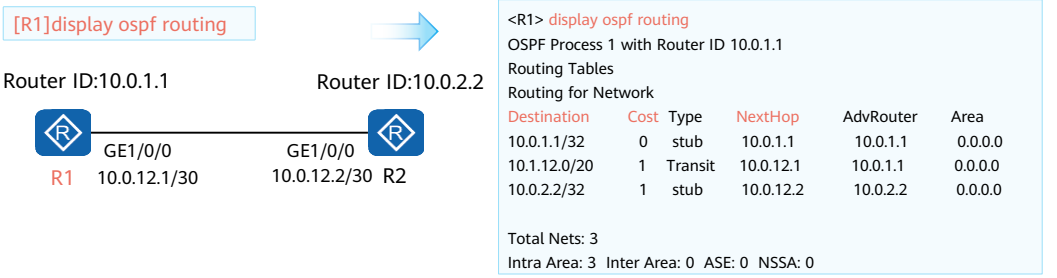
Type	LinkStateID	AdvRouter	Age	Len	Sequence	Metric
Router	10.0.2.2	10.0.2.2	98	36	8000000B	1
Router	10.0.1.1	10.0.1.1	92	36	80000005	1
Network	10.0.12.2	10.0.2.2	98	32	80000004	0

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Three OSPF Tables: OSPF Routing Table

Pay attention to the following information about the OSPF routing table:

- The OSPF routing table and the router routing table are different. In this example, the OSPF routing table contains three routes.
- An OSPF routing table contains information that guides packet forwarding, for example, destination, cost, and next hop.
- You can run the **display ospf routing** command to check the OSPF routing table.



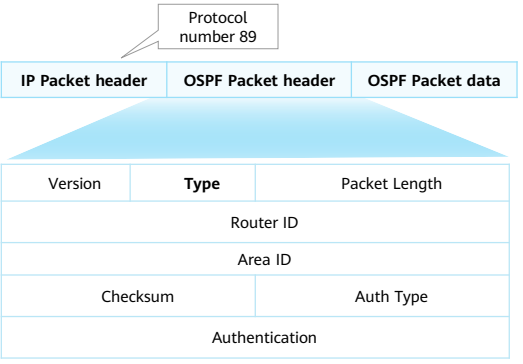
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The routing table of a router is called a global routing table. Not all OSPF routes can be added to the routing table of a router.

OSPF Packet Format and Type

- OSPF defines five types of packets. Different types of OSPF packets have the same header format.
- OSPF packets are encapsulated in IP packets. The protocol number in the IP header of OSPF packets is 89.

Type	Packet Name	Function
1	Hello	Discovers and maintains neighbor relationships.
2	Database Description	Exchanges brief LSDB information.
3	Link State Request	Requests specific link state information.
4	Link State Update	Sends detailed link state information.
5	Link State Ack	Acknowledges LSAs.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

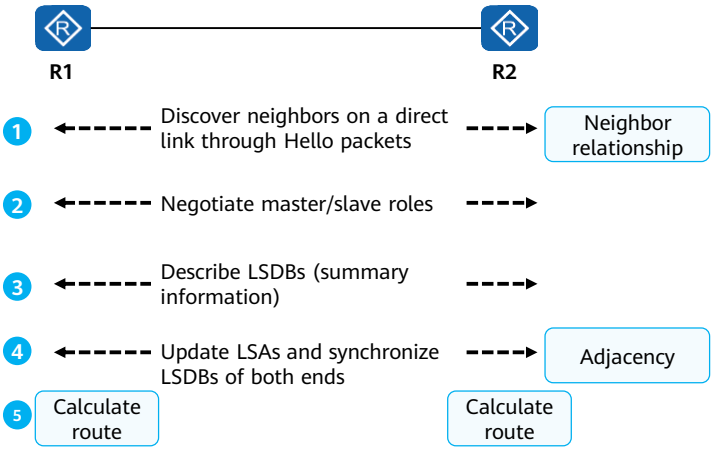
- Key fields:
  - Version: indicates the OSPF version. The value of this field is 2 for OSPFv2.
  - Router ID: indicates the router ID of the router that generates the packet.
  - Area ID: indicates the ID of the area to which the packet is advertised.
  - Type: indicates the packet type.
  - Packet length: indicates the length of an OSPF packet, in bytes.
  - Checksum: indicates the LSA checksum. It is used to check the entire OSPF packet, including the OSPF packet header.
  - Auth Type:
    - 0: no authentication
    - 1: plain-text password authentication
    - 2: cipher-text (MD5) authentication
  - Authentication: indicates information required for authentication. The value of this field varies according to the value of AuType.

Contents

1. Introduction to Dynamic Routing Protocols
2. Overview of OSPF
3. **OSPF Working Mechanism**
   - **Neighbor Relationship Establishment**
   - Adjacency Establishment
   - Functions of the DR and BDR
4. Basic OSPF Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Summary of OSPF Working Mechanism

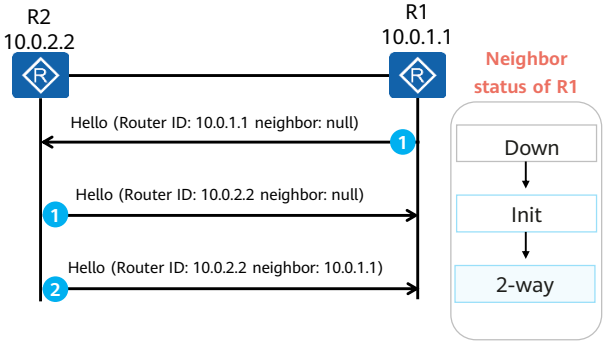


Steps 1 to 4 are performed through interaction between the two parties, and step 5 is performed independently.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Neighbor Relationship Establishment

- OSPF uses Hello packets to discover and establish neighbor relationships.
- On an Ethernet link, by default, OSPF sends Hello packets in multicast mode (destination address: 224.0.0.5).
- An OSPF Hello packet contains information such as the router ID and neighbor list of a router.



Down: Initial state of a neighbor, which indicates that no packets are received from the neighbor.

Init: The router has received a Hello packet from its neighbor, but its router ID is not in the neighbor list of the received Hello packet.

2-way: The router finds that its router ID exists in the neighbor list of the received Hello packet.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- R1 and R2 send Hello packets to each other. The first Hello packet contains an empty neighbor list.
- After R2 receives the Hello packet from R1, R2 finds that parameters in the Hello packet match those configured on R2. Then R2 adds R1 to its neighbor list when sending a Hello packet again.
- On an Ethernet link, Hello packets are transmitted in multicast mode.
  - The value 224.0.0.5 is the reserved IP multicast address of the OSPF device.
  - The value 224.0.0.6 is the IP multicast address reserved for Designated Routers (DRs) or Backup Designated Routers (BDRs).
- For links that do not support multicast, OSPF can send Hello packets in unicast mode.

Hello Packet

- Hello packets are used in the following scenarios:
  - Neighbor discovery: Hello packets are used to automatically discover neighboring routers.
  - Neighbor relationship establishment: The two ends negotiate parameters in Hello packets and establish a neighbor relationship.
  - Neighbor relationship holding: A router periodically sends and receives Hello packets to detect the operating status of neighbors.

Network Mask		
Hello Interval	***Options***	Router Priority
RouterDeadInterval		
Designated Router		
Backup Designated Router		
Neighbor		
...		

- **Key fields**
  - **Network Mask**: indicates the network mask of the interface that sends Hello packets.
  - **HelloInterval**: indicates the interval at which Hello packets are sent. The value is 10s typically.
  - **RouterDeadInterval**: indicates the expiration time of a neighbor relationship. If a device does not receive any Hello packets from its neighbors within a specified Dead interval, the neighbors are considered to be Down. The value is 40s typically.
  - **Neighbor**: indicates the router ID of a neighbor.
- **Description of other fields**
  - **Options:**
    - E: indicates whether external routes are supported.
    - MC: indicates whether to support forwarding of multicast data packets.
    - N/P: indicates whether the area is an NSSA.
  - **Router Priority**: indicates the DR priority. The default value is 1. If it is set to 0, the router cannot participate in DR or BDR election.
  - **Designated Router**: indicates the interface address of a DR.
  - **Backup Designated Router**: indicates the interface address of a BDR.

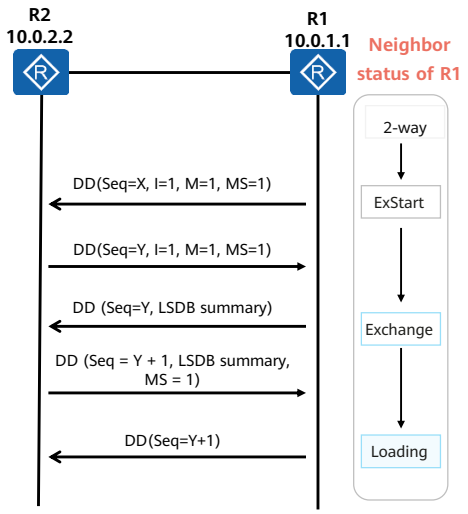
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Introduction to Dynamic Routing Protocols
2. Overview of OSPF
3. **OSPF Working Mechanism**
   - Neighbor Relationship Establishment
   - **Adjacency Establishment**
   - Functions of the DR and BDR
4. Basic OSPF Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Adjacency Establishment (1)



ExStart: The router starts to send DD packets to its neighbor. The DD packets sent in this state do not contain the link state description.

Exchange: A router and its neighbor exchange DD packets that contain link state summaries.

Loading: A router and its neighbor send LSR packets, LSU packets, and LSAck packets to each other.

Fields in DD packets

- **I:** If the DD packet is the first among multiple consecutive DD packets sent by a device, this field is set to 1. Otherwise, this field is set to 0.
- **M (More):** If the DD packet is the last among multiple consecutive DD packets sent by a device, this field is set to 0. Otherwise, this field is set to 1.
- **MS (Master/Slave):** When two OSPF routers exchange DD packets, they need to determine the master/slave relationship. The router with a larger router ID becomes the master router. The value 1 indicates that the sender is the master.
- **DD sequence number:** indicates the sequence number of a DD packet. The master and slave devices use sequence numbers to ensure the reliability and integrity of DD packet transmission.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The router IDs of R1 and R2 are 10.0.1.1 and 10.0.2.2 respectively, and the neighbor relationship has been established between R1 and R2. When the neighbor status of R1 changes to ExStart, R1 sends the first DD packet. In the DD packet, if the M-bit is set to 1, subsequent DD packets need to be sent. If the MS-bit is set to 1, R1 declares itself as the master router. The DD sequence number is randomly set to X, and the I-bit is set to 1, indicating that this is the first DD packet.
- Similarly, when the neighbor status of R2 changes to ExStart, R2 also sends the first DD packet. In this packet, the DD sequence number is randomly set to Y (I-bit=1, M-bit=1, MS-bit=1, and the meaning is the same as above). Because R2 has a larger router ID, R2 becomes the master router. After R1 receives this packet, R1 generates a Negotiation-Done event and changes the neighbor status from ExStart to Exchange.
- When the neighbor status of R1 changes to Exchange, R1 sends a new DD packet carrying the LSDB summary. The sequence number is set to the sequence number Y used by R2 in step 2. The I-bit is 0, indicating that R1 is not the first DD packet. The M-bit is 0, indicating that the packet is the last DD packet carrying the LSDB summary. The MS-bit is 0, indicating that R1 declares itself as the slave device. After R2 receives the packet, R2 changes the neighbor status from ExStart to Exchange.
- When the neighbor status of R2 changes to Exchange, R2 sends a new DD packet that contains the summary of the LSDB. The DD sequence number is set to (Y+1) and the MS-bit is set to 1, indicating that R2 declares itself as the master router.
- Although R1 does not need to send a new DD packet that contains the LSDB summary, R1 needs to acknowledge each DD packet sent by the master router. Therefore, R1 sends a new DD packet with the sequence number (Y+1) to R2. The packet is empty. After the packet is sent, R1 generates an Exchange-Done event and changes the neighbor status to Loading. After receiving the packet, R2 changes the neighbor status to Full. (Assume that the LSDB of R2 is the latest and does not need to request R1 to update the LSDB.)

DD Packet

A DD packet contains the LSA header information, including the LS type, LS ID, Advertising Router, LS Sequence Number, and LS Checksum.

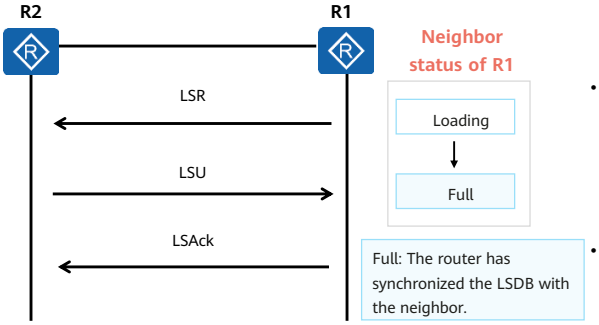
Interface MTU	Options	0 0 0 0 0 I M MS
DD sequence number		
LSA Header		

Description of other fields

- Interface MTU: indicates the maximum size of an IP packet that an interface can send without fragmenting the packet. The DD packets sent by two neighbors contain the MTU. If the MTU in the received DD packet is different from the local MTU, the DD packet is discarded. By default, MTU check is disabled on a Huawei device.
- Options: The field is the same as that in a Hello packet.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Adjacency Establishment (2)



- R1 starts to send LSR packets to R2 to request the link state information that is discovered through DD packets in Exchange state and does not exist in the local LSDB.
- R2 sends an LSU packet to R1. The LSU packet contains detailed information about the requested link state. After R1 receives the LSU packet and does not have other LSAs to be requested, R1 changes the neighbor status from Loading to Full.
- R1 sends an LSAck packet to R2 to acknowledge the LSU packet.

Question: If multiple routers are located on the same broadcast network, what are the problems in establishing adjacencies using the preceding method?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Introduction to Dynamic Routing Protocols
2. Overview of OSPF
3. **OSPF Working Mechanism**
   - Neighbor Relationship Establishment
   - Adjacency Establishment
   - **Functions of the DR and BDR**
4. Basic OSPF Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Functions of the DR and BDR

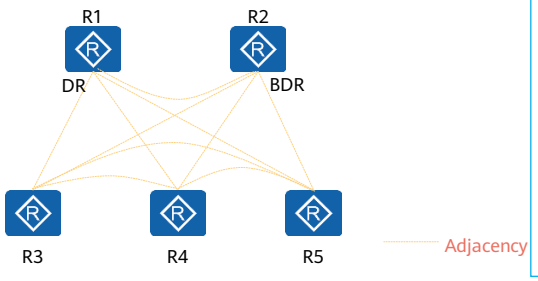
Problems on the MA Network

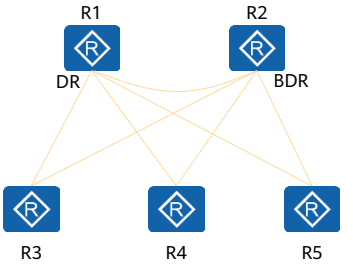
- n x (n-1)/2 adjacencies complicate management.
- Repeated LSA flooding wastes resources.

Solution

DR election on an MA network:

- A DR establishes and maintains adjacencies on an MA network and synchronizes LSAs.
- The DR establishes adjacencies with all other routers and exchanges link state information with them. Other routers do not directly exchange link state information.
- To prevent single points of failures (SPOFs), a BDR is elected to quickly take over services of the DR when the DR fails.



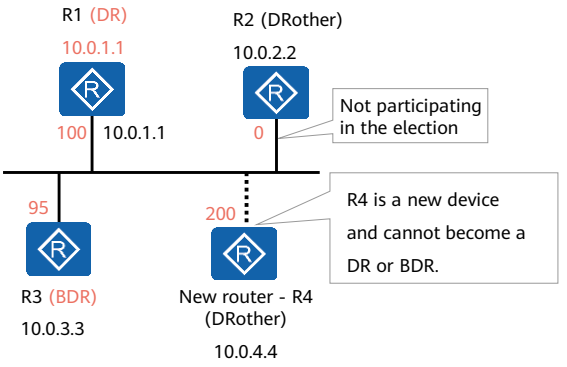


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Multiple access (MA) is classified into broadcast multi-access (BMA) and non-broadcast multiple access (NBMA). The network formed by Ethernet links is a typical BMA network. FR links are logically divided to form a typical NBMA network. (Note: FR-related information is not described here.)
- DRother: A router that is not a DR or a BDR is a DRother.

DR and BDR Election Rules

- DR or BDR election is in non-preemption mode.
- DR or BDR election is based on interfaces.
  - The greater the DR priority of an interface, the higher the priority.
  - If the DR priorities of interfaces are the same, the interface with a larger router ID is preferred.



- Question:
  - If the priorities of the four routers in the preceding figure are all set to 0, can OSPF work normally?
  - Which types of links form an MA network by default?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The DR and BDR election process on a broadcast or NBMA link is as follows:
  - When going Up, an interface sends Hello packets and enters the waiting state. In Waiting state, there is a WaitingTimer that has the same value as the DeadTimer. By default, the waiting timer duration is 40 seconds, which cannot be changed.
  - Before the waiting timer is triggered, sent Hello packets carry no DR or BDR field. During the waiting period, if the received Hello packets contain the DR or BDR field, the election is not triggered and routers directly leave the waiting state to start neighbor synchronization.
  - Assume that a DR and a BDR exist on the network. Any router newly connected to the network will accept the DR and BDR that exist on the network regardless of its router ID or DR priority.
  - If the DR fails and goes Down, the BDR takes over the role of the DR and the remaining devices whose priority is greater than 0 compete to become the new BDR.
  - The DR is elected based on the rules only when routers with different router IDs or DR priorities become Up and perform DR election simultaneously.

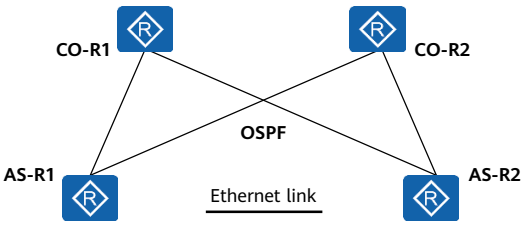
DR and BDR Election on Different Types of Networks

OSPF Network Type	Common Data Link Layer Protocol	Whether to Elect a DR	Whether to Establish an Adjacency with the Neighbor
Point-to-point	PPP and HDLC	No	Yes
Broadcast	Ethernet	Yes	The DR establishes adjacencies with the BDR and DRothers. The BDR establishes adjacencies with the DR and DRothers. The DRothers establish a neighbor relationship.
NBMA	FR		
P2MP	Manually specified protocol	No	Yes

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- P2MP: point-to-multipoint

Adjusting the OSPF Network Type of Device Interfaces as Needed



- The OSPF network type is automatically set based on the data link layer encapsulation of the interface.
- Routers in the figure are interconnected through Ethernet interfaces, so the network type of these interfaces is broadcast by default.
- Each link is a point-to-point (P2P) link, so it is unnecessary to elect the DR and BDR on a link.
- To improve OSPF efficiency and speed up the establishment of neighbor relationships, you can change the network type of these interconnected interfaces to P2P.

In the interface view, run the **ospf network** { **p2p** | **p2mp** | **broadcast** | **nbma** } command to change the network type of the interface.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The OSPF network type is automatically set according to the data link layer encapsulation of an interface. As shown in the figure, two routers AS-R1 and AS-R2 at the network access layer are connected to the routers CO-R1 and CO-R2 at the core layer through two links. OSPF is enabled on the interfaces of the four routers. These routers are interconnected through Ethernet interfaces, so the network type of these interfaces is broadcast by default. During the establishment of neighbor relationships, OSPF elects the DR and BDR on each Ethernet link.
- However, this is unnecessary and time-consuming (the DR and BDR election process involves a waiting timer, which increases the time for directly connected routers to establish adjacencies). These links are P2P connections logically, so DR or BDR election is unnecessary. To improve OSPF efficiency and speed up the establishment of neighbor relationships, you can change the network type of these interconnected interfaces to P2P.

Contents

1. Introduction to Dynamic Routing Protocols
2. Overview of OSPF
3. OSPF Working Mechanism
4. **Basic OSPF Configurations**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Commands (1)

1. Create an OSPF process and enter the OSPF view.

   [Huawei] **ospf** [*process-id* | **Router ID** *Router ID*]

   The router supports OSPF multi-process, and the process ID is configured locally. Two devices that use different OSPF process IDs can also establish an adjacency.

2. Create an OSPF area and enter the OSPF area view.

   [Huawei-ospf-1] **area** *area-id*

3. Enable OSPF in the OSPF area.

   [Huawei-ospf-1-area-0.0.0.0] **network** *network-address wildcard-mask*

   Run the following command to configure the network segment included in the area. The mask length of the interface IP address is larger than or equal to the mask length specified by the **network** command, and the primary IP address of the interface must be on the network segment specified by the **network** command. In this case, OSPF can be activated in the corresponding area on the interface.

4. Enable OSPF in the interface view.

   [Huawei-GigabitEthernet1/0/0] **ospf enable** *process-id* **area** *area-id*

   The **ospf enable** command takes precedence over the **network** command.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- By default, OSPF advertises the IP address of a loopback interface as a 32-bit host route, which is independent of the mask length configured on the loopback interface. Therefore, to configure OSPF to advertise the actual network segment of a loopback interface, set the network type of the interface to NBMA or broadcast on the loopback interface.

Configuration Commands (2)

5. Set a priority for an interface that participates in the DR election in the interface view.

```
[Huawei-GigabitEthernet1/0/0] ospf dr-priority priority
```

By default, the priority is 1.

6. Set the interval for sending Hello packets on an interface.

```
[Huawei-GigabitEthernet1/0/0] ospf timer Hello interval
```

By default, for a P2P or broadcast interface, the interval for sending Hello packets is 10 seconds; the dead interval after which an interface considers its OSPF neighbor invalid is four times the interval for sending Hello packets.

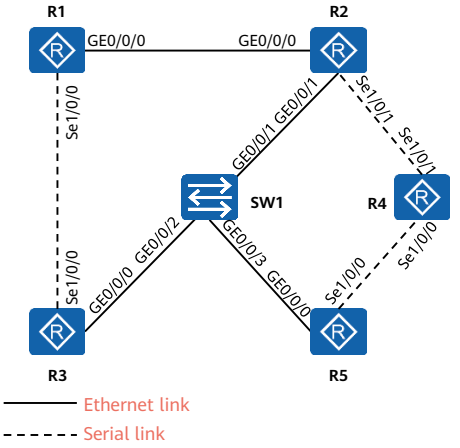
7. Set a network type for an OSPF interface.

```
[Huawei-GigabitEthernet1/0/0] ospf network-type { broadcast | nbma | p2mp | p2p }
```

By default, the network type of an interface is determined by the physical interface. The network type of an Ethernet interface is broadcast, and the network type of a serial interface or a POS interface (PPP or HDLC is used) is P2P.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

OSPF Configuration Examples



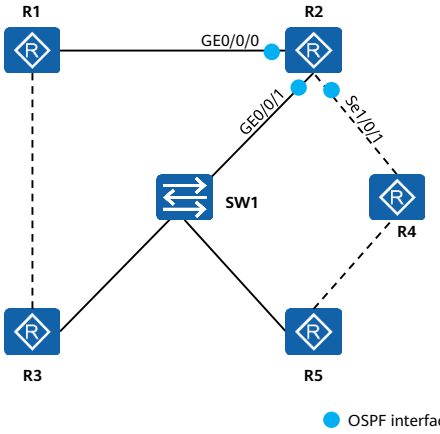
- Basic information: The router ID of each device is 10.0.x.x, where x is the router number. For example, the router ID of R5 is 10.0.5.5. The IP address for interconnection between devices is 10.0.xyz.x(y)/24, where *xyz* indicate the router numbers. The router numbers are in ascending order. For example, the IP address of GE0/0/1 on R2 is 10.0.235.2/24.
- Topology: Five routers work in area 0.

The configuration on R2 is used as an example.

```
[R2]ospf 1 router-id 10.0.2.2
[R2-ospf-1]area 0.0.0.0
[R2-ospf-1-area-0.0.0.0] network 10.0.12.0 0.0.0.255
[R2-ospf-1-area-0.0.0.0] network 10.0.24.2 0.0.0.0
[R2-ospf-1-area-0.0.0.0] network 10.0.35.2 0.0.0.0
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

OSPF Configuration Verification (1)



Run the **display ospf interface all** command to check information about all OSPF interfaces on the device.

- Time parameters, such as the interval for sending Hello packets and dead interval
- Link type and MTU of the interface
- Interface IP address of the DR and the priority of the DR for an Ethernet link

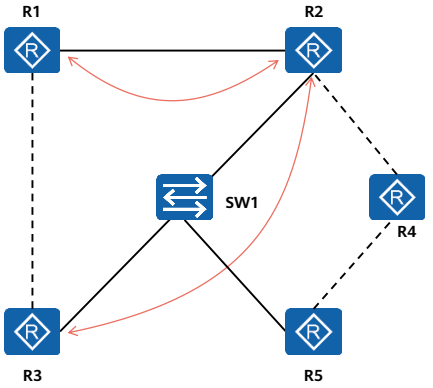
```
[R2]display ospf interface all
                OSPF Process 1 with Router ID 10.0.2.2
 Area: 0.0.0.0
 Interface: 10.0.12.2 (GigabitEthernet0/0/0)
 Cost: 1       State: DR       Type: Broadcast    MTU: 1500   Priority: 1
 Designated Router: 10.0.12.2
 Backup Designated Router: 10.0.12.1
 Timers: HELLO 10 , Dead 40 , Poll  120 , Retransmit 5 , Transmit Delay 1

 Interface: 10.0.235.2 (GigabitEthernet0/0/1)
 Cost: 1       State: DROther Type: Broadcast   MTU: 1500   Priority: 1
 Designated Router: 10.0.235.5
 Backup Designated Router: 10.0.235.3
 Timers: HELLO 10 , Dead 40 , Poll  120 , Retransmit 5 , Transmit Delay 1

 Interface: 10.0.24.2 (Serial1/0/1) --> 10.0.24.4
 Cost: 48      State: P-2-P      Type: P2P        MTU: 1500
 Timers: HELLO 10 , Dead 40 , Poll  120 , Retransmit 5 , Transmit Delay 1
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

OSPF Configuration Verification (2)



Run the **display ospf peer** command to check the neighbor status of the device.

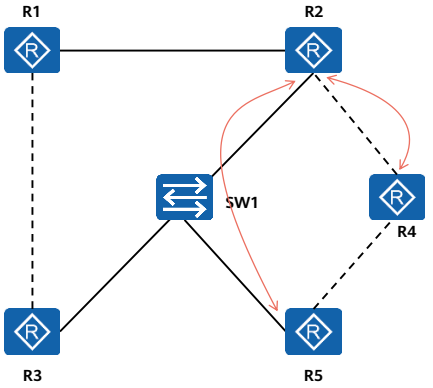
- Router ID of the neighboring router
- Neighbor status, such as FULL, TWO-WAY, and DOWN

```
<R2>display ospf peer
                OSPF Process 1 with Router ID 10.0.2.2
 Area 0.0.0.0 interface 10.0.12.2(GigabitEthernet0/0/0)'s neighbors
 Router ID: 10.0.1.1          Address: 10.0.12.1
   State: Full  Mode:Nbr is  Slave  Priority: 1
   DR: 10.0.12.2  BDR: 10.0.12.1  MTU: 0
   Dead timer due in 28  sec
   Retrans timer interval: 5
   Neighbor is up for 00:01:31
   Authentication Sequence: [ 0 ]

 Area 0.0.0.0 interface 10.0.235.2(GigabitEthernet0/0/1)'s neighbors
 Router ID: 10.0.3.3          Address: 10.0.235.3
   State: Full  Mode:Nbr is  Master  Priority: 1
   DR: 10.0.235.5  BDR: 10.0.235.3  MTU: 0
   Dead timer due in 30  sec
   Retrans timer interval: 5
   Neighbor is up for 00:01:31
   Authentication Sequence: [ 0 ]
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

OSPF Configuration Verification (3)



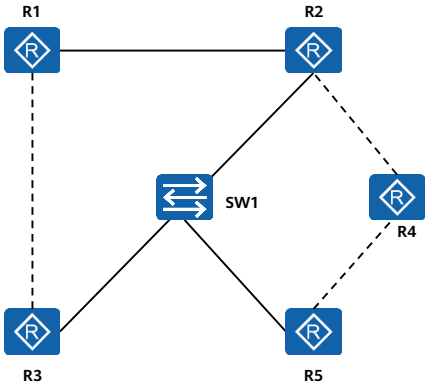
On a P2P network, DR or BDR election is not required. Therefore, when checking the OSPF neighbor table of R2, you can find that the DR/BDR field of Serial1/0/1 in the command output is **None**.

```
<R2>display ospf peer
                OSPF Process 1 with Router ID 10.0.2.2
Area 0.0.0.0 interface 10.0.235.2(GigabitEthernet0/0/1)'s neighbors
Router ID: 10.0.5.5          Address: 10.0.235.5
  State: Full  Mode:Nbr is  Master  Priority: 1
  DR: 10.0.235.5  BDR: 10.0.235.3  MTU: 0
  Dead timer due in 40  sec
  Retrans timer interval: 0
  Neighbor is up for 00:01:27
  Authentication Sequence: [ 0 ]

Area 0.0.0.0 interface 10.0.24.2(Serial1/0/1)'s neighbors
Router ID: 10.0.4.4          Address: 10.0.24.4
  State: Full  Mode:Nbr is  Master  Priority: 1
  DR: None   BDR: None   MTU: 0
  Dead timer due in 35  sec
  Retrans timer interval: 5
  Neighbor is up for 00:01:56
  Authentication Sequence: [ 0 ]
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

OSPF Configuration Verification (4)



Run the **display ospf lsdb** command to check the LSDB of the device.

- An LSDB consists of multiple types of LSAs. All LSAs have the same packet header format, in which key fields such as Type, LinkState ID, and AdvRouter are included. The next course will focus on LSA details.

<R2>display ospf lsdb

OSPF Process 1 with Router ID 10.0.2.2
Link State Database

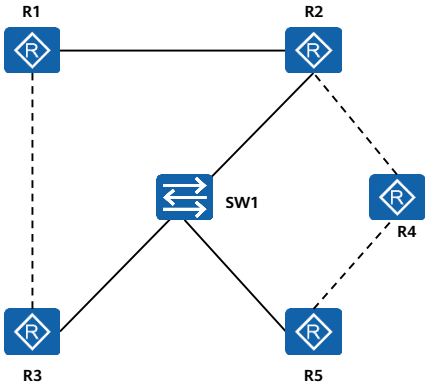
Area: 0.0.0.0

Type	LinkState ID	AdvRouter	Age	Len	Sequence	Metric
Router	10.0.4.4	10.0.4.4	662	72	80000006	48
Router	10.0.2.2	10.0.2.2	625	72	8000000C	1
Router	10.0.1.1	10.0.1.1	638	60	80000007	1
Router	10.0.5.5	10.0.5.5	634	60	8000000B	1
Router	10.0.3.3	10.0.3.3	639	60	80000009	1
Network	10.0.235.5	10.0.5.5	634	36	80000005	0
Network	10.0.12.2	10.0.2.2	629	32	80000003	0

Are LSDBs on other devices the same?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

OSPF Configuration Verification (5)



Run the **display ospf routing** command to check the OSPF routing table of the device.

- The OSPF routing table of R2 shows that R2 has learned the routes to the entire network through OSPF.

```
<R2>display ospf routing

        OSPF Process 1 with Router ID 10.0.2.2
                Routing Tables
```

Destination	Cost	Type	NextHop	AdvRouter	Area
10.0.12.0/24	1	Transit	10.0.12.2	10.0.2.2	0.0.0.0
10.0.24.0/24	48	Stub	10.0.24.2	10.0.2.2	0.0.0.0
10.0.235.0/24	1	Transit	10.0.235.2	10.0.2.2	0.0.0.0
10.0.13.0/24	49	Stub	10.0.12.1	10.0.1.1	0.0.0.0
10.0.13.0/24	49	Stub	10.0.235.3	10.0.3.3	0.0.0.0
10.0.45.0/24	49	Stub	10.0.235.5	10.0.5.5	0.0.0.0

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Single) Which of the following packets is used by OSPF to maintain neighbor relationships? ( )
   A. Hello
   B. Database Description
   C. LSR
   D. LSU
2. (Multiple) Which of the following network types are supported by OSPF? ( )
   A. P2P network
   B. P2MP network
   C. Broadcast network
   D. NBMA network

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. A
2. ABCD

Summary

- This course describes basic OSPF concepts, including the router ID, area, and cost. Routers running OSPF send link state information to each other to calculate the topology and routes.
- This course describes the process of establishing OSPF neighbor relationships and adjacencies. On an MA network, the DR and BDR need to be elected. There are five types of OSPF packets. All packets have the same packet header format. An OSPF router periodically sends Hello packets to discover and maintain neighbor relationships, and uses DD, LSR, LSU, and LSAck packets to synchronize LSDBs. Finally, this course introduces the simple configuration of a single OSPF area.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



OSPF Route Calculation

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- Open Shortest Path First (OSPF) routers in the same area have the same link state database (LSDB). In an area, OSPF uses the shortest path first (SPF) algorithm to calculate routes.
- As the network scale expands, routers consume more memory and CPU resources to calculate routes. Area partitioning can relieve the pressure on routers to some extent.
- On a large-scale network, there may be other routing protocols besides OSPF. OSPF can import external routes so that OSPF routers can know the routes to other areas.
- This course describes the principles of OSPF route calculation, including the calculation of intra-area routes, inter-area routes, and external routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Explain functions of key fields in link-state advertisements (LSAs).
  - Describe common LSA types and functions.
  - Illustrate the SPF algorithm.
  - Describe the principles for calculating intra-area and inter-area OSPF routes.
  - Describe the loop prevention mechanism of OSPF inter-area routes.
  - Describe the principles for calculating OSPF external routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

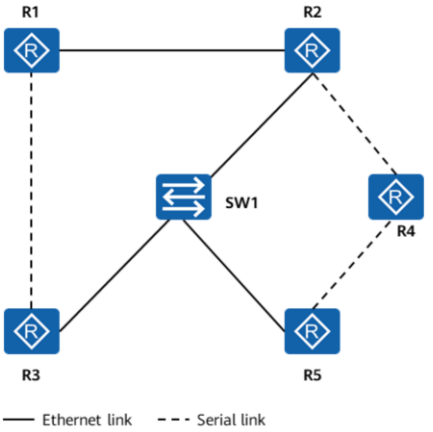
Contents

1. **Intra-Area Route Calculation**
   - **Introduction to LSAs**
   - Router-LSA
   - Network-LSA
   - SPF Calculation Process
2. Inter-Area Route Calculation
3. External Route Calculation

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

OSPF Review

All routers in the same area have the same LSDB. According to the command output of R2, multiple LSAs exist on the network.



<R2>display ospf lsdb

OSPF Process 1 with Router ID 10.0.2.2
Link State Database

Area: 0.0.0.0

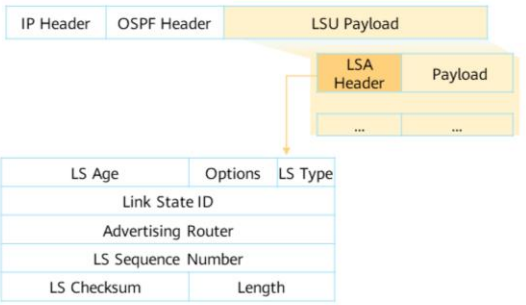
Type	LinkState ID	AdvRouter	Age	Len	Sequence	Metric
Router	10.0.4.4	10.0.4.4	662	72	80000006	48
Router	10.0.2.2	10.0.2.2	625	72	8000000C	1
Router	10.0.1.1	10.0.1.1	638	60	80000007	1
Router	10.0.5.5	10.0.5.5	634	60	8000000B	1
Router	10.0.3.3	10.0.3.3	639	60	80000009	1
Network	10.0.235.5	10.0.5.5	634	36	80000005	0
Network	10.0.12.2	10.0.2.2	629	32	80000003	0

What information does each LSA contain? Based on the preceding information, how does a router calculate routes?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Concepts of LSAs

- LSAs are the basis for OSPF to calculate routes.
- An OSPF Link State Update (LSU) packet can carry multiple types of LSAs.
- LSAs of different types have the same packet header.



Key Fields

- **LS Age**: indicates the lifetime of an LSA, in seconds.
- **Options**: Each bit corresponds to a feature supported by OSPF.
- **LS Type**: indicates the type of the local LSA.
- **Link State ID**: indicates the link state ID in the LSA header. This field varies with LSAs.
- **Advertising Router**: indicates the router ID of the device that generates the LSA.
- **LS Sequence Number**: indicates the sequence number in the LSA header. The value increases each time a new instance is generated.
- **LS Checksum**: is used to ensure data integrity and accuracy.
- **Length**: indicates the length of an LSA, including the length of the LSA header.

- **LS Type**, **Link State ID**, and **Advertising Router** uniquely identify an LSA.
- **LS Age**, **LS Sequence Number**, and **LS Checksum** are used to determine whether an LSA is old or new.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- LS Age: When an LSA is originated, the value of this field is 0. The value of this field increases as the LSA is flooded on the network. When the value of this field reaches the value of MaxAge (3600s by default), the LSA is not used for route calculation.
- LS Sequence Number: This field is used to determine whether an LSA is old or new or whether there are duplicate instances. The sequence number ranges from 0x80000001 to 0x7FFFFFFF. A router originates an LSA with the sequence number 0x80000001. The sequence number increases by 1 each time the LSA is updated. When the sequence number of the LSA reaches the maximum value, the LSA is regenerated and the sequence number is set to 0x80000001.

Common LSA Types

Type	Name	Description
1	Router-LSA	Every router on an OSPF network generates Router-LSAs. A Router-LSA describes a router's link state and cost and can be flooded only in the area to which the interface belongs.
2	Network-LSA	A designated router (DR) generates Network-LSAs. A Network-LSA describes all the routers that establish adjacencies with the DR on the MA network to which the DR is connected and the DR itself. The LSA can be flooded only in the area to which the interface belongs.
3	Network-summary-LSA	An area border router (ABR) generates Network-summary-LSAs. A Network-summary-LSA describes the route to the destination network segment of an area. It is used to transmit inter-area routes.
4	ASBR-summary-LSA	An ABR generates ASBR-summary-LSAs. An ASBR-summary-LSA describes routes to an ASBR, and is equivalent to a host route to an autonomous system border router (ASBR).
5	AS-external-LSA	An ASBR generates AS-external-LSAs. An AS-external-LSA describes routes to destinations outside an AS.
7	NSSA LSA	An ASBR generates NSSA LSAs. An NSSA LSA describes routes to destinations outside an AS. NSSA LSAs have similar functions as AS-external-LSAs, but are flooded in different areas. NSSA LSAs can be flooded only in the NSSA and cannot enter area 0. The ABR in the NSSA converts Type 7 LSAs into Type 5 LSAs and injects them into area 0.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In many cases, the type value is used to refer to the corresponding LSA. For example, Type 1 LSAs indicate Router-LSAs and Type 2 LSAs indicate Network-LSAs, and so on.

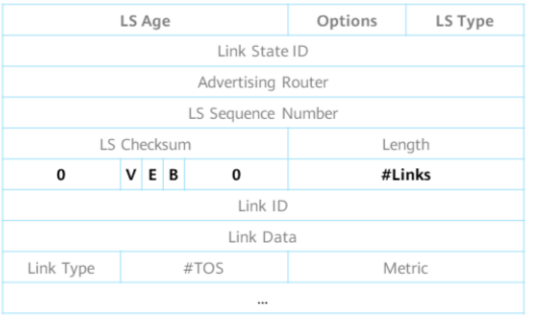
Contents

1. **Intra-Area Route Calculation**
   - Introduction to LSAs
   - **Router-LSA**
   - Network-LSA
   - SPF Calculation Process
2. Inter-Area Route Calculation
3. External Route Calculation

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Description of a Router-LSA (1)

- Router-LSA (Type 1 LSA): Every router on an OSPF network generates Router-LSAs. A Router-LSA describes a router's link state and cost.
- Router-LSAs can be flooded only in the area that the interface belongs to.



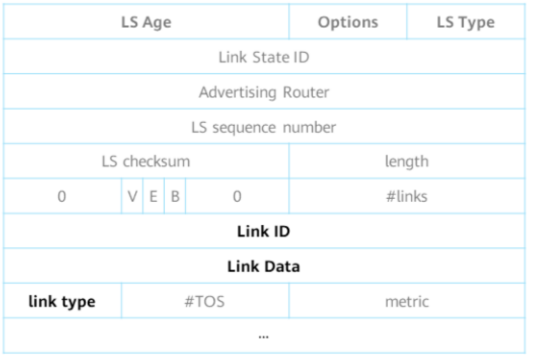
- V (virtual link): If the router that generates the LSA is the endpoint of a virtual link, this field is set to 1.
- E (external): If the router that generates the LSA is an ASBR, this field is set to 1.
- B (border): If the router that generates the LSA is an ABR, this field is set to 1.
- Links: indicates the number of links in the LSA. Router-LSAs use links to carry information about directly connected interfaces on routers.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Link State ID of Type 1 LSA is Router ID

Description of a Router-LSA (2)

- Router-LSAs use links to carry information about directly connected interfaces on routers.
- Each link contains the link type, link ID, link data, and metric.
- A router may use one or more links to describe an interface.

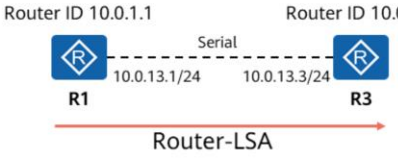


Link Type	Link ID	Link Data
Point-to-point (P2P): describes a P2P link between the local router and a neighboring router, which is included in the topology information.	ID of the neighboring router.	IP address of the router interface that advertises the Router-LSA.
TransNet: describes a connection from the local router to a transit network segment such as an MA or NBMA network segment, which is included in the topology information.	Interface IP address of the DR	IP address of the router interface that advertises the Router-LSA.
StubNet: describes a connection from the local router to a stub network segment, for example, a loopback interface, which is included in the network segment information.	Network IP address of the router interface that advertises the Router-LSA	Subnet mask of the stub network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Metric: indicates the cost.

Router-LSAs: P2P Networks



R1 sends a Router-LSA to R3. The Router-LSA carries topology and network segment information.

```
<R1>display ospf lsdb router self-originate
  Type      : Router
  Ls id     : 10.0.1.1
  Adv rtr   : 10.0.1.1
* Link ID: 10.0.3.3
  Data      : 10.0.13.1
  Link Type: P-2-P
  Metric    : 48
* Link ID: 10.0.13.0
  Data      : 255.255.255.0
  Link Type: StubNet
  Metric    : 48
  Priority  : Low
```

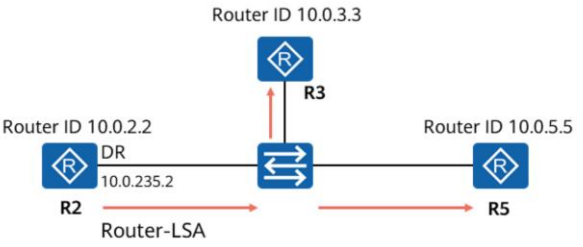
LSA header

First link in the Router-LSA: Describe the topology information.

Second link in the Router-LSA: Describe the network segment information.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Router-LSAs: TransNet Networks



R2 sends Router-LSAs carrying topology information to R3 and R5.

<R2>display ospf lsdb router self-originate

Type : Router
Ls id : 10.0.2.2
Adv rtr : 10.0.2.2

LSA header

* Link ID: 10.0.235.2
Data : 10.0.235.2
Link Type: TransNet
Metric : 1

Link contained in a Router-LSA: Describe topology information.

How is complete network segment information described on the TransNet network if there is no network mask information?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Intra-Area Route Calculation**
   - Introduction to LSAs
   - Router-LSA
   - **Network-LSA**
   - SPF Calculation Process
2. Inter-Area Route Calculation
3. External Route Calculation

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Description of a Network-LSA

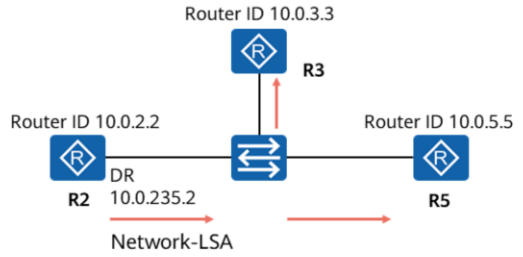
- Network-LSA (Type 2 LSA): is originated by a DR, describes the link state of the local network segment, and is advertised in an area to which the DR belongs.
- A Network-LSA records all OSPF routers that have established adjacencies with the DR on the network segment and carries the network mask of the network segment.

LS Age	Options	**LS Type**
Link State ID		
Advertising Router		
LS Sequence Number		
LS Checksum	Length	
Network Mask		
Attached Router		
...		

- **Link State ID**: indicates the IP address of an interface on the DR.
- **Network Mask**: indicates the subnet mask of the MA network.
- **Attached Router**: indicates the router ID of the device connected to the MA network. If multiple routers are connected to the MA network, multiple fields are used.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Network-LSAs Describe the MA Network



R2 sends Network-LSAs to R3 and R5, carrying topology and network segment information.

```
<R2>display ospf lsdb network self-originate

        OSPF Process 1 with Router ID 10.0.2.2
                Area: 0.0.0.0
            Link State Database

Type      : Network                          LSA header
Ls id     : 10.0.235.2
Adv rtr   : 10.0.2.2

                                             10.0.235.0/24

Net mask  : 255.255.255.0                    Topology
  Priority  : Low                            information
  Attached Router    10.0.2.2
  Attached Router    10.0.3.3
  Attached Router    10.0.5.5
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Intra-Area Route Calculation**
   - Introduction to LSAs
   - Router-LSA
   - Network-LSA
   - **SPF Calculation Process**
2. Inter-Area Route Calculation
3. External Route Calculation

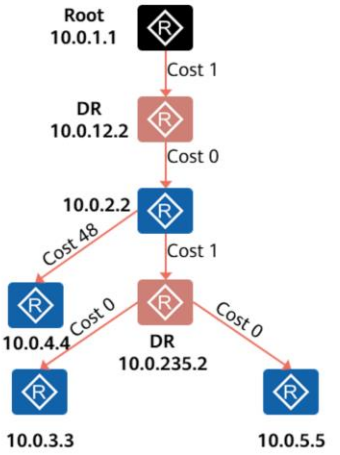
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

SPF Algorithm (1)

Phase 1: Construct an SPF tree.

- A router takes itself as the root of the SPF tree. Each time a new node is added, the router with the smallest cost is added to the SPF tree based on the topology information in the Router-LSA and Network-LSA. A router is identified by a router ID or DR ID.
- On a broadcast network, the cost between the DR and the router connected to it is 0.
- The SPF tree has only the unidirectional shortest path, which prevents routing loops in the OSPF area.

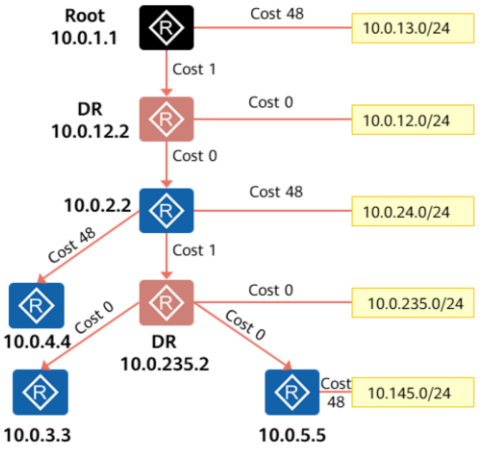


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SPF Algorithm (2)

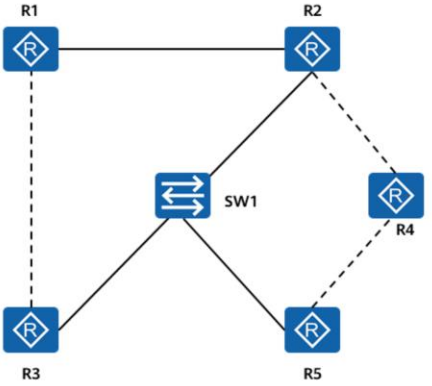
Phase 2: Calculate the optimal route.

- The routing information in the Router-LSA and Network-LSA is attached to the corresponding OSPF router to calculate the optimal route.
- The existing routing information is not added to the SPF tree.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

SPF Algorithm Example



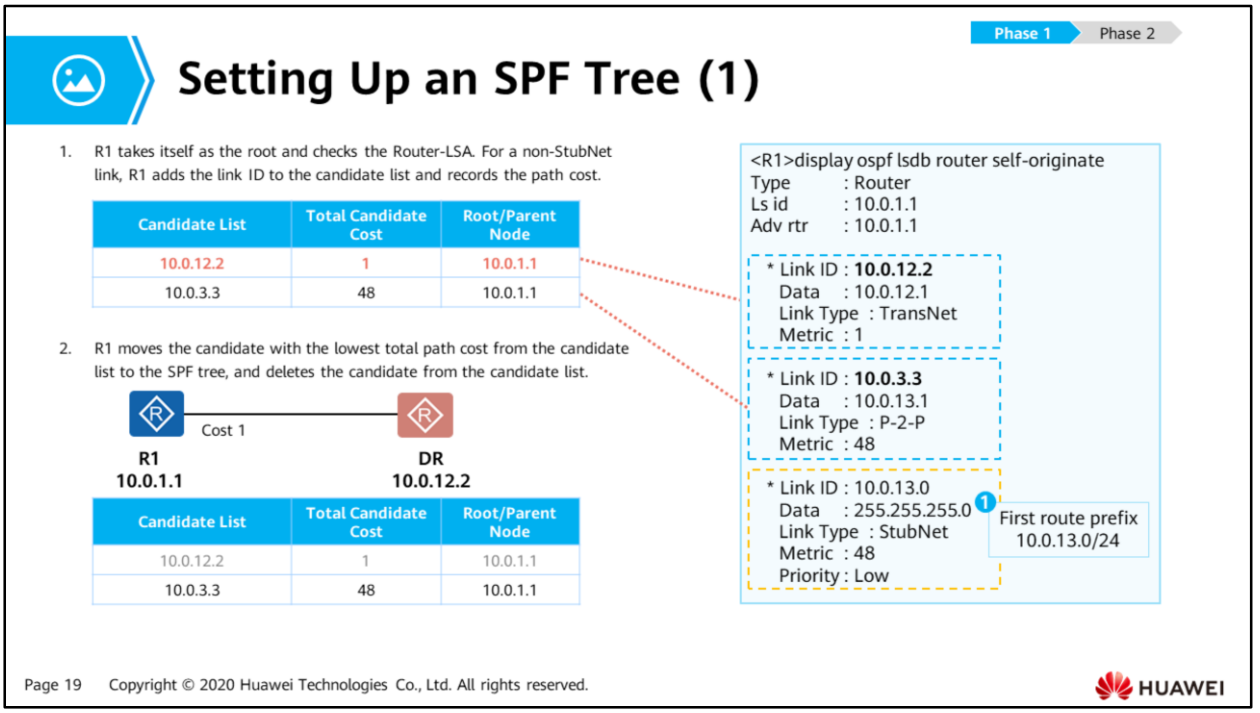
The following uses R1 as an example to describe the OSPF topology and route calculation process.

Neighbor to be ignored

Neighbor to be added

Route prefix

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



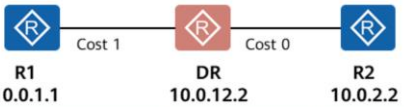
- Note:
  - The total candidate cost is the sum of the metric described in the LSA and the cost of the route from the parent node to the root node.
  - The candidate list records the neighbor list.

Setting Up an SPF Tree (2)

3. R1 continues to query the Network-LSAs generated by the DR and adds the topology information described in the Network-LSAs to the candidate list. If the node described in the LSAs already exists in the SPF tree, the node does not need to be added to the topology.

Candidate List	Total Candidate Cost	Root/Parent Node
10.0.3.3	48	10.0.1.1
10.0.2.2	1+0	10.0.12.2

4. R1 moves the candidate with the lowest total path cost from the candidate list to the SPF tree, and deletes the candidate from the candidate list.



Candidate List	Total Candidate Cost	Root/Parent Node
10.0.3.3	48	10.0.1.1
10.0.2.2	1+0	10.0.12.2

```
<R1>display ospf lsdb network 10.0.12.2

Type       : Network
Ls id      : 10.0.12.2
Adv rtr    : 10.0.2.2

Net mask   : 255.255.255.0
  Priority : Low
    Attached Router    10.0.2.2

    Attached Router    10.0.1.1
```

2 Second route prefix 10.0.12.0/24

The node does not need to be added to the topology because the node already exists in the SPF tree.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Setting Up an SPF Tree (3)

5. Based on Router-LSAs generated by R2, R1 records neighbor information in the candidate list.

Candidate List	Total Candidate Cost	Root/Parent Node
10.0.3.3	48	10.0.1.1
10.0.235.2	1+0+1	10.0.2.2
10.0.4.4	1+0+48	10.0.2.2

6. R1 moves the candidate with the lowest total path cost from the candidate list to the SPF tree, and deletes the candidate from the candidate list.



Candidate List	Total Candidate Cost	Root/Parent Node
10.0.3.3	48	10.0.1.1
10.0.235.2	1+0+1	10.0.2.2
10.0.4.4	1+0+48	10.0.2.2

```
<R1>display ospf lsdb router 10.0.2.2

Type      : Router
Ls id     : 10.0.2.2
Adv rtr   : 10.0.2.2

* Link ID: 10.0.12.2
  Data   : 10.0.12.2
  Link Type: TransNet
  Metric : 1

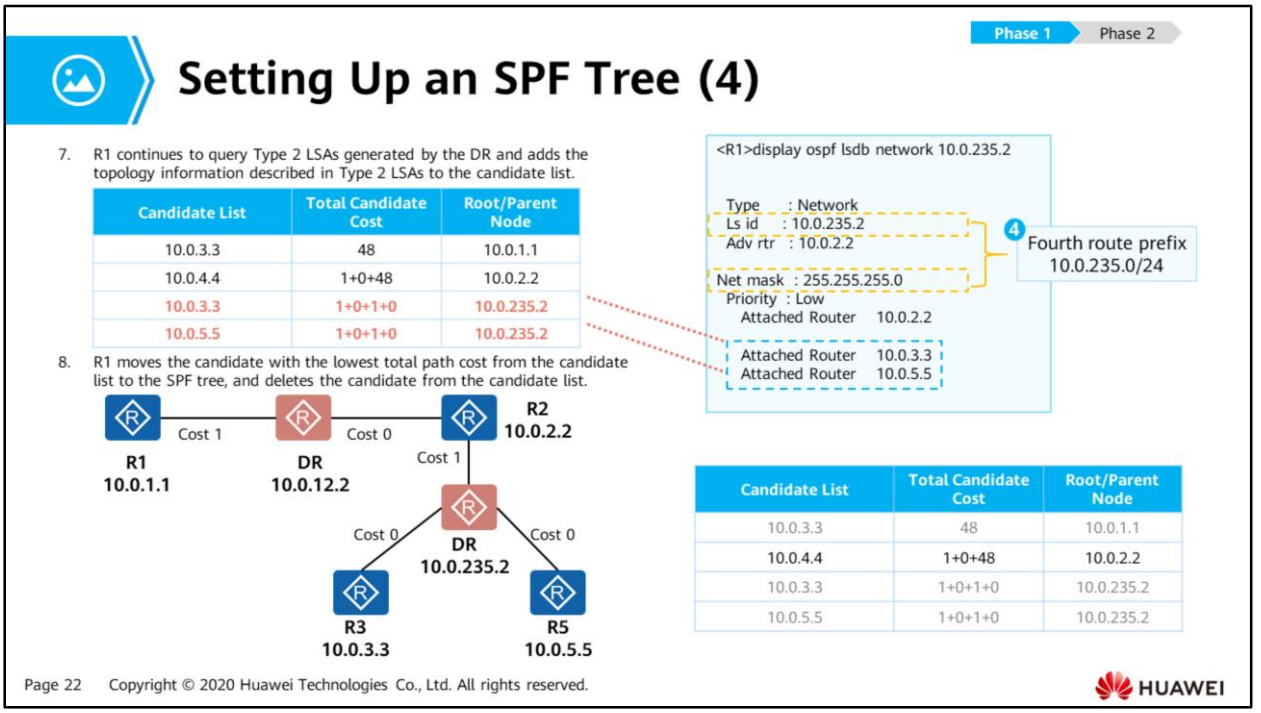
* Link ID: 10.0.4.4
  Data   : 10.0.24.2
  Link Type: P-2-P
  Metric : 48

* Link ID: 10.0.235.2
  Data   : 10.0.235.2
  Link Type: TransNet
  Metric : 1

* Link ID: 10.0.24.0
  Data   : 255.255.255.0
  Link Type: StubNet
  Metric : 48
  Priority : Low
```

3 Third route prefix 10.0.24.0/24

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

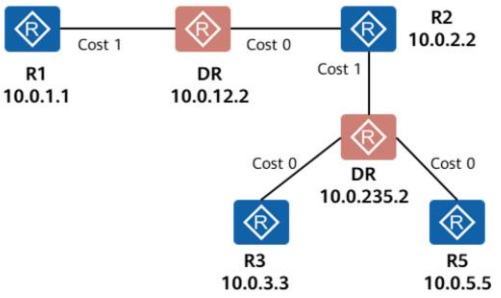


- R3 has two different cost values in the candidate list: 48 and 2. Therefore, R3 adds the route with the smallest cost to the SPF tree and deletes the route from the candidate list.

Setting Up an SPF Tree (5)

9. R1 queries Type 1 LSAs of R3. All neighbors of R3 are in the SPF tree, and the topology does not change.

Candidate List	Total Candidate Cost	Root/Parent Node
10.0.4.4	1+0+48	10.0.2.2




```
<R1>display ospf lsdb router 10.0.3.3

Type      : Router
Ls id     : 10.0.3.3
Adv rtr   : 10.0.3.3

 * Link ID: 10.0.235.2
   Data   : 10.0.235.3
   Link Type: TransNet
   Metric : 1

 * Link ID: 10.0.1.1
   Data   : 10.0.13.3
   Link Type: P-2-P
   Metric : 48

 * Link ID: 10.0.13.0
   Data   : 255.255.255.0
   Link Type: StubNet
   Metric : 48
   Priority : Low
```

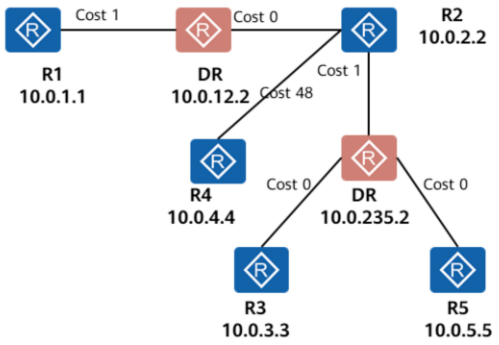
The two nodes do not need to be added to the topology because the two nodes already exist in the SPF tree.

1 Same as the first route prefix 10.0.13.0/24

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Setting Up an SPF Tree (6)

10. R1 queries Type 1 LSAs of R5, adds R4 to the SPF tree, and deletes R4 from the candidate list.



Candidate List	Total Candidate Cost	Root/Parent Node
10.0.4.4	1+0+48	10.0.2.2
10.0.4.4	1+0+1+0+48	10.0.5.5

```
<R1>display ospf lsdb router 10.0.5.5

Type      : Router
Ls id     : 10.0.5.5
Adv rtr   : 10.0.5.5

 * Link ID: 10.0.235.2
   Data   : 10.0.235.5
   Link Type: TransNet
   Metric : 1

 * Link ID: 10.0.4.4
   Data   : 10.0.45.5
   Link Type: P-2-P
   Metric : 48

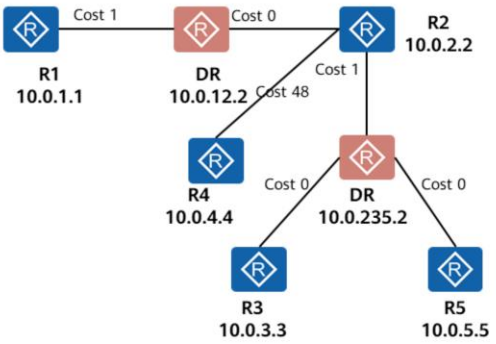
 * Link ID: 10.0.45.0
   Data   : 255.255.255.0
   Link Type: StubNet
   Metric : 48
   Priority : Low
```

5 Fifth route prefix 10.0.45.0/24

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Setting Up an SPF Tree (7)

11. R1 queries Type 1 LSAs of R4. All neighbors are in the SPF tree. In addition, all LSAs have been calculated, and the SPF tree has been established.




```
<R1>display ospf lsdb router 10.0.4.4

  Type      : Router
  Ls id     : 10.0.4.4
  Adv rtr   : 10.0.4.4

  * Link ID: 10.0.2.2
    Data   : 10.0.24.4
    Link Type: P-2-P
    Metric : 48
  * Link ID: 10.0.5.5
    Data   : 10.0.45.4
    Link Type: P-2-P
    Metric : 48

  * Link ID: 10.0.24.0
    Data   : 255.255.255.0
    Link Type: StubNet
    Metric : 48
    Priority : Low

  * Link ID: 10.0.45.0
    Data   : 255.255.255.0
    Link Type: StubNet
    Metric : 48
    Priority : Low
```

2 Same as the second route prefix 10.0.24.0/24

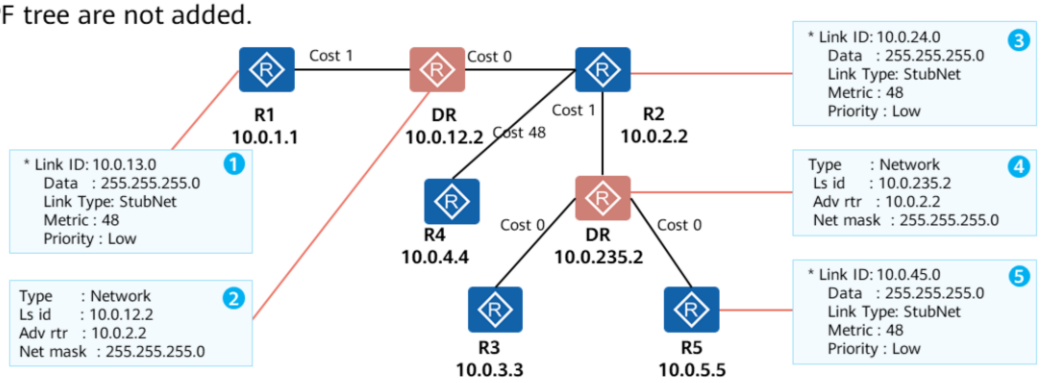
5 Same as the second route prefix 10.0.45.0/24

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Phase 1 Phase 2

Calculating the Optimal Route

Starting from the root node, routing information in the LSA of each node is added according to the sequence in which nodes are added to the SPF tree. The routes that have been added to the SPF tree are not added.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- In the second phase, the router calculates the optimal route based on routing information in the Router-LSA and Network-LSA.
- Starting from the root node, routing information in the LSA of each node is added according to the sequence in which nodes are added to the SPF tree:
  - In the Router-LSA of R1 at 10.0.1.1, there is one network. The network ID/subnet mask is 10.0.13.0/24, and the metric is 48.
  - In the Network-LSA of DR at 10.0.12.2, the network ID/subnet mask is 10.0.12.0/24, and the metric is 1 (1+0).
  - In the Router-LSA of R2 at 10.0.2.2, there is one network. The network ID/subnet mask is 10.0.24.0/24, and the metric is 49 (1+0+48).
  - In the Network-LSA of DR at 10.0.235.2, the network ID/subnet mask is 10.0.235.0/24, and the metric is 2 (1+0+1).
  - In the Router-LSA of R3 at 10.0.3.3, there is one network. The network ID/subnet mask is 10.0.13.0/24, which already exists on R1 and can be ignored.
  - In the Router-LSA of R5 at 10.0.5.5, there is one network. The network ID/subnet mask is 10.0.45.0/24, and the metric is 50 (1+0+0+1+48).
  - In the Router-LSA of R4 at 10.0.4.4, there are two networks. One network ID/subnet mask is 10.0.24.0/24, which already exists on R2 and is not added. The other network ID/subnet mask is 10.0.45.0/24, which already exists on R5 and is not added.

Verifying the Configuration

```
<R1>display ospf routing

        OSPF Process 1 with Router ID 10.0.1.1
                Routing Tables

 Routing for Network
 Destination     Cost  Type       NextHop         AdvRouter       Area
 10.0.12.0/24    1     Transit    10.0.12.1       10.0.1.1        0.0.0.0
 10.0.13.0/24    48    Stub       10.0.13.1       10.0.1.1        0.0.0.0
 10.0.24.0/24    49    Stub       10.0.12.2       10.0.2.2        0.0.0.0
 10.0.45.0/24    50    Stub       10.0.12.2       10.0.5.5        0.0.0.0
 10.0.235.0/24   2     Transit    10.0.12.2       10.0.5.5        0.0.0.0
```

OSPF routing table of R1

```
<R1>display ip routing-table

Destination/Mask    Proto   Pre  Cost      NextHop
      10.0.12.0/24  Direct  0    0         10.0.12.1
      10.0.13.0/24  Direct  0    0         10.0.13.1
      10.0.24.0/24  OSPF    10   49        10.0.12.2
      10.0.45.0/24  OSPF    10   50        10.0.12.2
     10.0.235.0/24  OSPF    10   2         10.0.12.2
```

Global routing table of R1

Question: Why are not all routes in the OSPF routing table of R1 displayed in the global routing table?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Multiple) Which of the following link types are included in a Router-LSA?( )
   A. P2P
   B. TransNet
   C. StubNet
   D. Vlink
2. (TorF) After the SPF algorithm is used to calculate routes, the OSPF routes that are considered as the optimal routes are added to the routing table of the router.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. ABCD
2. False

Section Summary

- OSPF calculates the routing table based on the LSDB. The LSDB may contain multiple types of LSAs, and all LSAs have the same packet header format.
- OSPF routers in the same area have the same LSDB. When there is only one area, there are two types of LSAs in the area: Router-LSA and Network-LSA.
- Each router generates a Router-LSA that describes the information about the directly connected interface of the router.
- On an MA network, a DR generates a Network-LSA to describe the router IDs (including the DR itself) of all the routers that access the MA network and the mask of the network.

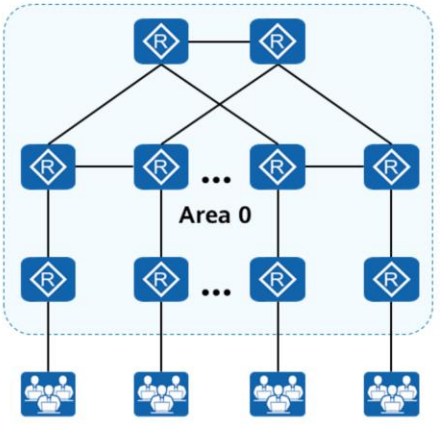
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Intra-Area Route Calculation
2. **Inter-Area Route Calculation**
   - **Inter-Area Route Calculation Process**
   - Inter-Area Routing Loop Prevention Mechanism
   - Functions and Configurations of Virtual Links
3. External Route Calculation

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

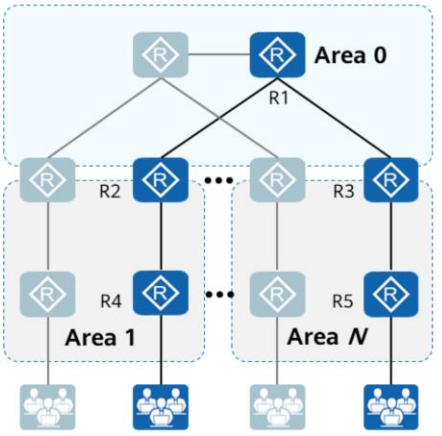
Problems in a Single OSPF Area on a Large-Scale Network



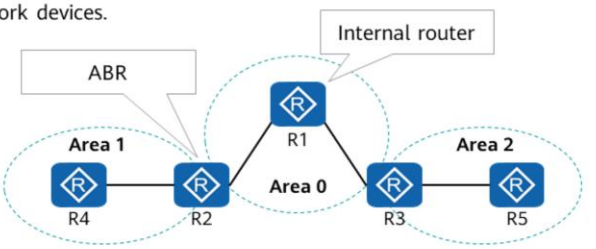
- A network consisting of a series of consecutive OSPF routers is called an OSPF domain.
- OSPF requires that routers on the network synchronize their LSDBs to keep consistent network information.
- When the network scale increases continuously, the LSDB becomes oversized. A device calculates routes based on the LSDB, which increases the burden. In addition, the size of the routing table of the router increases, deteriorating the performance of the router.
- When the network topology changes, the changes need to be spread to the entire network and may trigger route recalculation on the entire network.
- The single-area design makes it impossible to deploy route summarization.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Area Partitioning



- Router-LSAs and Network-LSAs are flooded only within an area, so area partitioning reduces the memory and CPU usage of network devices.

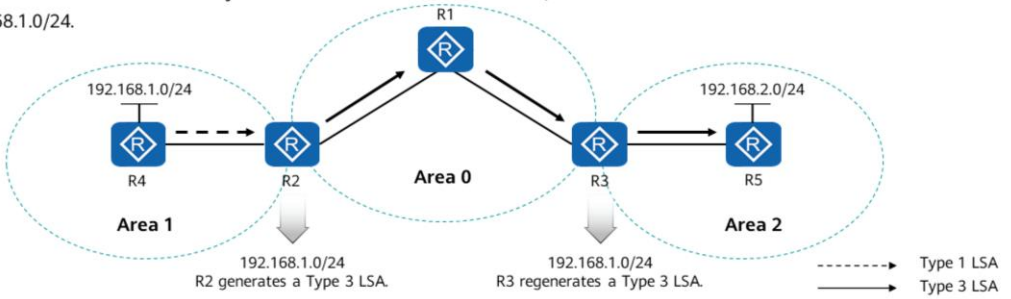


- After areas are partitioned, routers can be classified into the following types:
  - Internal router: An internal router has all its interfaces belonging to the same OSPF area, for example, R1, R4, and R5.
  - Area border router (ABR): An ABR connects to two or more areas, for example, R2 and R3.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Inter-Area Routing Information Transmission

- OSPF inter-area routing information is transmitted through Type 3 Network-Summary-LSAs generated by ABRs.
- Transmission of the route to 192.168.1.0/24 is used as an example.
  - R2 calculates a route (intra-area route) to 192.168.1.0/24 based on the Router-LSA and Network-LSA flooded in area 1, and advertises the route to area 0 through a Network-Summary-LSA. R3 can calculate the inter-area route to 192.168.1.0/24 based on the Network-Summary-LSA.
  - R3 generates a Network-Summary-LSA and advertises it to area 2. Then, routers in all OSPF areas can learn the route to 192.168.1.0/24.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Description of a Network-Summary-LSA

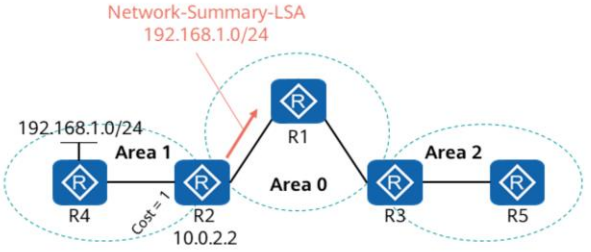
An ABR generates Network-summary-LSAs (Type 3 LSAs) and advertises them within its area to notify destination addresses of routes from this area to other areas.

LS Age	Options	**LS Type**
Link State ID		
Advertising Router		
LS Sequence Number		
LS Checksum	Length	
Network Mask		
0	**Metric**	
...		

- Main Fields
  - LS Type: The value 3 indicates a Network-summary-LSA.
  - Link State ID: indicates the destination network address of a route.
  - Advertising Router: indicates the router ID of the device that generates the LSA.
  - Network Mask: indicates the network mask of a route.
  - Metric: indicates the cost of the route to the destination address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Example of a Network-Summary-LSA



This LSA is generated by R2 and is used to advertise inter-area routes destined for 192.168.1.0/24 to area 0.

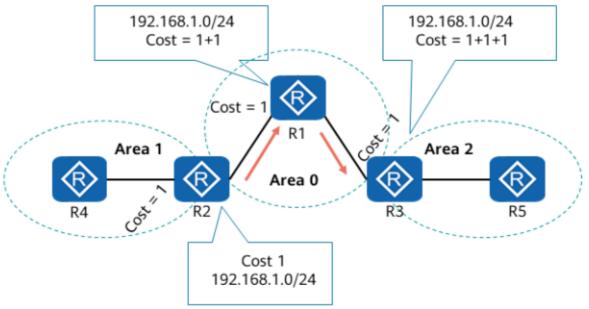
```
<R2>display ospf lsdb summary 192.168.1.0
            OSPF Process 1 with Router ID 10.0.2.2
              Area: 0.0.0.0
              Link State Database

Type          : Sum-Net
Ls id         : 192.168.1.0
Adv rtr       : 10.0.2.2
Ls age        : 86
Len           : 28
Options       : E
seq#          : 80000001
chksum        : 0x7c6d
Net mask      : 255.255.255.0
Tos0          metric: 1
Priority      : Low
```

Routing information 192.168.1.0/24

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Calculation on R1 and R3

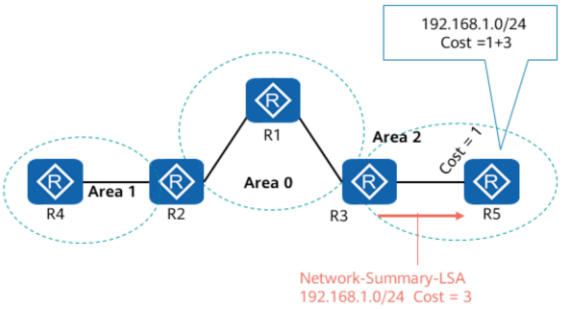


The cost of the route from R2 to 192.168.1.0/24 is 1, so the cost of the route to 192.168.1.0/24 advertised by R2 to area 0 is 1.

1. Through SPF calculation, the cost of the route from R1 to R2 is 1, and that of the route from R3 to R2 is 2.
2. R1 and R3 calculate routes based on the received Network-Summary-LSAs.
   - R1 adds the cost of the route to R2 and the cost carried in the Network-summary-LSA. Therefore, the cost of the route from R1 to 192.168.1.0/24 is 2.
   - R3 adds the cost of the route to R2 and the cost carried in the Network-summary-LSA. Therefore, the cost of the route from R3 to 192.168.1.0/24 is 3.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Calculation on R5

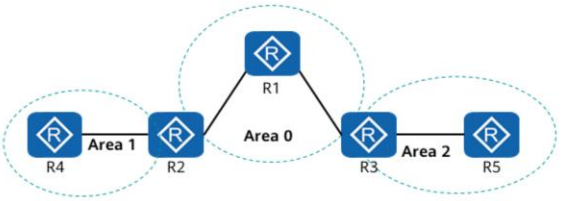


The cost of the route from R3 to 192.168.1.0/24 is 3, so the cost of the route to 192.168.1.0/24 advertised by R3 to area 2 is 3.

1. R3 functions as an ABR. It calculates the route to 192.168.1.0/24 based on the Network-Summary-LSA flooded in area 0, and reinjects the Network-Summary-LSA to area 2. The Network-Summary-LSA contains the cost (value 3) of the route to the network segment.
2. Through SPF calculation, the cost of the route from R5 to R3 is 1, so the cost of the route from R5 to 192.168.1.0/24 is 4.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Inter-Area Route Calculation Result Verification




```
<R1>display ip routing-table
Destination/Mask    Proto   Pre  Cost
    192.168.1.0/24  OSPF    10   2
```

```
<R3>display ip routing-table
Destination/Mask    Proto   Pre  Cost
    192.168.1.0/24  OSPF    10   3
```

```
<R5>display ip routing-table
Destination/Mask    Proto   Pre  Cost
    192.168.1.0/24  OSPF    10   4
```

Check the routing tables on R1, R3, and R5.

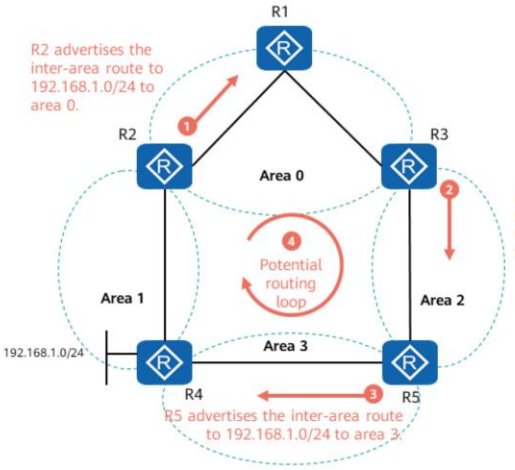
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Intra-Area Route Calculation
2. **Inter-Area Route Calculation**
   - Inter-Area Route Calculation Process
   - **Inter-Area Routing Loop Prevention Mechanism**
   - Functions and Configurations of Virtual Links
3. External Route Calculation

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

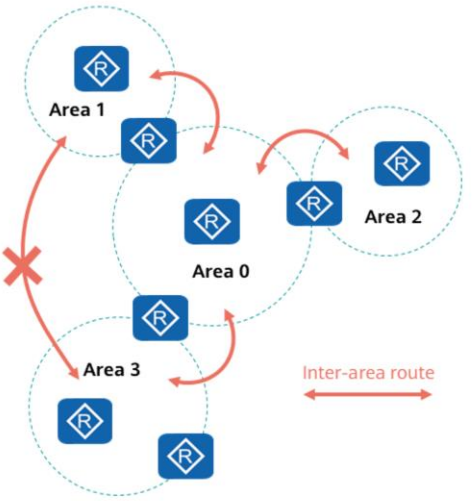
Inter-Area Routing Loops



- The process of advertising OSPF inter-area routes is similar to that of advertising routes of distance-vector routing protocols.
- OSPF also requires a loop prevention mechanism for inter-area route advertisement.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Inter-Area Routing Loop Prevention Mechanism (1)

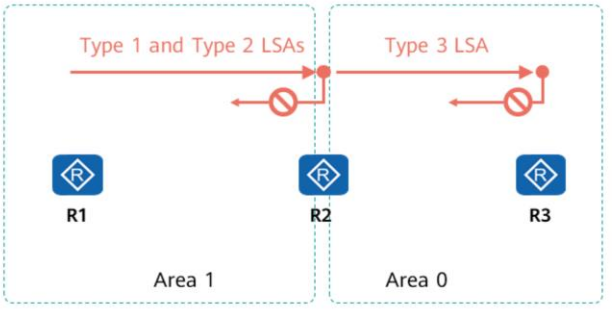


- OSPF requires that all non-backbone areas be directly connected to area 0 and inter-area routes be advertised through area 0.
- Inter-area routes cannot be advertised between two non-backbone areas. As such, a star topology is logically formed in the OSPF area architecture.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- OSPF requires that at least one interface of the ABR should belong to the backbone area.

Inter-Area Routing Loop Prevention Mechanism (2)

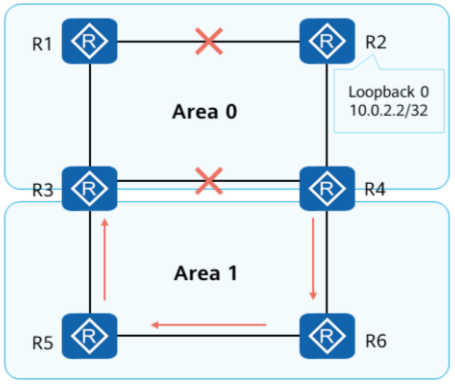


An ABR does not reinject Type 3 LSAs that describe routes to a network segment in an area to the area.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Inter-Area Routing Loop Prevention Mechanism (3)

Type 3 LSAs received by an ABR from a non-backbone area cannot be used to calculate inter-area routes.



The link between R1 and R2 and the link between R3 and R4 are disconnected, forming the non-contiguous backbone area.

- R4 advertises the route to 10.0.2.2/32 to area 1 in the form of Type 3 LSA.
- R5 and R6 can calculate the route to 10.0.2.2/32 based on the Type 3 LSA.
- After receiving Type 3 LSAs from a non-backbone area, R3 does not calculate routes or advertise them to other areas.
- In this case, R1 and R3 cannot communicate with the device at 10.0.2.2/32.

How to solve the problem of the non-contiguous backbone area?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

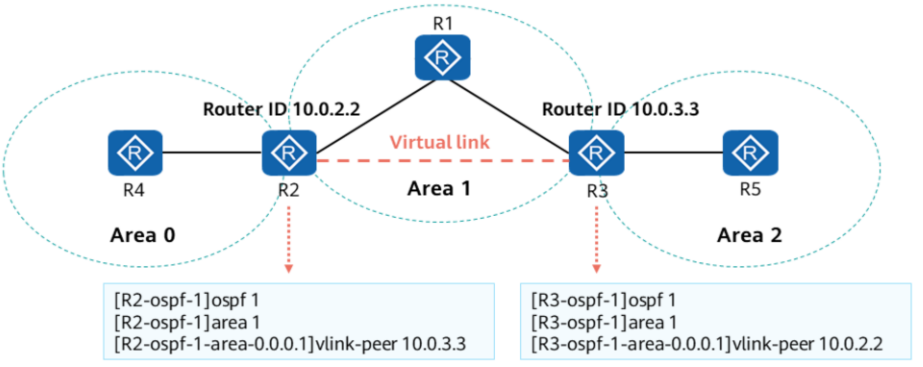
Contents

1. Intra-Area Route Calculation
2. **Inter-Area Route Calculation**
   - Intra-Area Route Calculation Process
   - Inter-Area Routing Loop Prevention Mechanism
   - **Functions and Configurations of Virtual Links**
3. External Route Calculation

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Functions and Configurations of Virtual Links

- OSPF requires that the backbone area should be contiguous physically or logically using virtual links.
- A virtual link can be established between any two ABRs, but the two ABRs must have interfaces connected to the same non-backbone area.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Note: The virtual link enables OSPF routers to communicate through a non-backbone area, which may cause routing loops in some scenarios. Therefore, you are not advised to deploy an OSPF virtual link.



Quiz

1. (TorF) A Network-summary-LSA can describe only one route.
2. (Essay) How does OSPF prevent inter-area routing loops?

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

1. True
2. An OSPF network is partitioned into the backbone area and non-backbone areas. All non-backbone areas are directly connected to the backbone area, and there is only one backbone area. Non-backbone areas communicate with each other through the backbone area. In addition, the Type 3 LSAs from the backbone area do not return to the backbone area.

Section Summary

- OSPF introduces the concept of multi-area to support larger-scale networking.
- OSPF uses Type 3 LSAs to describe inter-area routing information.
- To prevent inter-area routing loops, OSPF defines multiple rules.
- An OSPF virtual link is a virtual and logical link deployed between two OSPF routers. It traverses a non-backbone area to connect another non-backbone area to the backbone area (area 0). The virtual link is only used in the scenario where a non-backbone area is not directly connected to area 0.

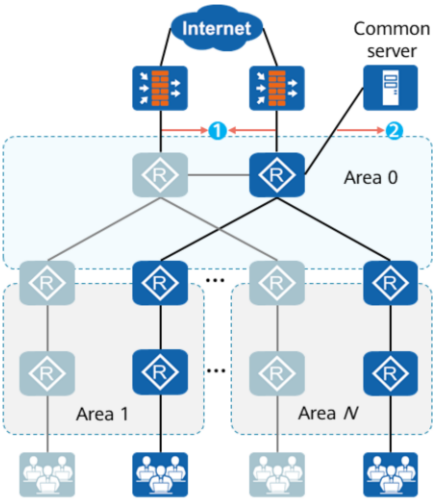
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Intra-Area Route Calculation
2. Inter-Area Route Calculation
3. **External Route Calculation**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Importing External Routes

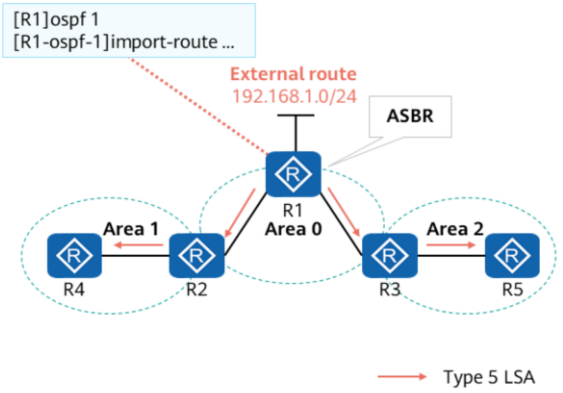


The OSPF protocol is not enabled for some links on the network.

- Routers connect to external networks using static routes or BGP routes.
- The OSPF protocol is not enabled for the link directly connected to the server.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Concepts for External Route Import



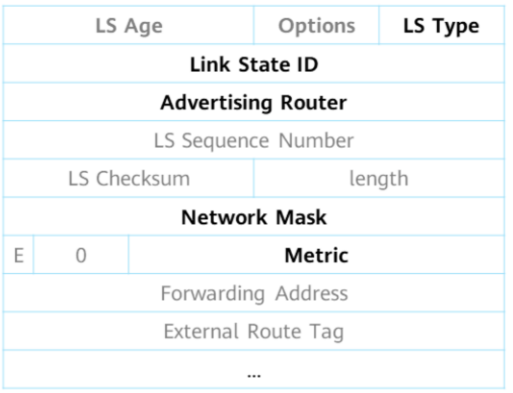
- Autonomous system boundary router (ASBR): An OSPF device that imports external routes will become an ASBR, for example, R1 in the figure.
- An ASBR floods external routes in AS-external-LSAs (Type 5 LSAs) on an OSPF network.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Run the following commands in the OSPF process view to import external routes: BGP routes, IS-IS routes, OSPF routes, direct routes, and static routes can be imported.
  - **import-route** { **limit** *limit-number* | { **bgp** [**permit-ibgp**] | **direct** | **unr** | **rip** [*process-id-rip*] | **static** | **isis** [*process-id-isis*] | **ospf** [*process-id-ospf*] } [**cost** *cost* | **type** *type* | **tag** *tag* | **route-policy** *route-policy-name*] * }

Description of AS-external LSAs

AS-external-LSA (Type 5 LSA): Originated by an ASBR, an AS-External-LSA describes routes to destinations outside an AS, and is flooded to all areas except stub areas and NSSAs.



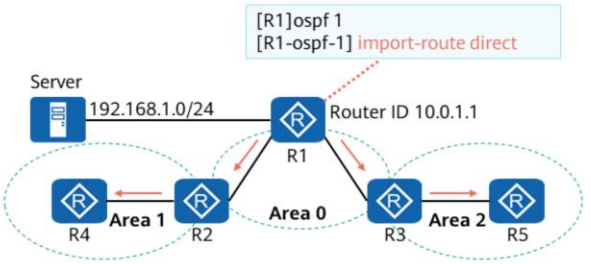
Main fields

- LS Type: The value 5 indicates an AS-external-LSA.
- Link State ID: indicates the destination network address of an external route.
- Advertising Router: indicates the router ID of the device that originates the LSA.
- Network Mask: indicates the network mask of a route.
- E: indicates the metric type used by the external route.
  - 0: The metric type is Metric-Type-1.
  - 1: The metric type is Metric-Type-2.
- Metric: indicates the cost of the route to the destination network.
- Forwarding Address (FA): Packets destined for the advertised destination address are forwarded to the address specified by this field.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Forwarding Address: When the value of this field is 0.0.0.0, traffic destined for the external network segment is sent to the ASBR that imports the external route. If the value of this field is not 0.0.0.0, traffic is sent to this forwarding address. This field is used to avoid the sub-optimal path problem in some special scenarios.
- External Route Tag: indicates the external route tag and is often used to deploy routing policies.

Example of an AS-external-LSA



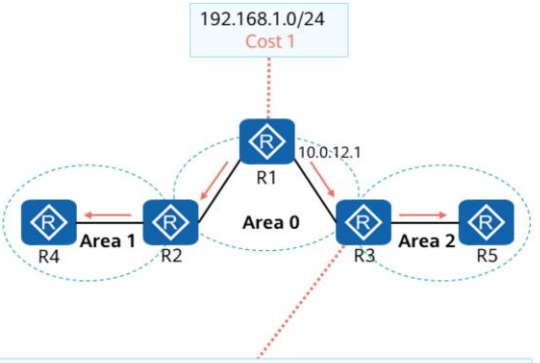
R1 is directly connected to the server on the network segment 192.168.1.0/24. After the direct route is imported to OSPF on R1, R1 injects an AS-external-LSA that describes the route to 192.168.1.0/24 to OSPF. The AS-external-LSA is flooded in the entire OSPF area.

```
<R1>display ospf lsdb ase self-originate
            OSPF Process 1 with Router ID 10.0.1.1
                Link State Database
Type                : External
Ls id               : 192.168.1.0
Adv rtr             : 10.0.1.1
Ls age              : 1340
Len                 : 36
Options             :  E
seq#                : 80000004
chksum              : 0xb5cc
Net mask            : 255.255.255.0
TOS 0               Metric: 1
E type              : 2
Forwarding Address  : 0.0.0.0
Tag                 : 1
Priority            : Low
```

Routing information 192.168.1.0/24

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Calculation on R3



[R3]display ospf routing

Destination	Cost	Type	NextHop	AdvRouter
192.168.1.0/24	1	Type2	10.0.12.1	10.0.1.1

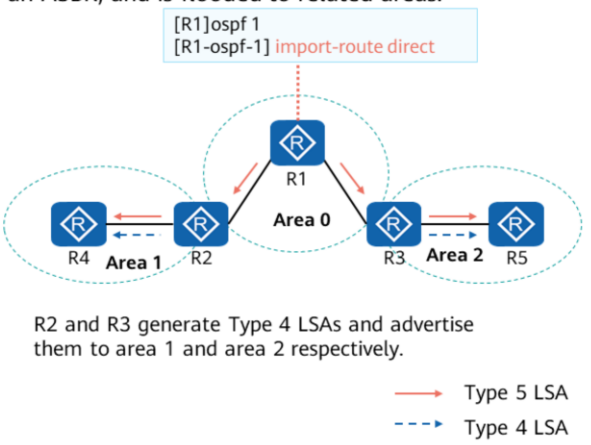
1. External routes are attached to R1.
2. According to the Advertising Router field in the Type 5 LSA, R3 determines that the packet needs to pass through R1 before reaching the external network.
3. R3 performs intra-area SPF calculation and adds R1 to the SPF tree. R3 sets the next-hop router of the route to the external network to R1.
4. The calculation method of R2 is similar to that of R3.

- R3 and R1 (ASBR) are in the same area, so R3 can calculate the route to the ASBR based on the flooded Router-LSAs and Network-LSAs.
- If R5 and R1 are not in the same area, how does R5 calculate external routes?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Description of an ASBR-Summary-LSA

ASBR-Summary-LSA (Type 4 LSA): Originated by an ABR, an ASBR-summary-LSA describes the routes to an ASBR, and is flooded to related areas.

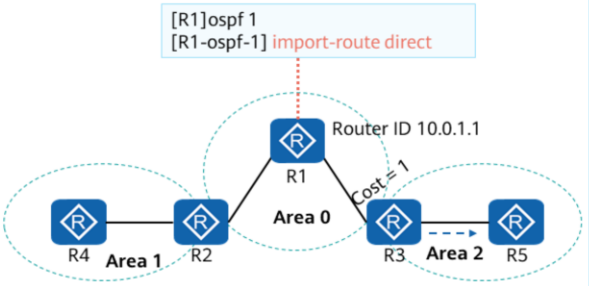


LS Age	Options	LS Type
Link State ID		
Advertising Router		
LS Sequence Number		
LS Checksum	Length	
Network Mask		
0	Metric	
...		

- Main fields
  - LS Type: The value 4 indicates an ASBR-summary LSA.
  - Link State ID: indicates the router ID of an ASBR.
  - Advertising Router: indicates the router ID of the device that originates the LSA.
  - Network Mask: This parameter is reserved.
  - Metric: indicates the cost of the route to the destination address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Example of an ASBR-Summary LSA



The following uses the ASBR-summary LSA advertised by R3 to area 2 as an example.

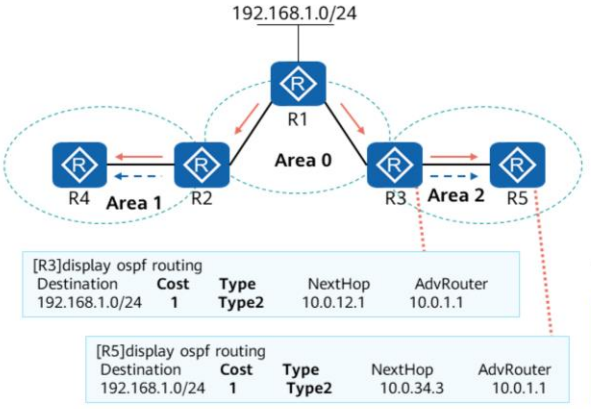
```
<R3>display ospf lsdb asbr self-originate
Area: 0.0.0.2
          Link State Database
Type      : Sum-Asbr
Ls id     : 10.0.1.1
Adv rtr   : 10.0.3.3
Ls age    : 15
Len       : 28
Options   :  E
seq#      : 80000005
chksum    : 0xf456
Tos 0     metric: 1
```

Router ID of the ASBR

Cost of the route from R3 to the ASBR

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Calculation on R5



[R3]display ospf routing

Destination	Cost	Type	NextHop	AdvRouter
192.168.1.0/24	1	Type2	10.0.12.1	10.0.1.1

[R5]display ospf routing

Destination	Cost	Type	NextHop	AdvRouter
192.168.1.0/24	1	Type2	10.0.34.3	10.0.1.1

1. External routes are attached to R1.
2. R5 determines that the Type 5 LSA is generated by R1 according to the Advertising Router field in the Type 5 LSA. However, R1 does not exist in the SPF tree of R5.
3. Based on the Type 4 LSAs generated by R3, R5 sets the next hop of the route to the external network to R3.
4. Route calculation on R4 is similar to that on R5.

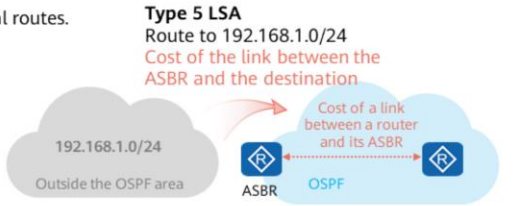
On the physical network, the costs of the routes from R3 and R5 to the external network are different, but cannot be represented in the routing table.

Note: The cost of all device interfaces is 1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Distinguishing Two Types of OSPF External Routes (1)

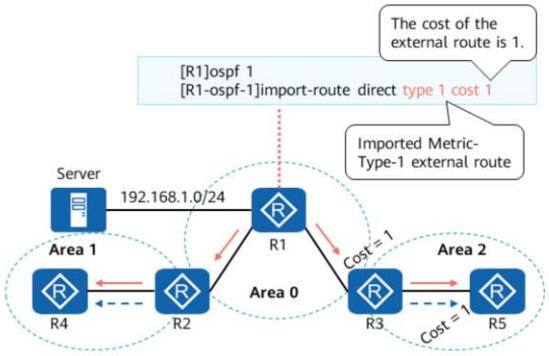
- **Metric-Type-1**
  - When the cost of an external route equals that of an AS internal route and can be compared with the cost of an OSPF route, the external route has high reliability and can be configured as a Metric-Type-1 external route.
  - The cost of a Metric-Type-1 external route equals that of a link between a router and its ASBR plus the cost of the link between the ASBR and the destination.
- **Metric-Type-2**
  - When the cost of the route from an ASBR to the destination outside an AS is much greater than the cost of the internal route to the ASBR, the external route has low reliability and can be configured as a Metric-Type-2 external route.
  - Metric-Type-2 external routes have the same cost as AS external routes.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Distinguishing Two Types of OSPF External Routes (2)

Type	Description	Cost
Metric-Type-1	High reliability.	Cost of the AS internal route + Cost of the AS external route
Metric-Type-2 (default)	Low reliability. The cost of the AS external route is much greater than the cost of the AS internal route.	Cost of the AS external route



[R3]display ospf routing

Destination	Cost	Type	Tag	NextHop	AdvRouter
192.168.1.0/24	**2**	**Type1**	1	10.0.12.1	10.0.1.1

The external route type is Metric-Type-1. In the routing table of R3, the cost of the route to 192.168.1.0/24 is the cost of the external route plus the cost of the internal route from R3 to R1 (ASBR).

[R5]display ospf routing

Destination	Cost	Type	Tag	NextHop	AdvRouter
192.168.1.0/24	**3**	**Type1**	1	10.0.34.3	10.0.1.1

The cost of the route to 192.168.1.0/24 in the routing table of R5 is equal to the cost of the external route plus the cost of the internal route from R5 to R1 (ASBR).

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Multiple) Which of the following types of LSAs may be originated by an ABR?( )
   A. ASBR-summary-LSA
   B. AS-external-LSA
   C. Router-LSA
   D. Network-summary-LSA
2. (Single) Which of the following routes has the highest priority for the same route prefix?( )
   A. Intra-area route
   B. Inter-area route
   C. Type 1 external route
   D. Type 2 external route

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. ABCD
2. A

Section Summary

- On a large-scale network, there may be multiple routing protocols. OSPF can import AS external routes to the local AS so that routers in the AS can learn the routes to other ASs.
- After an ASBR imports external routes into OSPF, the ASBR uses Type 5 LSAs to describe these routes. Routers in the same area as the ASBR can calculate external routes based on Type 5 LSAs as well as Type 1 and Type 2 LSAs in the area. Routers that are not in the same area as the ASBR need to use Type 4 LSAs to calculate external routes.
- After OSPF imports external routes, you can set the metric type of the routes to Metric-Type-1 or Metric-Type-2. The cost calculation methods and priorities of external routes with different metric types are different, Metric-Type-1 routes have a higher priority than Metric-Type-2 routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Summary

- This course describes calculation of OSPF intra-area routes, inter-area routes, and external routes.
- Intra-area routes, inter-area routes, Metric-Type-1 external routes, and Metric-Type-2 external routes are in descending order of priority.
- LSAs carry OSPF link state information. As the network scale expands, network services become complex. As OSPF areas are partitioned and a large number of external routes are imported, the LSDB of a router may contain a large number of Type 1, Type 2, Type 3, Type 4, and Type 5 LSAs, which may cause the device performance to deteriorate.
- What technical means does OSPF use to prevent the device performance from deteriorating while ensuring network reachability?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



OSPF Special Areas and Other Features

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- On a large-scale network, an Open Shortest Path First (OSPF) router needs to maintain the database composed of intra-area, inter-area, and external routes. As the network scale expands, the link state database (LSDB) size also increases. If an area does not need to provide traffic forwarding services for other areas, routers in this area do not need to maintain link status information outside the area.
- By partitioning areas, OSPF can reduce the LSDB size of routers in an area. Low-end routers in non-backbone areas at the border of an Autonomous System (AS) cannot accommodate the LSDB. OSPF special areas can further reduce the number of Link State Advertisements (LSAs) and the size of the routing table.
- This course describes OSPF special areas and OSPF-related features.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe types and characteristics of OSPF special areas.
  - Describe application scenarios and advantages of OSPF route summarization.
  - Configure OSPF packet authentication.

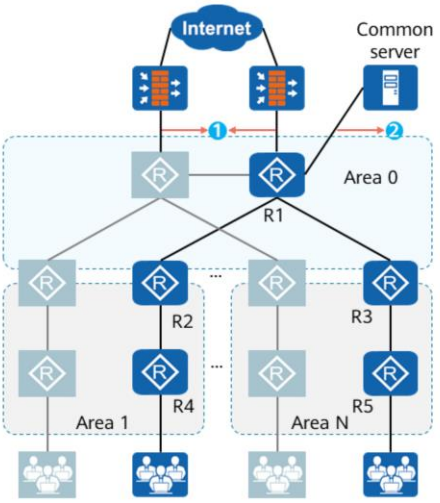
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Stub Area and Totally Stub Area**
2. NSSA and Totally NSSA
3. Summarization of Inter-Area and External Routes
4. OSPF Features

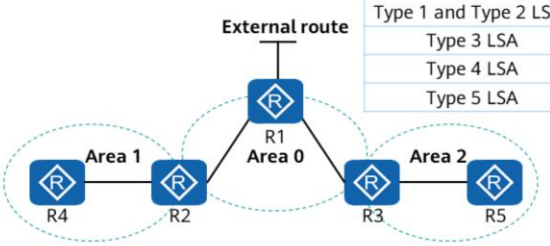
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Problems Caused by the Increase of the Network Scale



OSPF routers calculate intra-area, inter-area, and external routes based on LSAs on a network. When the network scale increases, the LSDB size of the device also increases, which makes route calculation more difficult and lowers device performance.

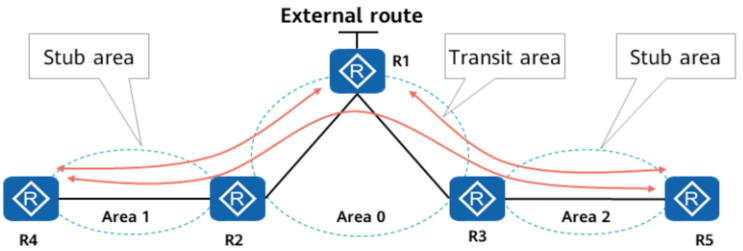
LSDB of R5
Type 1 and Type 2 LSAs
Type 3 LSA
Type 4 LSA
Type 5 LSA



How to reduce the number of LSAs without affecting IP reachability?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Transit Area and Stub Area



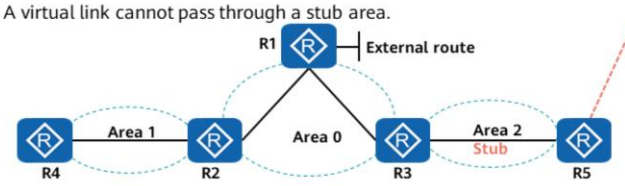
OSPF areas are classified into the following types:

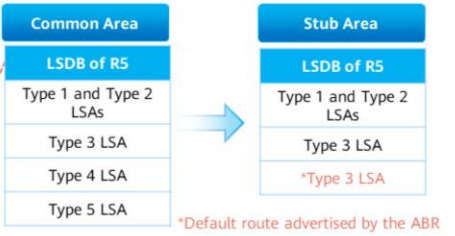
- Transit area: carries traffic originated from and destined for the local area as well as traffic whose source and destination IP addresses do not belong to the local area, for example, traffic in area 0.
- Stub area: carries only the traffic initiated by and destined for the local area, for example, traffic in areas 1 and 2.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Stub Area

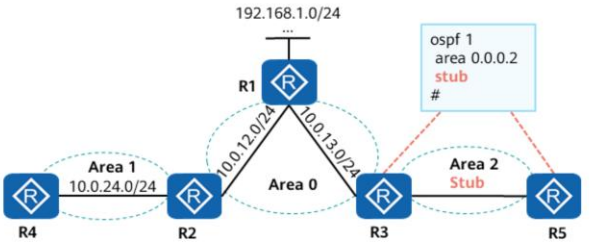
- The Area Border Router (ABR) in a stub area does not advertise the received AS external routes to the stub area, which significantly decreases the LSDB and routing table sizes of routers in the stub area.
- To ensure the reachability of AS external routes, the ABR generates a default route (carried by Type 3 LSAs) and advertises the route to non-ABRs in the stub area.
- Note the following when configuring a stub area:
  - A backbone area cannot be configured as a stub area.
  - Stub area attributes need to be configured on all the routers in the area that needs to be configured as a stub area.
  - AS external routes cannot be imported or received within a stub area.
  - A virtual link cannot pass through a stub area.





Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Routing Table of the Stub Area and Type 3 LSAs




```
ospf 1
 area 0.0.0.2
 stub
#
```

```
<R5>display ospf lsdb
         OSPF Process 1 with Router ID 10.0.5.5
                 Link State Database
                 Area: 0.0.0.2
Type      LinkState ID    AdvRouter       Metric
Sum-Net   0.0.0.0         10.0.3.3        1
Sum-Net   10.0.13.0       10.0.3.3        1
Sum-Net   10.0.24.0       10.0.3.3        3
Sum-Net   10.0.12.0       10.0.3.3        2
```

Type 4 and Type 5 LSAs do not exist, but Type 3 LSAs that carrying inter-area routes still exist.

```
<R5>display ospf routing

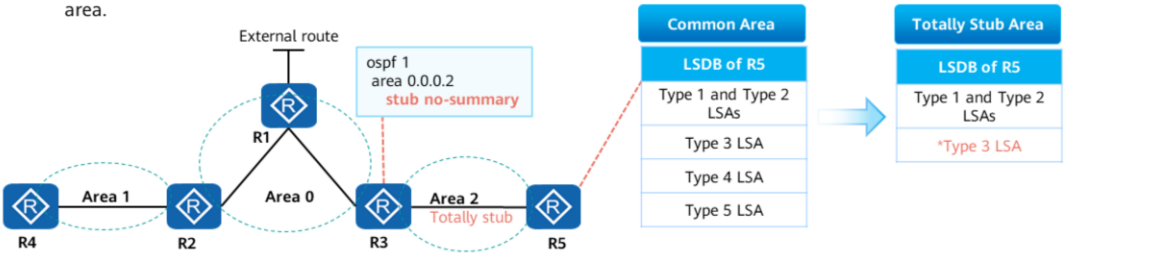
         OSPF Process 1 with Router ID 10.0.5.5
                 Routing Tables
Routing for Network
Destination     Cost  Type        NextHop         AdvRouter       Area
10.0.35.0/24    1     Transit     10.0.35.5       10.0.5.5        0.0.0.2
0.0.0.0/0       2     Inter-area  10.0.35.3       10.0.3.3        0.0.0.2
10.0.12.0/24    3     Inter-area  10.0.35.3       10.0.3.3        0.0.0.2
10.0.13.0/24    2     Inter-area  10.0.35.3       10.0.3.3        0.0.0.2
10.0.24.0/24    4     Inter-area  10.0.35.3       10.0.3.3        0.0.0.2
```

- R1 functions as an ASBR to import multiple external network segments. If area 2 is a common area, R3 imports Type 5 and Type 4 LSAs to this area.
- After area 2 is configured as a stub area:
  - R3 does not inject Type 5 and Type 4 LSAs into area 2.
  - R3 sends a Type 3 LSA carrying the default route to area 2. Although routers in area 2 do not know the specific route to the AS external network, they can still reach the AS external network through the default route.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Totally Stub Area (1)

- A totally stub area does not allow AS external routes or inter-area routes to be advertised within the area.
- Other routers except the ABR in a totally stub area use default routes (carried by Type 3 LSAs) delivered by the ABR to each other areas and AS external network.
- When configuring a totally stub area, pay attention to the following points:
  - The difference between the stub area and the totally stub area is that no-summary is specified on the ABR for the totally stub area.

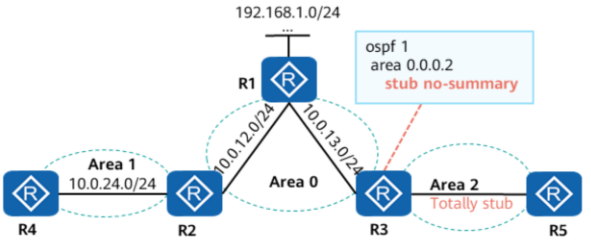


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Question: Why is the **no-summary** parameter not required for non-ABR devices?

Totally Stub Area (2)

- The totally stub area accesses other areas and AS external networks through default routes.
- The topology and route changes outside an AS or in other OSPF areas do not cause routers in the totally stub area to recalculate routes. This prevents device performance from deteriorating.




```
<R5>display  ospf lsdb
            OSPF Process 1 with Router ID 10.0.5.5
                        Link State Database
                        Area: 0.0.0.2
Type        LinkState ID  AdvRouter        Metric
Sum-Net     0.0.0.0       10.0.3.3         1
```

After area 2 is configured as a totally stub area:

- R3 does not inject Type 5 and Type 4 LSAs into area 2.
- R3 does not inject Type 3 LSAs into area 2, but injects a default route carried by Type 3 LSAs into area 2.
- R5 uses the default route to reach the AS external network and other OSPF areas.

- The stub area and totally stub area solve the problem of maintaining a large LSDB in the end area. In certain scenarios, they are not the best solutions.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

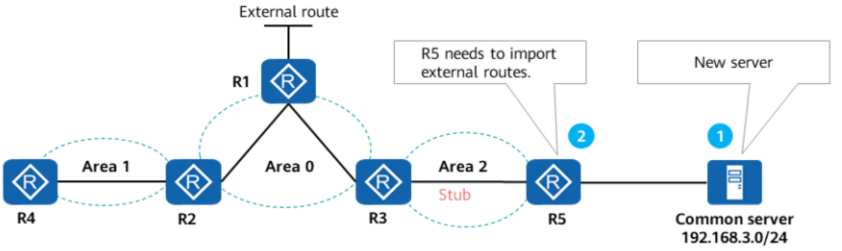
Contents

1. Stub Area and Totally Stub Area
2. **NSSA and Totally NSSA**
3. Summarization of Inter-Area and External Routes
4. OSPF Features

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Problems in the Stub Area and Totally Stub Area

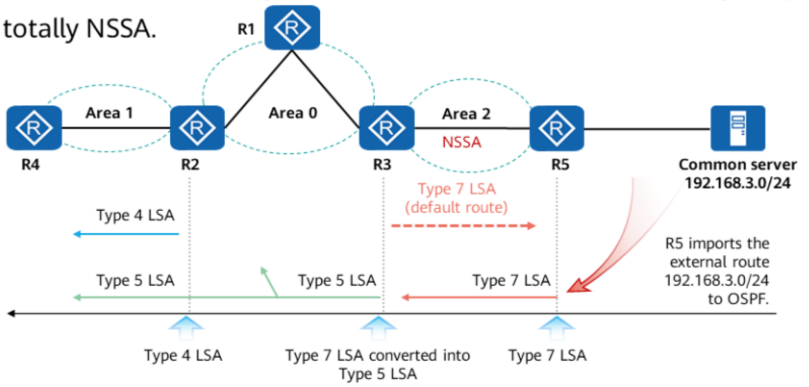
- As defined in OSPF, the stub area cannot import external routes. This mechanism prevents a large number of external routes from consuming resources of devices in the stub area.
- Stub and totally stub areas cannot apply to the scenario where external routes need to be imported and resource consumption caused by external routes also needs to be avoided.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

NSSA and Totally NSSA

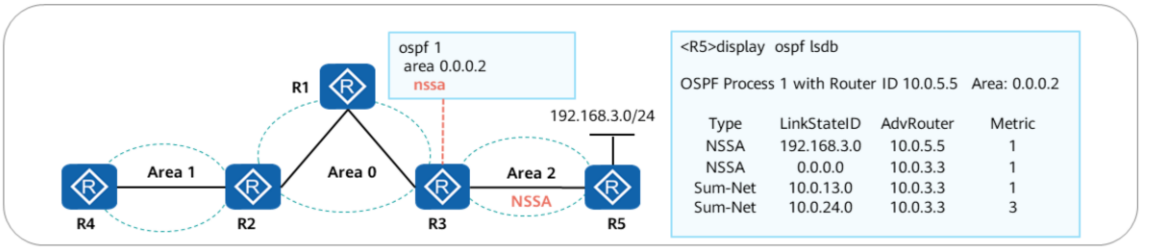
- An NSSA can import external routes but cannot learn external routes imported from other areas on the OSPF network.
- The difference between the totally NSSA and NSSA is that no-summary is specified on the ABR for the totally NSSA.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- An NSSA can import external routes but cannot learn external routes from other areas on the OSPF network.
  - Type 7 LSAs are generated by an Autonomous System Boundary Router (ASBR) in an NSSA and advertised only within the NSSA. After an ASBR in an NSSA imports external routes to the NSSA, the ASBR generate Type 7 LSAs to carry information about the external routes.
  - The Type 7 LSAs are advertised only within the NSSA.
  - When Type 7 LSAs reach an ABR in the NSSA, the ABR translates the Type 7 LSAs into Type 5 LSAs, imports them to the backbone area, and floods them to the entire AS.
  - The ABR in an NSSA prevents external routes imported from other areas from being imported into the NSSA. That is, Type 4 and Type 5 LSAs do not exist in the NSSA. To enable routers in the NSSA to reach the AS external network through the backbone area, the ABR in the NSSA automatically injects a default route carried by a Type 7 LAS into the NSSA.

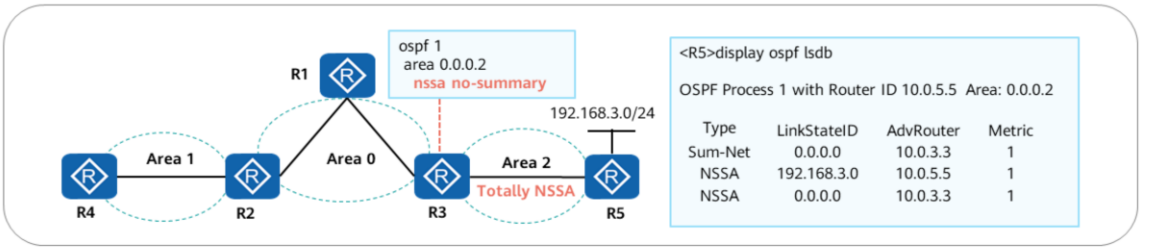
LSDBs of the NSSA and Totally NSSA



<R5>display ospf lsdb

OSPF Process 1 with Router ID 10.0.5.5 Area: 0.0.0.2

Type	LinkStateID	AdvRouter	Metric
NSSA	192.168.3.0	10.0.5.5	1
NSSA	0.0.0.0	10.0.3.3	1
Sum-Net	10.0.13.0	10.0.3.3	1
Sum-Net	10.0.24.0	10.0.3.3	3



<R5>display ospf lsdb

OSPF Process 1 with Router ID 10.0.5.5 Area: 0.0.0.2

Type	LinkStateID	AdvRouter	Metric
Sum-Net	0.0.0.0	10.0.3.3	1
NSSA	192.168.3.0	10.0.5.5	1
NSSA	0.0.0.0	10.0.3.3	1

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Scenario 1 (area 2 is configured as an NSSA): When R5 imports external route 192.168.3.0/24 to the NSSA, R5 functions as an ASBR to generate Type 7 LSAs and flood them in area 2. R3 generates a default route that is carried by a Type 7 LSA and imports the route to area 2. The routers in area 2 still receive the Type 3 LSAs imported by R3 and calculate the inter-area routes to other areas.
- Scenario 2 (area 2 is configured as a totally NSSA): A totally NSSA is similar to an NSSA. The difference is that the ABR in a totally NSSA prevents Type 3 LSAs from entering the totally NSSA. In this scenario, R3 does not inject inter-area routes into area 2. Therefore, in the LSDB of R5, there is only one Type 3 LSA that carries the default route.

OSPF LSA Review

Type	Name	Description
1	Router-LSA	Every router on an OSPF network generates Router-LSAs. A Router-LSA describes a router's link state and cost and can be flooded only in the area that the interface belongs to.
2	Network-LSA	A DR generates Network-LSAs. A Network-LSA describes all the routers that establish adjacencies with the DR on the MA network to which the DR is connected and the DR itself. The LSA can be flooded only in the area to which the interface belongs.
3	Network-summary-LSA	An ABR generates Network-summary-LSAs. A Network-summary-LSA describes the route to the destination network segment of an area. It is used to transmit inter-area routes.
4	ASBR-summary-LSA	An ABR generates ASBR-summary-LSAs. An ASBR-Summary-LSA describes routes to an ASBR, and is equivalent to a host route to an ASBR.
5	AS-external-LSA	An ASBR generates AS-external-LSAs. An AS-external-LSA describes routes to destinations outside an AS.
7	NSSA LSA	An ASBR generates NSSA LSAs. An NSSA LSA describes routes to destinations outside an AS. NSSA LSAs have similar functions as AS-external-LSAs, but are flooded in different areas. NSSA LSAs can be flooded only in the NSSA and cannot enter area 0. The ABR in the NSSA converts Type 7 LSAs into Type 5 LSAs and injects them into area 0.

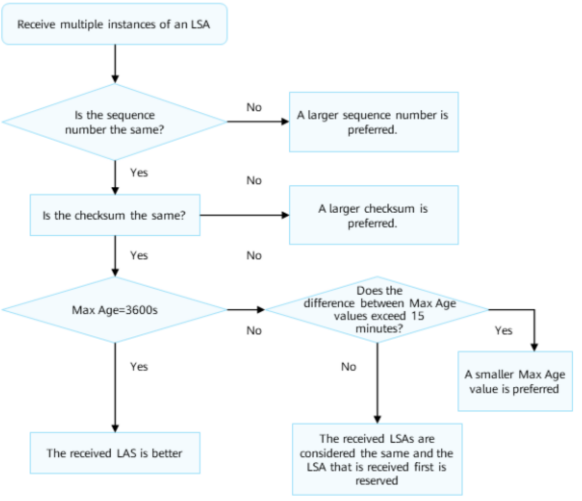
- The use of special areas reduces the LSDB size on the device, thus preventing device performance from deteriorating and minimizing the impact of network faults.
- How to reduce LSA flooding and LSDB size in a common area to optimize the OSPF network while ensuring IP reachability?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

LSA Processing on a Router

OSPF synchronizes LSDBs by exchanging LSAs. After receiving LSAs, a router processes the LSAs according to the following principles:

- If the received LSA does not exist in the local LSDB, the router updates the LSDB and floods the LSA.
- If the LSA already exists in the local LSDB but an update is received, the router updates the LSDB and floods the LSA.
- If the received LSA is the same as that in the LSDB, the router ignores the LSA and does not flood it.
- If the received LSA is damaged, for example, the checksum is incorrect, the router does not accept it.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

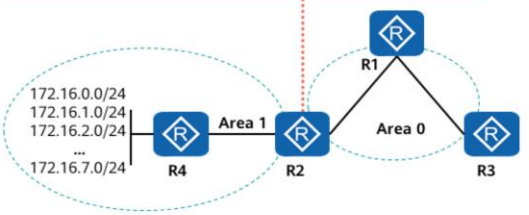
1. Stub Area and Totally Stub Area
2. NSSA and Totally NSSA
3. **Summarization of Inter-Area and External Routes**
4. OSPF Features

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Summarization on an ABR

- Route summarization indicates that a group of routes with the same prefix is summarized into one route to reduce the size of the routing table and optimize the usage of device resources. The routes before summarization are called refined or specific routes, the route summarized by the group of routes is called summarized route.
- OSPF route summarization:
  - Route summarization on an ABR: Inter-area routes are summarized.
  - Route summarization on an ASBR: Imported external routes are summarized.

```
[R2]ospf
[R2-ospf-1]area 1
[R2-ospf-1-area-0.0.0.1]abr-summary 172.16.0.0 255.255.248.0
```




```
<R2>display  ospf lsdb
            area 0
Type       LinkState ID    AdvRouter
Sum-Net    172.16.0.0      10.0.2.2
Sum-Net    172.16.1.0      10.0.2.2
Sum-Net    172.16.2.0      10.0.2.2
Sum-Net    172.16.3.0      10.0.2.2
Sum-Net    172.16.4.0      10.0.2.2
Sum-Net    172.16.5.0      10.0.2.2
Sum-Net    172.16.6.0      10.0.2.2
Sum-Net    172.16.7.0      10.0.2.2
```

Before route summarization

```
<R2>display  ospf lsdb
            area 0
Type          LinkState ID
AdvRouter
Sum-Net       172.16.0.0
10.0.2.2
```

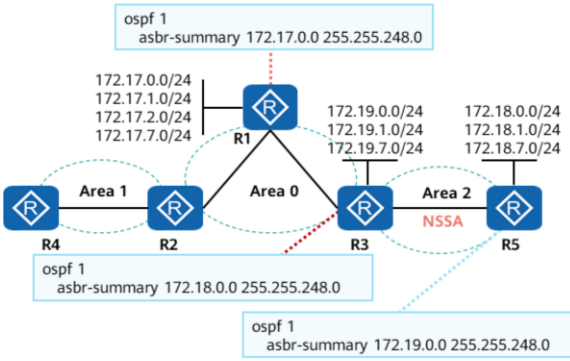
After route summarization

- After route summarization, R2 (ABR) advertises only the summarized route 172.16.0.0/21 to area 0.
- If the network segment corresponding to a specific route flaps (Up/Down), the impact of the topology change is limited to area 1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Route Summarization on an ASBR

- After route summarization is configured on an ASBR, the ASBR summarizes the external routes imported by itself.
- An ASBR in an NSSA can also summarize the external routes imported into the NSSA.
- In an NSSA, the ABR translates Type 7 LSAs to Type 5 LSAs. In this case, the ABR is also an ASBR. If route summarization is configured, the ABR summarizes Type 5 LSAs converted from Type 7 LSAs.



<R3>display ospf lsdb

OSPF Process 1 with Router ID 10.0.3.3

Area: 0.0.0.2

Type	LinkStateID	AdvRouter
NSSA	0.0.0.0	10.0.3.3
NSSA	172.19.0.0	10.0.3.3
NSSA	172.18.0.0	10.0.5.5

Area: 0.0.0.0

Type	LinkState ID	AdvRouter
External	172.19.0.0	10.0.3.3
External	172.18.0.0	10.0.3.3
External	172.17.0.0	10.0.1.1

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- R1, R3, and R5 summarize the imported external routes.

Contents

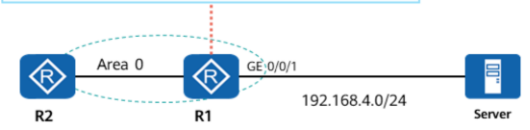
1. Stub Area and Totally Stub Area
2. NSSA and Totally NSSA
3. Summarization of Inter-Area and External Routes
4. **OSPF Features**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Silent Interface

- This can enhance the networking adaptability of OSPF and reduce system resource consumption.
- A silent interface has the following features:
  - A silent interface does not receive or send OSPF packets.
  - Direct routes of a silent interface can still be advertised.

```
[R1]ospf
[R1-ospf-1]area 0
[R1-ospf-1-area-0.0.0.0]network 192.168.4.0 0.0.0.255
[R1-ospf-1]silent-interface GigabitEthernet 0/0/1
```




```
<R1>display ospf interface GigabitEthernet 0/0/1

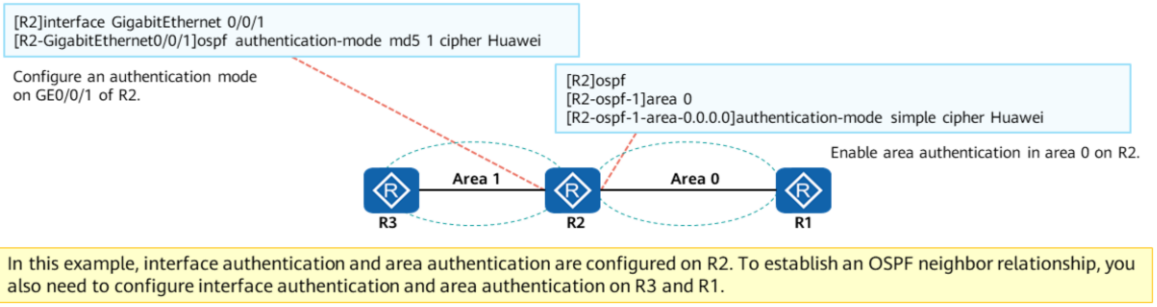
                OSPF Process 1 with Router ID 10.0.1.1
Interface: 10.0.13.1 (GigabitEthernet0/0/1)
Cost: 1        State: Waiting    Type: Broadcast    MTU: 1500
Priority: 1
Designated Router: 0.0.0.0
Backup Designated Router: 0.0.0.0
Timers: Hello 10 , Dead 40 , Poll  120 , Retransmit 5 , Transmit Delay 1
Silent interface, No hellos
```

- GE0/0/1 of R1 is on the network segment configured using the **network** command.
- The route to the interface is advertised to OSPF so that other devices can access the server.
- The interface is not connected to any other OSPF router, so the administrator configures the interface as a silent interface. The interface does not send or receive Hello packets, which prevents the performance of the server from deteriorating.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

OSPF Packet Authentication

- OSPF supports packet authentication. Only the OSPF packets that pass the authentication can be received.
- The router supports two OSPF packet authentication modes. When both modes are available, the interface authentication mode is preferred.
  - Area authentication: The authentication mode and password on all routers in the same area must be the same.
  - Interface authentication: The authentication mode and password of the interfaces directly connected to neighboring routers must be the same.



In this example, interface authentication and area authentication are configured on R2. To establish an OSPF neighbor relationship, you also need to configure interface authentication and area authentication on R3 and R1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- In the OSPF area view, configure an authentication mode for the OSPF area.
  - Run **authentication-mode simple** [**plain** *plain-text* | [**cipher**] *cipher-text*] to configure an authentication mode for the OSPF area.
    - **plain**: plain-text password
    - **cipher**: cipher-text password For MD5 or HMAC-MD5 authentication, the cipher-text mode is used by default.
- Configure an authentication mode on an interface.
  - Run **ospf authentication-mode simple** [**plain** *plain-text* | [**cipher**] *cipher-text*] to configure an authentication mode on an OSPF interface.

Quiz

1. (Multiple) Which of the following special areas are defined by OSPF?
   A. Stub Area
   B. Totally Stub Area
   C. Not-So-Stubby Area (NSSA)
   D. Totally NSSA
2. (Essay) What is the difference between a stub area and a totally stub area?
3. (Essay) Which router is configured with the inter-area route summarization function?

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. ABCD
2. A stub area does not allow Type 4 and Type 5 LSAs, but allows Type 3 LSAs. A totally stub area does not allow Type 4 and Type 5 LSAs or Type 3 LSAs. It allows only Type 3 LSAs that carry default routes.
3. It is configured on the ABR.

Summary

- This chapter describes OSPF special areas. Routers in the stub area reach the external network through the default route. Routers in the totally stub area reach the external network and the OSPF inter-area network through the default route. The routers in the NSSA and totally NSSA can import external routes.
- This chapter also describes OSPF features, including the silent interface, authentication, and processing mode when an OSPF router receives LSAs.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



IS-IS Implementation and Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- Intermediate System-to-Intermediate System (IS-IS) is a dynamic routing protocol initially designed by the International Organization for Standardization (ISO) for its Connectionless Network Protocol (CLNP).
- To support IP routing, the Internet Engineering Task Force (IETF) extended and modified IS-IS in RFC 1195. This modification enables IS-IS to apply to TCP/IP and OSI environments. This type of IS-IS is called Integrated IS-IS. IS-IS stated in this course refers to Integrated IS-IS, unless otherwise stated.
- This course describes the basic concepts, working mechanism, and configuration methods of IS-IS.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe basic concepts of IS-IS.
  - Understand how IS-IS is implemented.
  - Describe the differences between IS-IS and OSPF.
  - Perform basic IS-IS configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Basic Concepts of IS-IS**
2. Working Mechanism of IS-IS
3. Basic IS-IS Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Overview of IS-IS

IS-IS is a part of the connectionless network service (CLNS) in the OSI protocol stack defined by the ISO.

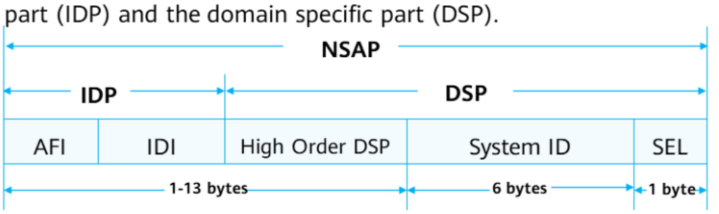


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- IS-IS is a link-state routing protocol. IS-IS is similar to OSPF in many aspects. For example, directly connected devices running IS-IS discover each other by sending Hello packets, establish adjacencies, and exchange link-state information.
- A NET includes the following elements:
  - CLNP: is similar to the IP protocol in TCP/IP.
  - IS-IS: is similar to OSPF in TCP/IP.
  - ES-IS: is similar to ARP or ICMP in TCP/IP.
- End system (ES): is similar to a host on the IP network.
- ES-IS: End System to Intermediate System
- CLNP and ES-IS are not involved on the IP network, so they are not described in this course.

NSAP

- The network service access point (NSAP) is an address defined by the OSI to locate resources. It provides interfaces between the network layer and upper-layer applications. The NSAP is composed of the initial domain part (IDP) and the domain specific part (DSP).

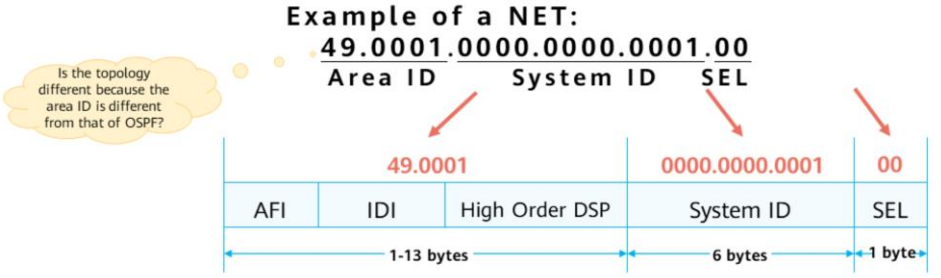


- The IDP is similar to the network ID in an IP address. It is defined by the ISO and consists of the authority and format identifier (AFI) and the initial domain identifier (IDI). The AFI indicates the address allocation authority and address format, and the IDI identifies a domain.
- The DSP is similar to the subnet ID and host address in an IP address. The DSP consists of the High Order DSP (HODSP), system ID, and NSAP Selector (SEL). The HODSP is used to divide areas, the system ID identifies a host, and the SEL indicates the service type.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

NET

- A network entity title (NET) indicates the network layer information of an IS and consists of an area ID and a system ID. It is used for route calculation. A NET can be considered as a special NSAP (the NSAP whose SEL is 00).
- The length of the NET field is the same as that of an NSAP. Its maximum length is 20 bytes and its minimum length is 8 bytes.
- When IS-IS runs on an IP network, you only need to configure NETs. Devices can obtain area IDs and system IDs based on NETs.

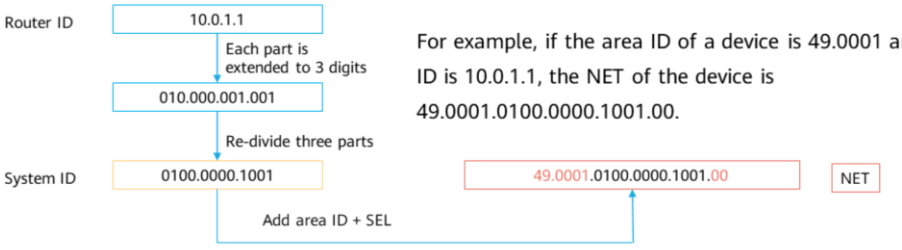


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- An area ID consists of the IDP and HODSP in the DSP. It can identify a routing domain and an area in the routing domain. Therefore, they are called an area address, which is equivalent to the area number in OSPF.
    - In most cases, a router can be configured with only one area address. The area address of all nodes in an area must be the same. To support seamless combination, division, and transformation of areas, a device can be configured with a maximum of three area addresses in an IS-IS process by default.
- A system ID uniquely identifies a host or a router in an area. The fixed length of the system ID on the device is 6 bytes.

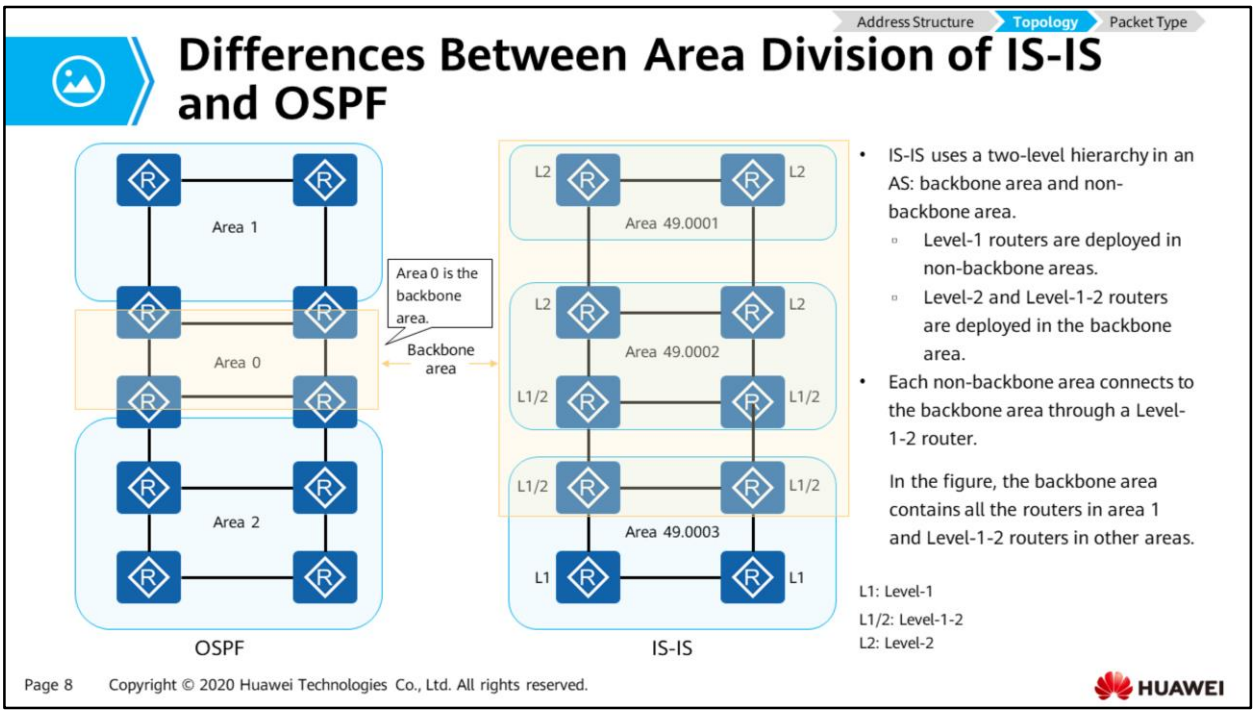
Example of a NET

- Each network device running IS-IS must have at least one NET. A device can be configured with multiple NETs, but the system IDs of these NETs must be the same.
- On Huawei network devices, the length of the system ID is always 6 bytes. In an IS-IS routing domain, the system ID of a device must be unique. To facilitate management, configure the system ID based on the router ID.



For example, if the area ID of a device is 49.0001 and the router ID is 10.0.1.1, the NET of the device is 49.0001.0100.0000.1001.00.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

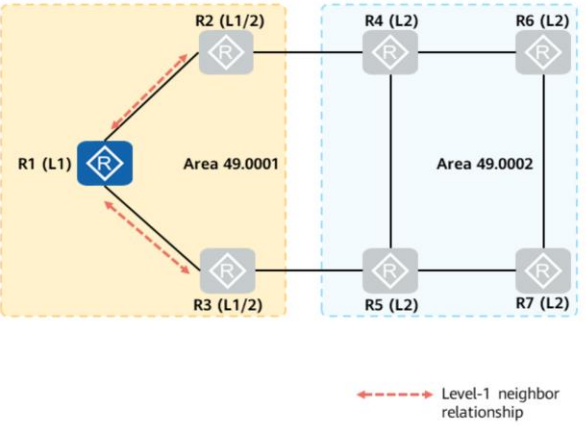


- During the learning of OSPF, we have learned the advantages of multi-area and hierarchical network design. For a link-state routing protocol, a device running the protocol advertises link state information to the network, collects and stores the flooded link-state information, and performs calculation based on the information to obtain routing information. If the multi-area deployment mode is not used, more and more link state information will be flooded on the network as the network scale increases. All devices on the network will bear heavier burdens, and route convergence will become slower. This also results in low network scalability.
- The two types of topologies show the differences between IS-IS and OSPF:
    - In IS-IS, each router belongs to only one area. In OSPF, different interfaces of a router may belong to different areas.
    - In IS-IS, no area is defined as the backbone area. In OSPF, area 0 is defined as the backbone area.
    - In IS-IS, Level-1 and Level-2 routes are calculated using the SPF algorithm to generate the shortest path tree (SPT). In OSPF, the SPF algorithm is used only in the same area, and inter-area routes are forwarded by the backbone area.

IS-IS Router Types (1)

Level-1 router

- A Level-1 router (for example, R1 in the figure) is an internal router in an IS-IS area. It establishes Level-1 neighbor relationships with only Level-1 and Level-1-2 routers in the same area. A Level-1 router cannot establish the neighbor relationship with a Level-2 router.
- A Level-1 router maintains only the Level-1 LSDB, which contains only the routing information of the local area. The Level-1 router must access the IS-IS backbone area through a Level-1-2 router to access other areas.

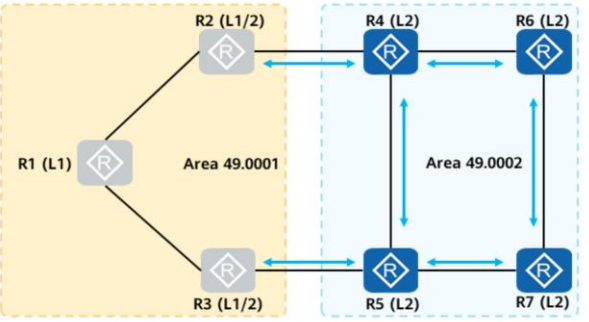


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IS-IS Router Types (2)

Level-2 router

- Level-2 routers (such as R4, R5, R6, and R7 in the figure) are IS-IS backbone routers. They can establish neighbor relationships with Level-2 routers or Level-1-2 routers in the same area or different areas. A Level-2 router maintains a Level-2 LSDB that contains all routing information of the IS-IS domain.
- All Level-2 routers form the backbone network of the routing domain. They establish Level-2 neighbor relationships and are responsible for inter-area communication. Level-2 routers in the routing domain must be physically contiguous to ensure the continuity of the backbone network.

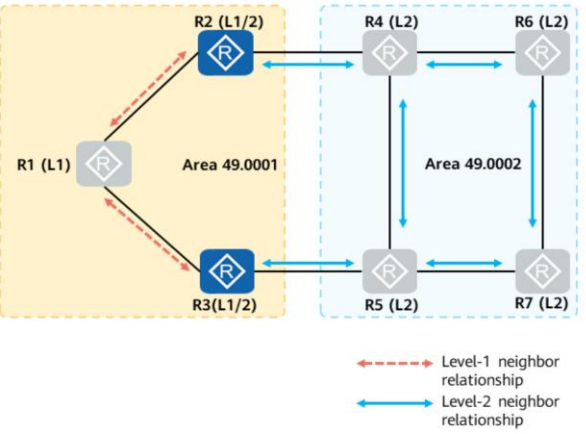


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

IS-IS Router Types (3)

Level-1-2 router

- Similar to an ABR in OSPF, a Level-1-2 router is also a part of the IS-IS backbone network.
- A Level-1-2 router maintains two LSDBs: a Level-1 LSDB and a Level-2 LSDB. The Level-1 LSDB saves for intra-area routing and the Level-2 LSDB saves for inter-area routing.
- A router that belongs to both a Level-1 area and a Level-2 area is called a Level-1-2 router (for example, R2 and R3). It can establish Level-1 neighbor relationships with Level-1 and Level-1-2 routers in the same area. It can also establish Level-2 neighbor relationships with Level-2 and Level-1-2 routers in different areas.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

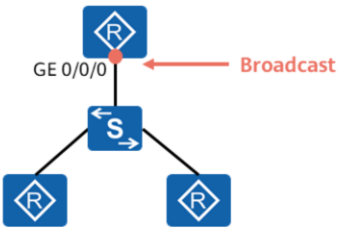
HUAWEI

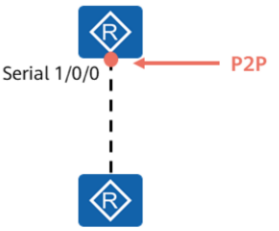
- When IS-IS is configured on a Huawei router, the router type is Level-1-2 by default. You can run commands to change the router type.

Network Types Supported by IS-IS

IS-IS automatically determines the default network type of an interface based on the data link layer encapsulation type of the interface. IS-IS supports the following network types:

- Broadcast: such as Ethernet.
- Point-to-point (P2P): such as PPP and HDLC



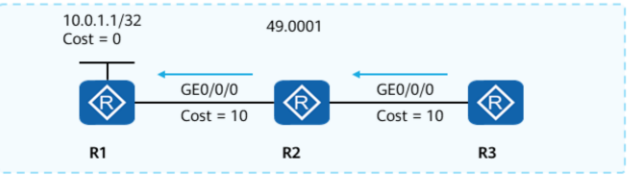


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- For a Non-Broadcast Multi-Access (NBMA) network, you should configure its sub-interfaces as P2P interfaces.

IS-IS Cost

- IS-IS uses the cost as the route metric. A smaller cost indicates a better path. The cost of an IS-IS link is related to an interface of the device. Similar to OSPF, each IS-IS-enabled interface maintains the interface cost. Unlike OSPF, the cost of an IS-IS-enabled interface is not related to the interface bandwidth by default. In actual deployment, IS-IS can adjust the cost based on the interface bandwidth. The default cost is 10 regardless of the interface bandwidth.
- The cost of an IS-IS path is the sum of the costs of all links from the local router to the destination network segment.
- The costs of IS-IS interfaces can be determined in the following modes (in descending order of priority):
  - Interface cost: configured for a specified interface.
  - Global cost: configured for all interfaces.
  - Automatically calculated cost: automatically calculated based on the interface bandwidth.



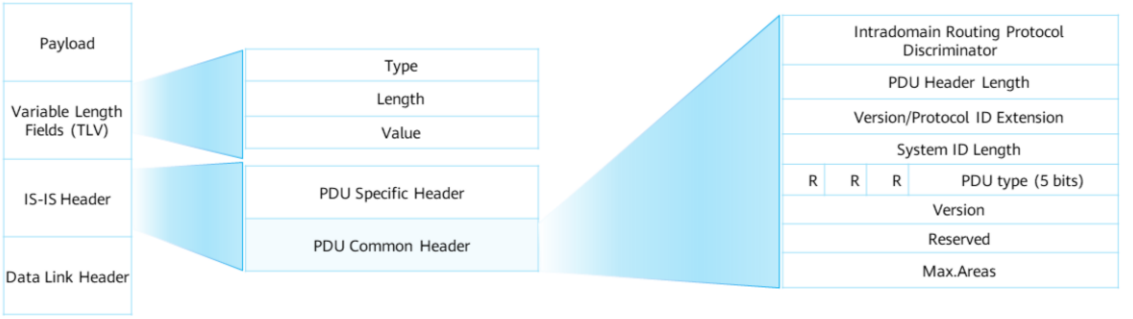
As shown in the figure, IS-IS runs on the entire network. The cost of the route from R3 to 10.0.1.1/32 is 20 (10 + 10 + 0).

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In ISO 10589, the maximum metric value of an IS-IS interface can only be 63 and the IS-IS cost type is narrow. A small range of metrics cannot meet the requirements on large-scale networks. As defined in RFC 3784, the cost of an IS-IS interface can be extended to 16777215. In this case, the IS-IS cost type is wide.
- By default, the cost type of Huawei routers is narrow.
- The following lists the TLVs used in narrow mode:
  - TLV 128 (IP Internal Reachability TLV): carries IS-IS routes in a routing domain.
  - TLV 130 (IP External Reachability TLV): carries IS-IS routes outside a routing domain.
  - TLV 2 (IS Neighbors TLV): carries neighbor information.
- The following lists the TLVs used in wide mode:
  - TLV 135 (Extended IP Reachability TLV): replaces the earlier IP reachability TLV and carries IS-IS routing information. This TLV expands the route metric and carries sub-TLVs.
  - TLV 22 (IS Extended Neighbors TLV): carries neighbor information.

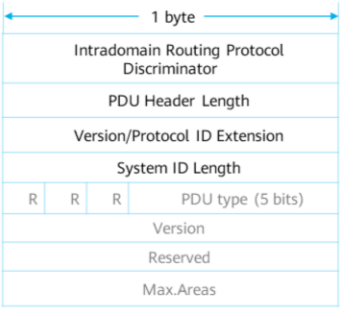
IS-IS Packet Format

- IS-IS packets are encapsulated in frames at the data link layer.
- A protocol data unit (PDU) consists of an IS-IS header and variable length fields.
- The IS-IS header can be further classified into the common header and specific header. For all PDUs, the common headers are the same, but the dedicated headers vary according to the PDU type.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Description of the PDU Common Header



Common header

Main fields:

- Intradomain Routing Protocol Discriminator: The fixed value is 0x83.
- Length Indicator: indicates the length of the IS-IS header (including the common header and dedicated header), in bytes.
- Version/Protocol ID Extension: The fixed value is 0x01.
- System ID Length: indicates the length of the system ID in the NSAP or NET. If it is set to 0, the system ID has 6 bytes.
- R (Reserved): The fixed value is 0.
- Version: The fixed value is 0x01.
- Maximum Area Address: indicates the maximum number of the addresses of supported areas. An integer value ranging from 1 to 254 indicates the maximum number of area addresses allowed by an IS-IS process. The default value 0 indicates that the IS-IS process supports a maximum of three area addresses.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IS-IS Packet Types

- IS-IS PDUs are classified into four types: IS-IS Hello PDUs (IIH), link state PDUs (LSPs), Complete Sequence Number PDUs (CSNPs), and Partial Sequence Number PDU (PSNPs).
- IIHs: are used to set up and maintain neighbor relationships. Among them, Level-1 LAN IIHs apply to the Level-1 routers on broadcast LANs; Level-2 LAN IIHs apply to the Level-2 routers on broadcast LANs; and P2P IIHs apply to non-broadcast networks.
- LSPs: are used to exchange link-state information. There are two types of LSPs: Level-1 and Level-2 LSPs.
- SNPs: describe the LSPs in all or some databases to help synchronize and maintain all LSDBs. SNPs are classified into CSNPs and PSNPs, which can be further classified into Level-1 CSNPs, Level-2 CSNPs, Level-1 PSNPs, and Level-2 PSNPs.



Type Value	Description
15	L1 LAN IIH
16	L2 LAN IIH
17	P2P IIH
18	L1 LSP
20	L2 LSP
24	L1 CSNP
25	L2 CSNP
26	L1 PSNP
27	L2 PSNP

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Common IS-IS TLVs

- A TLV is a data structure that consists of the Type, Length, and Value fields.
- The TLV structure features high flexibility and scalability in packet construction. The TLV is used to fix the overall structure of a packet. To add a new feature, you only need to add a new TLV without changing the overall structure of the packet.

TLV Type	Name	PDU Type
1	Area Addresses	IIH, LSP
2	IS Neighbors (LSP)	LSP
4	Partition Designated Level2 IS	L2 LSP
6	IS Neighbors (MAC Address)	LAN IIH
7	IS Neighbors (SNPA Address)	LAN IIH
8	Padding	IIH
9	LSP Entries	SNP
10	Authentication Information	IIH, LSP, SNP
128	IP Internal Reachability Information	LSP
129	Protocols Supported	IIH, LSP
130	IP External Reachability Information	LSP
131	Inter-Domain Routing Protocol Information	L2 LSP
132	IP Interface Address	IIH, LSP

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A TLV is also called a code-length-value (CLV).

Contents

1. Basic Concepts of IS-IS
2. **Working Mechanism of IS-IS**
   - **Neighbor Relationship Setup**
   - LSDB Synchronization
   - Route Calculation
3. Basic IS-IS Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Rules for Establishing IS-IS Neighbor Relationships

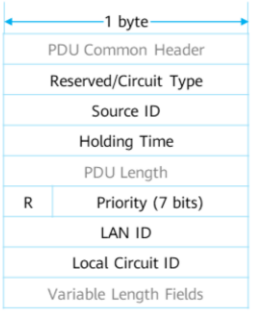
- Basic rules for establishing an IS-IS neighbor relationship are as follows:
  - Only neighboring routers of the same level can set up the neighbor relationship with each other.
  - Level-1 routers must have the same area IDs.
  - Network types of IS-IS interfaces on both ends of a link must be consistent.
  - IP addresses of IS-IS interfaces on both ends of a link must be on the same network segment by default.
- IS-IS runs on the data link layer and was initially designed for CLNP. Therefore, the establishment of IS-IS neighbor relationships is not related to IP addresses. When IS-IS runs over IP, IS-IS needs to check the IP address of its neighbor. If secondary IP addresses are assigned to the interfaces, the routers can set up the IS-IS neighbor relationship as long as either the primary or secondary IP address is on the same network segment as the local and remote interfaces.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- By simulating an Ethernet interface as a P2P interface, a router can establish a P2P neighbor relationship.
- When IP addresses of IS-IS interfaces on both ends of a link are on different network segments, a neighbor relationship can still be established on the two interfaces if the interfaces are configured not to check the IP addresses in received Hello packets.
  - For P2P interfaces, you can configure the interfaces not to check the IP addresses.
  - For Ethernet interfaces, you must simulate Ethernet interfaces as P2P interfaces and then configure the interfaces not to check the IP addresses.
- Generally, one interface only needs one primary IP address. In some special cases, one interface needs additional secondary IP addresses. For example, a router connects to a physical network through an interface, and hosts on this network belong to two network segments. To enable the router to communicate with all hosts on the physical network, configure a primary IP address and a secondary IP address for this interface. You can configure multiple IP addresses for a Layer 3 interface on a router, one as the primary IP address, and the others as secondary IP addresses. Each Layer 3 interface can have a maximum of 31 secondary IP addresses.

IIH

IIHs are used to set up and maintain neighbor relationships. Among them, Level-1 LAN IIHs apply to the Level-1 routers on broadcast LANs; Level-2 LAN IIHs apply to the Level-2 routers on broadcast LANs; and P2P IIHs apply to non-broadcast networks.



Reserved/Circuit Type: indicates the type of the router (01 indicates L1, 10 indicates L2, and 11 indicates L1/L2).

Source ID: indicates the system ID of the router that sends Hello packets.

Holding Time: If no Hello packet is received from the neighbor within the holding time, the neighbor relationship is terminated.

Priority: specifies the priority of the Designated Intermediate System (DIS). The value ranges from 0 to 127. A larger value indicates a higher priority. This field is carried only in LAN IIHs on a broadcast network. P2P IIHs on a P2P network do not have this field or the R bit preceding this field.

LAN ID: includes the system ID and pseudonode ID of the DIS. This field is carried only in LAN IIHs on a broadcast network. P2P IIHs on a P2P network do not contain this field.

Local Circuit ID: This field is carried only in P2P IIHs of the P2P network. LAN IIHs on a broadcast network do not have this field.

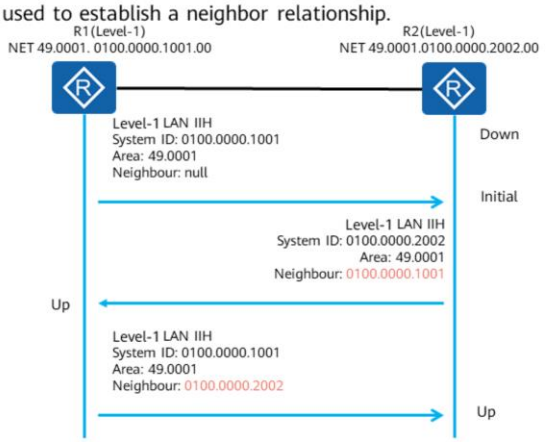
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Neighbor Relationship Setup | LSDB Synchronization | Route Calculation

Process of Establishing an IS-IS Neighbor Relationship on a Broadcast Network

Two IS-IS routers must establish a neighbor relationship before exchanging packets to implement routing. The modes for establishing IS-IS neighbors vary depending on the network type. On a broadcast network, three-way handshake is used to establish a neighbor relationship.



R1 and R2 are connected through GE interfaces. The process of establishing a neighbor relationship between the two directly connected Level-1 routers is as follows:

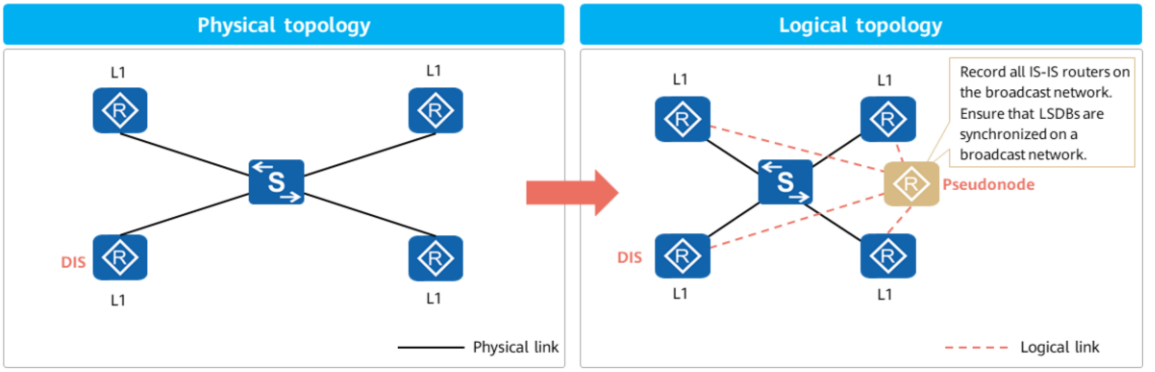
1. In Down state, R1 multicasts a Level-1 LAN IIH. The neighbor list in the packet is empty.
2. After receiving the packet, R2 sets the neighbor relationship status to Initial. R2 then responds to R1 with a Level-1 LAN IIH, indicating that R1 is a neighbor of R2.
3. R1 receives this packet and sets the status of the neighbor relationship with R2 to Up. R1 then sends R2 a Level-1 LAN IIH that identifies R2 used as a neighbor of R1.
4. R2 receives this packet and sets the status of the neighbor relationship with R1 to Up. R1 and R2 establish a neighbor relationship successfully.
5. The network is a broadcast network, so a DIS needs to be elected. After the neighbor relationship is established, routers wait for two intervals before sending IIHs to elect the DIS.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Multicast addresses of Level-1 and Level-2 IIHs are 01-80-C2-00-00-14 and 01-80-C2-00-00-15 respectively.
- Down: It is the initial status of the neighbor relationship.
- Initial: The IIH is received, but the neighbor list in the packet does not contain the system ID of the router.
- UP: The IIS is received and the neighbor list contains the system ID of the router.

DIS and Pseudonode

- In a broadcast network, IS-IS needs to elect a DIS from all the routers.
- DISs are used to create and update pseudonodes and generate LSPs of pseudonodes to describe available network devices. The pseudonode is used to simulate a virtual node on the broadcast network and is not an actual router. In IS-IS, a pseudonode is identified by the system ID of the DIS and Circuit ID (its value is not 0).



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

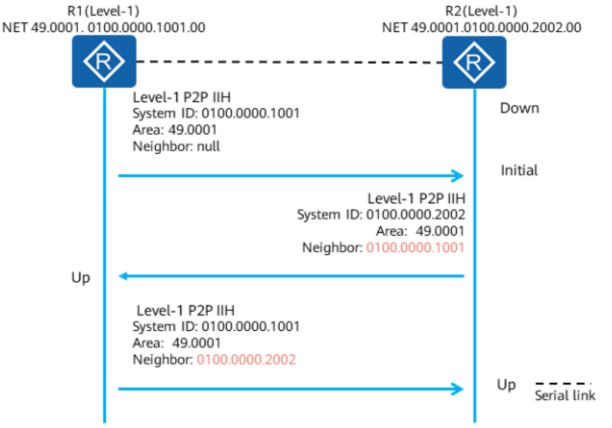
DIS in IS-IS and DR in OSPF

- Level-1 and Level-2 DISs are elected separately. You can configure different priorities for DISs of different levels.
- DIS election rules are as follows:
  - The router with the highest priority is elected as the DIS.
  - If there are multiple routers with the same highest priority on a broadcast network, the one with the highest MAC address is chosen.
- The interval for the DIS to send IIHs is one third of the interval for a common router. This ensures that the faulty DIS can be quickly discovered.
- DIS election in IS-IS differs from designated router (DR) election in OSPF:
  - On an IS-IS broadcast network, the router with priority 0 also takes part in DIS election. In OSPF, the router with priority 0 does not take part in DR election.
  - In IS-IS, when a new router that is qualified for being a DIS connects to a broadcast network, the router is elected as the new DIS and the previous pseudonode is deleted. This causes a new flooding of LSPs. In OSPF, when a new router connects to a network, it is not immediately elected as the DR even if it has the highest DR priority.
  - On an IS-IS broadcast network, routers (including non-DIS routers) of the same level on a network segment set up neighbor relationships. In OSPF, routers set up neighbor relationships with only the DR and backup designated router (BDR).

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Process of Establishing a Neighbor Relationship on a P2P Network

- On a P2P network, the two-way handshake mechanism is used to establish a neighbor relationship. When a router receives an IIH from the peer, the router declares the neighbor relationship status to be Up and establishes a neighbor relationship.
- The two-way handshake mechanism has defects. When IS-IS is used on a P2P network, Huawei devices use three-way handshake to establish neighbor relationships by default. In this mode, three P2P IIHs are sent to establish a neighbor relationship.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

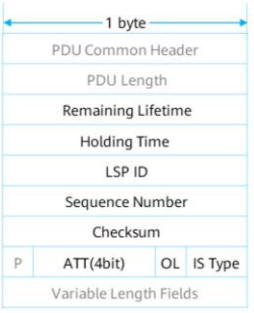
1. Basic Concepts of IS-IS
2. **Working Mechanism of IS-IS**
   - Neighbor Relationship Setup
   - **LSDB Synchronization**
   - Route Calculation
3. Basic IS-IS Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

LSP

- LSPs are used to exchange link state information. There are two types of LSPs: Level-1 and Level-2 LSPs. A Level-1 router transmits Level-1 LSPs; a Level-2 router transmits Level-2 LSPs; and a Level-1-2 router can transmit both Level-1 and Level-2 LSPs.
- Level-1 and Level-2 LSPs have the same format.



Remaining Lifetime: indicates the lifetime of an LSP, in seconds.

LSP ID: consists of the system ID, pseudonode ID, and LSP fragment number.

Sequence Number: indicates the LSP sequence number. When a router is started, the sequence number of the first LSP sent by the router is 1. When a new LSP is generated, the sequence number of the LSP is equal to the sequence number of the previous LSP plus 1. Therefore, newer LSPs have larger sequence numbers.

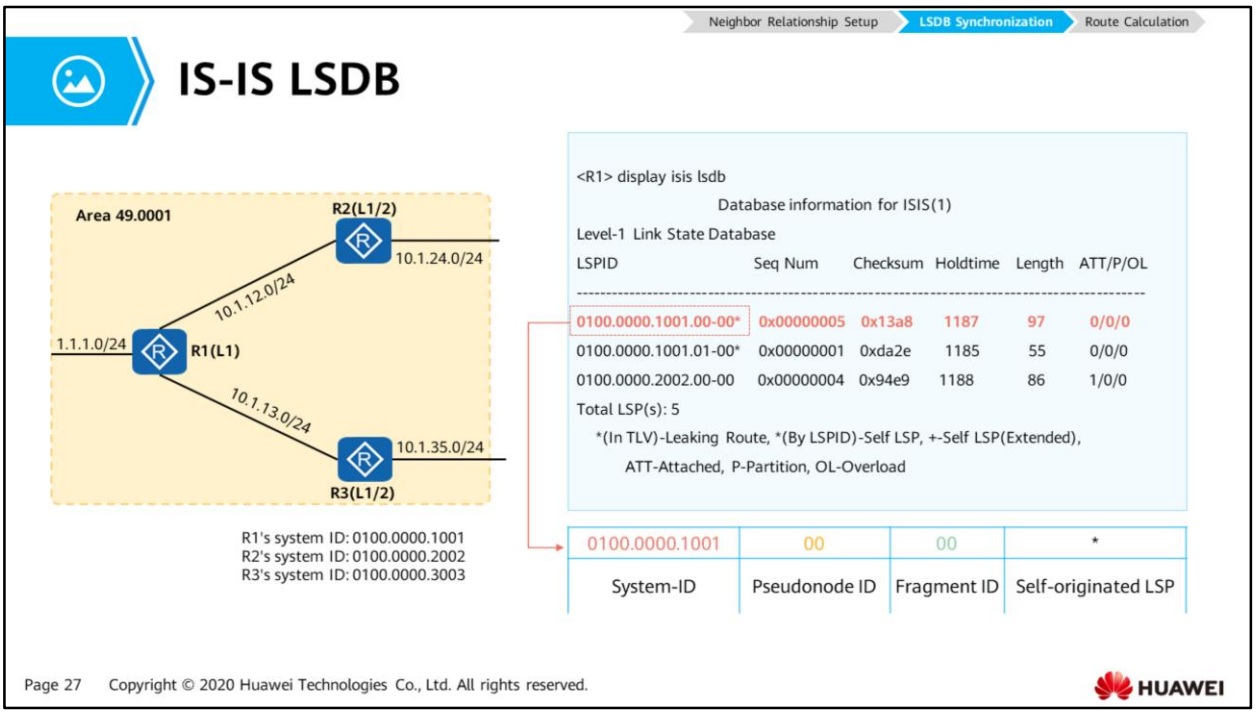
Checksum: indicates the checksum of an LSP.

ATT (Attachment): is generated by a Level-1-2 router to identify whether the originating router is connected to other areas. Though this field is also included in Level-1 or Level-2 LSPs sent by Level-1 or Level2 routers, it is associated only with Level-1 LSPs sent by Level-1-2 routers.

OL (1 bit): indicates the LSDB overload. LSPs with the overload bit are still flooded on the network, but these LSPs are ignored during the calculation of the routes that pass through a router in overload state. After the overload bit is set on a router, other routers ignore the router when performing SPF calculation. If a router is experiencing a memory shortage, it automatically configures the overload bit in LSPs to be sent.

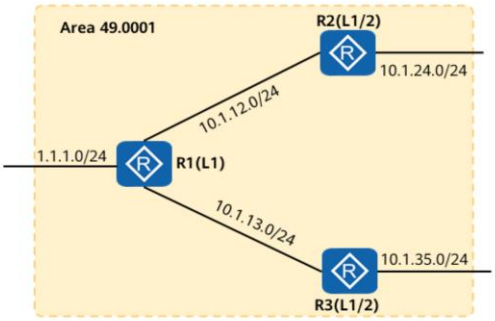
IS Type (2 bits): indicates the type of the router generating the LSP. It is used to specify whether the router is a Level-1 router or a Level-2 router (01 indicates Level-1; 11 indicates Level-2).

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Pseudonode ID: If the value of this parameter is not 0, the LSP is generated by a pseudonode.
- Fragment ID: When an IS-IS router needs to advertise the LSPs that contain much information, the IS-IS router generates multiple LSP fragments to carry more IS-IS information. The fragment ID is used to distinguish different LSP fragments.

Checking the LSP of the Non-pseudonode

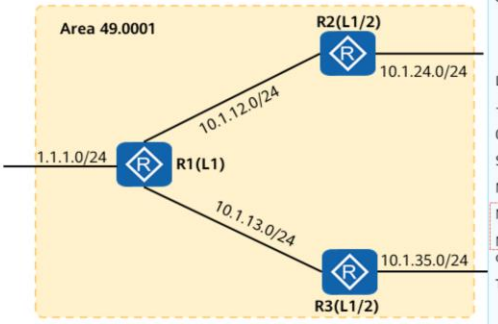



```
<R1> display isis lsdb 0100.0000.1001.00-00 verbose
                Database information for ISIS(1)
Level-1 Link State Database
LSPID                 Seq Num     Checksum  Holdtime    Length  ATT/P/OL
---------------------------------------------------------------------------
0100.0000.1001.00-00* 0x0000000e  0x9a75    1072        113     0/0/0
  SOURCE              0100.0000.1001.00
  NLPID               IPV4
  AREA ADDR           49.0123
  INTF ADDR           10.1.12.1                          //Describe interface information.
  INTF ADDR           10.1.13.1
  NBR ID              0100.0000.1001.01      COST: 10    //Describe the neighbor relationship.
  NBR ID              0100.0000.3003.01      COST: 10
  IP-Internal         10.1.12.0 255.255.255.0 COST: 10
  IP-Internal         10.1.13.0 255.255.255.0 COST: 10   //Describe the route information.
  IP-Internal         1.1.1.0 255.255.255.0  COST: 10
Total LSP(s): 1
```

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- AREA ADDR: indicates the area ID of the device that generates the LSP.
- INTF ADDR: indicates the interface address described in the LSP.
- NBR ID: indicates the neighbor information described in the LSP.
- IP-Internal: indicates the network segment information described in the LSP.

Checking the LSP of a Pseudonode




```
<R1> display isis lsdb 0100.0000.1001.01-00 verbose
                Database information for ISIS(1)
                Level-1 Link State Database
LSPID                 Seq Num     Checksum    Holdtime    Length    ATT/P/OL
---------------------------------------------------------------------------
0100.0000.1001.01-00* 0x00000009  0xca36      431         55        0/0/0
SOURCE                0100.0000.1001.01
NLPID                 IPV4
NBR ID                0100.0000.1001.00 COST: 0  //Describe all IS-IS routers on the broadcast network.
NBR ID                0100.0000.2002.00 COST: 0  //The cost of the route from the pseudonode to
other routers is 0.
Total LSP(s): 1
  *(In TLV)-Leaking Route, *(By LSPID)-Self LSP, +-Self LSP(Extended),
      ATT-Attached, P-Partition, OL-Overload
```

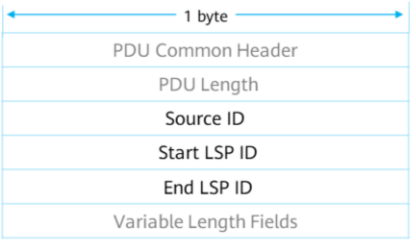
The LSP of the pseudonode contains only neighbor information but not routing information.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

CSNP

A CSNP contains the summary of all LSPs in the LSDB of a router. The router determines whether to synchronize its LSDB by exchanging CSNPs.

- On a broadcast network, CSNPs are sent by a DIS periodically (at a default interval of 10 seconds).
- On a P2P link, CSNPs are sent only during initial establishment of a neighbor relationship.



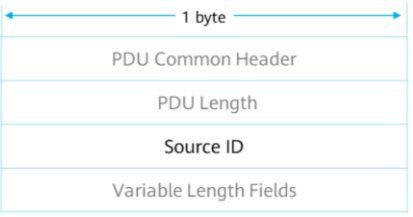
Source ID: indicates the system ID of the router that sends the CSNP.
Start LSP: indicates the ID of the first LSP in a CSNP.
End LSP: indicates the ID of the last LSP in a CSNP.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PSNP

Unlike a CSNP, a PSNP contains only the summary of some LSPs.

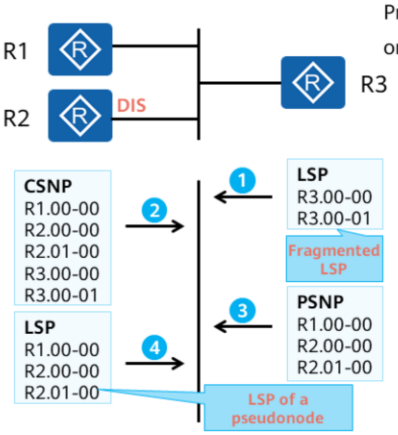
- When the LSDB is not synchronized, the PSNP is used to request the neighbor to send a new LSP.
- On a P2P network, when an LSP is received, a PSNP is used to acknowledge the received LSP.



Source ID: indicates the system ID of the router that sends the PSNP.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

LSP Synchronization on a Broadcast Network



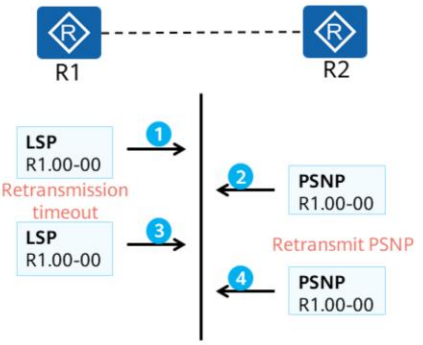
On a broadcast network, CSNPs are sent only by the DIS in multicast mode.

Process of synchronizing LSDBs between a new router and the DIS on a broadcast network:

1. The new router R3 sends IIHs to establish neighbor relationships with the other routers in the broadcast domain. R3 establishes neighbor relationships with R1 and R2, waits for the timeout of the LSP refresh timer, and then sends its LSP to a multicast address (01-80-C2-00-00-14 for a Level-1 router and 01-80-C2-00-00-15 for a Level-2 router). All neighbors on the network can receive the LSP.
2. The DIS on the network segment adds the received LSP to its LSDB. After the CSNP timer expires, the DIS sends CSNPs to synchronize the LSDBs on the network.
3. R3 receives the CSNPs from the DIS, checks its LSDB, and sends a PSNP to request the LSPs it does not have.
4. The DIS receives the PSNP and sends R3 the required LSPs for LSDB synchronization.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

LSP Synchronization on a P2P Network



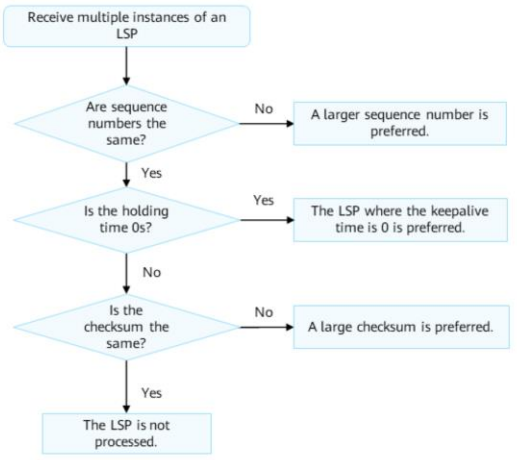
IS-IS LSDB synchronization on a P2P network:

- R1 establishes a neighbor relationship with R2.
- R1 and R2 send a CSNP to each other. If the LSDB of the neighbor and the received CSNP are not synchronized, the neighbor sends a PSNP to request the required LSP.
- Assume that R2 requests a corresponding LSP from R1.
  1. R1 sends the required LSP to R2, starts the LSP retransmission timer, and waits for a PSNP from R2 as an acknowledgement for the received LSP.
  2. If R1 does not receive the PSNP from the neighbor after the LSP retransmission timer expires, it resends the LSP.
  3. R1 resends the LSP.
  4. After receiving the LSP, R2 sends a PSNP for acknowledgment.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

LSP Processing Mechanism



- IS-IS synchronizes LSDBs by exchanging LSPs. After receiving LSPs, a router processes the LSPs according to the following rules:
- If the received LSP is better than the local LSP or the local device does not receive any LSP:
  - On a broadcast network, the router adds the LSP to its LSDB and multicasts the LSP.
  - On a P2P network, the router adds the LSP to its LSDB and sends a PSNP to acknowledge the receipt of the LSP. Then the router sends the new LSP to all its neighbors except the neighbor that sends the LSP.
- If the received LSP cannot be compared with the local LSP, the router does not process the LSP.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- All routers in the IS-IS routing domain can generate LSPs. The following events trigger the generation of a new LSP:
  - The neighbor is Up or Down.
  - The related interface goes Up or Down.
  - The imported IP routes change.
  - Inter-area IP routes change.
  - The interface is assigned a new metric value.
  - LSPs are updated periodically (update interval: 15 minutes).

Contents

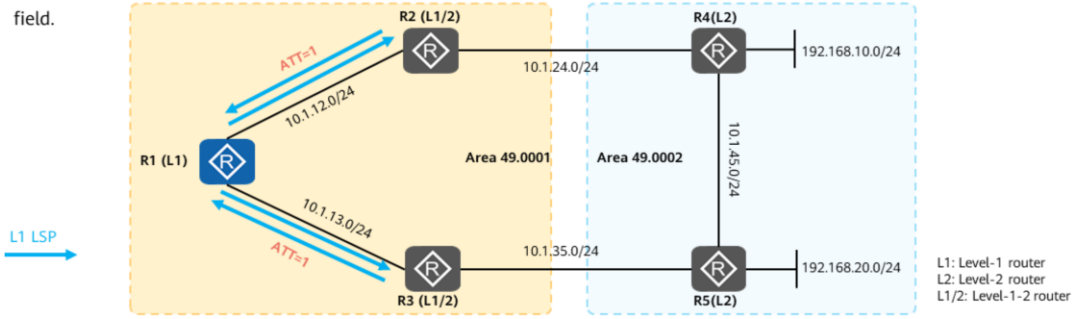
1. Basic Concepts of IS-IS
2. **Working Mechanism of IS-IS**
   - Neighbor Relationship Setup
   - LSDB Synchronization
   - **Route Calculation**
3. Basic IS-IS Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Route Calculation on a Level-1 Router

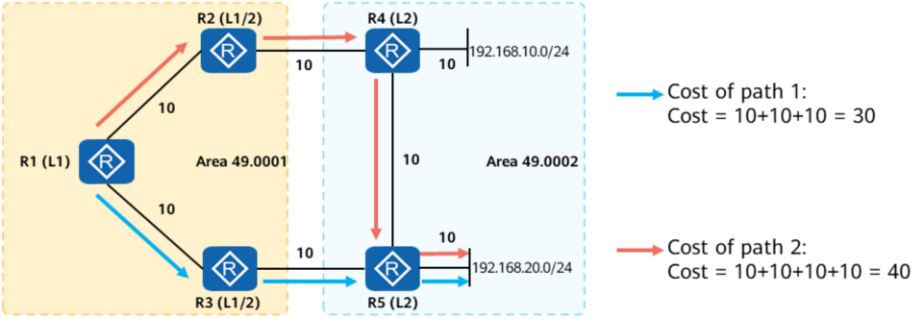
- R1 is a Level-1 router and maintains only the Level-1 LSDB. The LSDB contains Level-1 LSPs generated by R1 and level-1 LSPs of R2 and R3.
- R1 calculates the topology of Area 49.0001 and the routes to each network segment in the area based on the Level-1 LSPs in the LSDB.
- R2 and R3 are Level-1-2 routers in Area 49.0001. They set the ATT field in the Level-1 LSPs they send to the area to announce that they can reach other areas through themselves. As a Level-1 router, R1 calculates the default route to R2 or R3 based on the ATT field.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Sub-optimal Path of a Level-1 Router

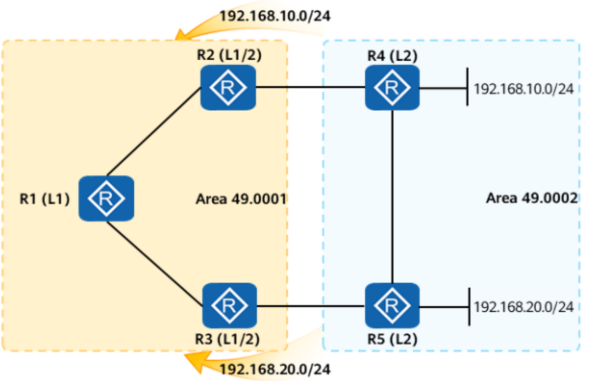
By default, R1 can reach the destination outside an area only through the default route pointing to R2 or R3. However, the costs of the routes from R1 to R2 and R3 are the same. Therefore, when R1 sends a data packet to 192.168.20.0/24, path 2 may be selected. As a result, the sub-optimal path problem occurs.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Leaking

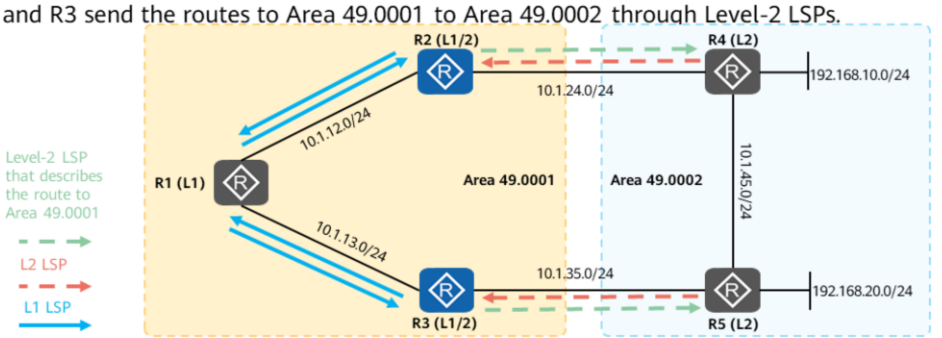
- By default, a Level-1-2 router does not advertise the routes to other areas to the local Level-1 area.
- Through route leaking, inter-area routes can be transmitted to the Level-1 area through the Level-1-2 router. In this case, the Level-1 router can learn the specific routes of other areas and calculate the optimal path.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Calculation on a Level-1-2 Router

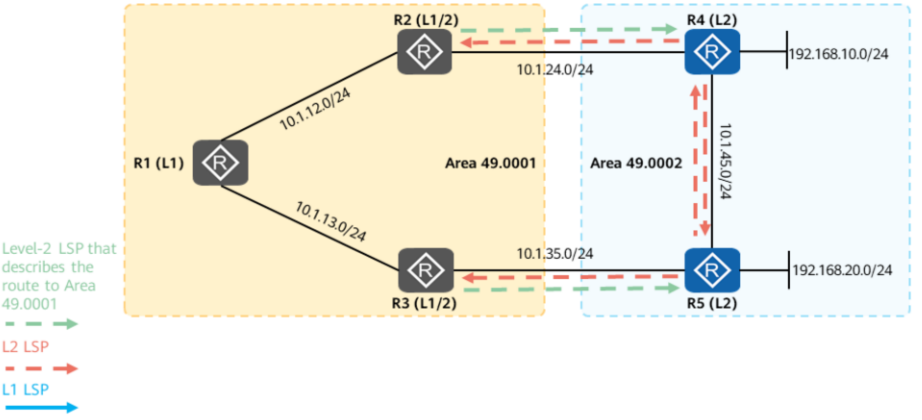
- R2 and R3 maintain Level-1 LSDBs and can calculate routes in Area 49.0001 based on LSPs in the LSDBs.
- R2 and R3 maintain Level-2 LSDBs and can calculate routes in Area 49.0002 based on LSPs in the LSDBs.
- R2 and R3 send the routes to Area 49.0001 to Area 49.0002 through Level-2 LSPs.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Route Calculation on a Level-2 Router

R4 and R5 used as Level-2 routers maintain only Level-2 LSDBs and can calculate routes to network segments on the entire network based on the LSDBs.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Basic Concepts of IS-IS
2. Working Mechanism of IS-IS
3. **Basic IS-IS Configurations**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic IS-IS Configurations (1)

1. Create an IS-IS process and enter the IS-IS process view.

   [Huawei] **isis** [*process-id*]

   process-id specifies the ID of an IS-IS process. If *process-id* is not specified, the default IS-IS process ID is 1.

2. Configure a NET.

   [Huawei-isis-1] **network-entity** *net*

   Generally, you only need to configure one NET for an IS-IS process. When an area needs to be redefined, for example, the area needs to be merged with other areas or divided into sub-areas, configure multiple NETs to ensure route correctness. A maximum of three area addresses can be configured in an IS-IS process, and therefore, you can configure only a maximum of three NETs. When you configure multiple NETs, ensure that their system IDs are the same.

3. Configure a device level.

   [Huawei-isis-1] **is-level** { **level-1** | **level-1-2** | **level-2** }

   By default, the level of the device is Level-1-2.
   If the levels of IS-IS devices are changed during network operation, the IS-IS process will be restarted, and IS-IS neighbor relationships will be disconnected. Setting the levels of devices when configuring IS-IS is recommended.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic IS-IS Configurations (2)

4. Enter the interface view.

   [Huawei]**interface** *interface-type interface-number*

   interface-type specifies the interface type and *interface-number* specifies the interface number.

5. Enable IS-IS on the interface.

   [Huawei-GigabitEthernet0/0/1] **isis enable** [*process-id*]

   After this command is run, IS-IS establishes neighbor relationships and floods LSPs through this interface.

6. Configure the level of the interface.

   [Huawei-GigabitEthernet0/0/1] **isis circuit-level** [**level-1** | **level-1-2** | **level-2**]

   By default, the level of an interface is Level-1-2.

   By default, Level-1 and Level-2 neighbor relationships will be established between two Level-1-2 devices. To allow the two Level-1-2 devices to establish only the Level-1 or Level-2 neighbor relationship, change the level of interfaces.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic IS-IS Configurations (3)

7. Set the network type of the interface to P2P.

   [Huawei-GigabitEthernet0/0/1]**isis circuit-type p2p**

   By default, the network type of an interface is determined by the physical type of the interface.

   After a broadcast interface is simulated as a P2P interface, the interval for sending IIHs, the number of IIHs that IS-IS does not receive from a neighbor before the neighbor is declared Down, interval for resending LSPs on a P2P link, and various IS-IS authentication modes are restored to the default settings; other configurations such as the DIS priority, DIS name, and interval for sending CSNPs on a broadcast network become invalid.

8. Restore the default network type of the IS-IS interface.

   [Huawei-GigabitEthernet0/0/1] **undo isis circuit-type**

   After the **undo isis circuit-type** command is run to restore the default network type of an IS-IS interface, the interval for sending IIHs, number of IIHs that IS-IS does not receive from a neighbor before the neighbor is declared Down, interval for retransmitting LSPs on a P2P link, various IS-IS authentication modes, DIS priority, and interval for sending CSNPs on a broadcast network are restored to the default settings.

9. Change the DIS priority of an IS-IS interface.

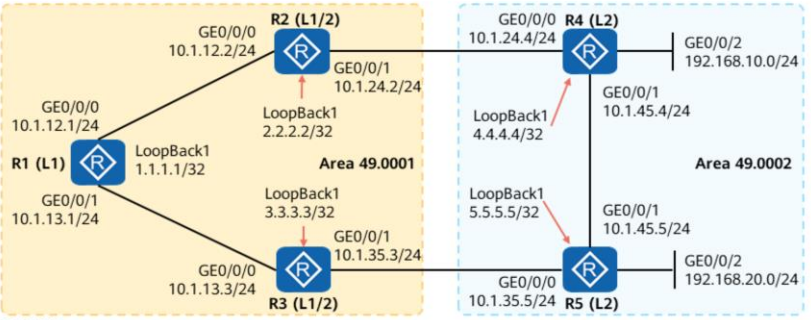
   [Huawei-GigabitEthernet0/0/1] **isis dis-priority** *priority* [**level-1** | **level-2**]

   By default, the DIS priority of an IS-IS interface is 64.

   The **isis dis-priority** command sets a priority for an interface to run for the DIS on a specified level.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: IS-IS Configuration



Networking Requirements

As shown in the figure, five routers are deployed on the network. The five routers need to communicate with each other. R1 can only process a small amount of data because it has lower performance than the other routers. R1 is required to select the optimal path to access network segments 192.168.10.0/24 and 192.168.20.0/24.

Configuration Roadmap

Enable IS-IS on each router so that the routers can be interconnected. Configure R1 as a Level-1 router to enable it to maintain less data. In addition, configure R2 and R3 as Level-1-2 routers to interconnect with Level-2 routers of R4 and R5.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

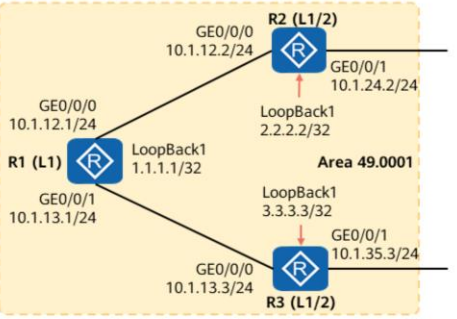
Case: Configurations of R1, R2, and R3

R1 configuration:

```
[R1] isis 1
[R1-isis-1] is-level level-1
[R1-isis-1] network-entity 49.0010.0100.1001.00
[R1-isis-1] quit
[R1] interface gigabitethernet 0/0/0
[R1-GigabitEthernet0/0/0] isis enable 1
[R1-GigabitEthernet0/0/0] interface gigabitethernet 0/0/1
[R1-GigabitEthernet0/0/1] isis enable 1
```

R2 configuration:

```
[R2] isis 1
[R2-isis-1] is-level level-1-2
[R2-isis-1] network-entity 49.0020.0200.2002.00
[R2-isis-1] quit
[R2] interface gigabitethernet 0/0/0
[R2-GigabitEthernet0/0/0] isis enable 1
[R2-GigabitEthernet0/0/0] interface gigabitethernet 0/0/1
[R2-GigabitEthernet0/0/1] isis enable 1
```



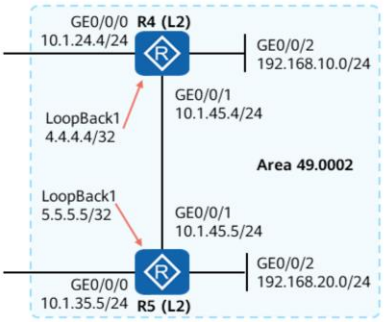
The level of R1 needs to be set to Level-1 and the levels of R2 and R3 to Level-1-2. The configuration of R3 is similar to that of R2 and is not provided here.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Configurations of R4 and R5

R4 configuration:

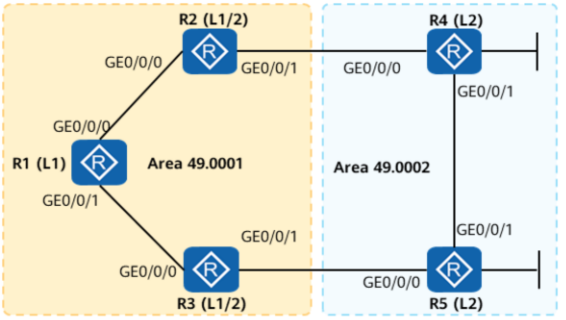
```
[R4] isis 1
[R4-isis-1] is-level level-2
[R4-isis-1] network-entity 49.0040.0400.4004.00
[R4-isis-1] quit
[R4] interface gigabitethernet 0/0/0
[R4-GigabitEthernet0/0/0] isis enable 1
[R4-GigabitEthernet0/0/0] interface gigabitethernet 0/0/1
[R4-GigabitEthernet0/0/1] isis enable 1
[R4-GigabitEthernet0/0/1] interface gigabitethernet 0/0/2
[R4-GigabitEthernet0/0/2] isis enable 1
```



The level of R4 and R5 needs to be set to Level-2. The configuration of R5 is similar to that of R4 and is not provided here.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Configuration Verification (Checking the IS-IS Neighbor Relationship Table)



<R4> display isis peer

Peer information for ISIS(1)

System Id	Interface	Circuit Id	State	HoldTime	Type	PRI
0020.0200.2002	GE0/0/0	0040.0400.4004.01	Up	25s	L2	64
0050.0500.5005	GE0/0/1	0050.0500.5005.01	Up	13s	L2	64

Total Peer(s): 2

R4 sets up a Level-2 neighbor relationship.

<R1> display isis peer

Peer information for ISIS(1)

System Id	Interface	Circuit Id	State	HoldTime	Type	PRI
0020.0200.2002	GE0/0/0	0010.0100.1001.01	Up	29s	L1	64
0030.0300.3003	GE0/0/1	0030.0300.3003.01	Up	9s	L1	64

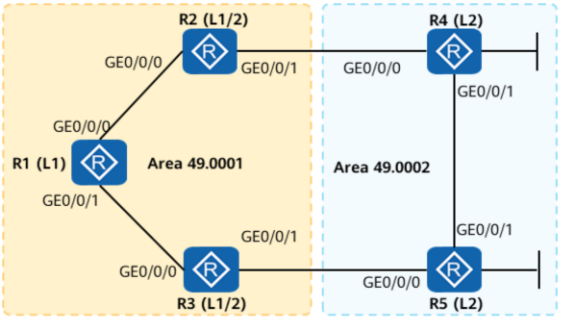
Total Peer(s): 2

R1 sets up a Level-1 adjacency.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- System Id: indicates the system ID of the neighbor.
- Interface: describes the router interface through which the neighbor relationship is established.
- Type: indicates the type of the neighbor relationship.
- PRI: indicates the DIS priority of the interface.

Case: Configuration Verification (Checking the IS-IS Routing Table)



<R1> display isis route

Route information for ISIS(1)

ISIS(1) Level-1 Forwarding Table

IPV4 Destination	IntCost	ExtCost	ExitInterface	NextHop
0.0.0.0/0	10	NULL	GE0/0/0	10.1.12.2
			GE0/0/1	10.1.13.3
10.1.24.0/24	20	NULL	GE0/0/0	10.1.12.2
10.1.13.0/24	10	NULL	GE0/0/1	Direct
10.1.12.0/24	10	NULL	GE0/0/0	Direct
10.1.35.0/24	20	NULL	GE0/0/1	10.1.13.3

The routing table of R1 contains specific routes in the area and two equal-cost default routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Configuration Verification (Checking the IS-IS Routing Table)

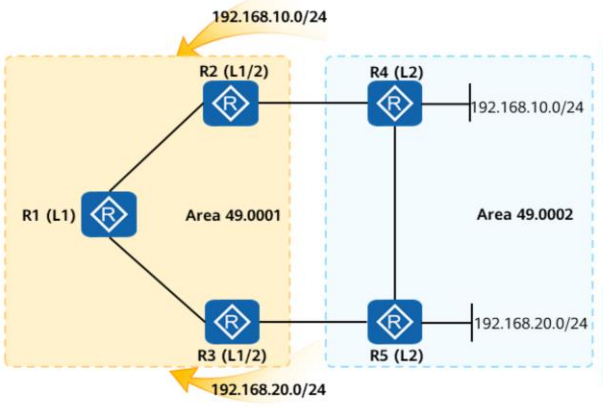
```
<R3>display isis route
                Route information for ISIS(1)
                ------------------------------
                ISIS(1) Level-1 Forwarding Table
                ------------------------------
IPV4 Destination   IntCost  ExtCost  ExitInterface   NextHop       Flags
--------------------------------------------------------------------------------
0.0.0.0/0          20       NULL
10.1.24.0/24       30       NULL     GE0/0/0         10.1.13.1     A/-/L/-
10.1.13.0/24       10       NULL     GE0/0/0         Direct        D/-/L/-
10.1.12.0/24       20       NULL     GE0/0/0         10.1.13.1     A/-/L/-
10.1.35.0/24       10       NULL     GE0/0/1         Direct        D/-/L/-
Flags: D-Direct, A-Added to URT, L-Advertised in LSPs, S-IGP Shortcut,
U-Up/Down Bit Set
......
```

```
......
                ISIS(1) Level-2 Forwarding Table
                ------------------------------
IPV4 Destination   IntCost  ExtCost  ExitInterface   NextHop       Flags
--------------------------------------------------------------------------------
10.1.24.0/24       30       NULL
10.1.13.0/24       10       NULL     GE0/0/0         Direct        D/-/L/-
10.1.12.0/24       40       NULL
192.168.20.0/24    20       NULL     GE0/0/1         10.1.35.5     A/-/-/-
10.1.45.0/24       20       NULL     GE0/0/1         10.1.35.5     A/-/-/-
192.168.10.0/24    30       NULL     GE0/0/1         10.1.35.5     A/-/-/-
10.1.35.0/24       10       NULL     GE0/0/1         Direct        D/-/L/-
Flags: D-Direct, A-Added to URT, L-Advertised in LSPs, S-IGP Shortcut,
U-Up/Down Bit Set
```

R3 functions as a Level-1-2 router and has Level-1 and Level-2 routes in the routing table.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Leaking Configuration



R2 configuration:

```
[R2] ip ip-prefix 1 permit 192.168.10.0 24
[R2] isis 1
[R2-isis-1] import-route isis level-2 into level-1 filter-policy ip-prefix 1
```

R3 configuration:

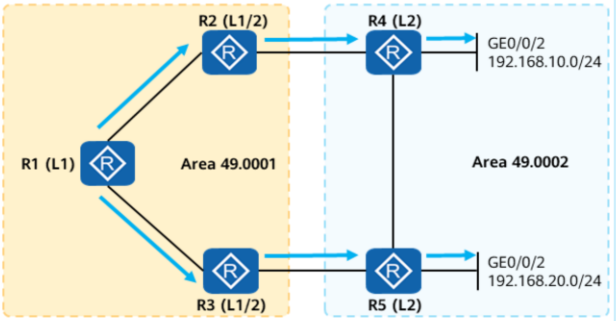
```
[R3] ip ip-prefix 1 permit 192.168.20.0 24
[R3] isis 1
[R3-isis-1] import-route isis level-2 into level-1 filter-policy ip-prefix 1
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The filtering policy will be described in subsequent courses.

Route Leaking Verification

Check the routing table of R1. You can see that two specific routes to 192.168.10.0/24 and 192.168.20.0/24 are added. The costs of the two routes are 30. When a data packet reaches the two network segments through R1, no sub-optimal path is generated.




```
<R1> display ip routing-table protocol isis
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Public routing table : ISIS
         Destinations : 5        Routes : 6

ISIS routing table status : <Active>
         Destinations : 5        Routes : 6

Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

0.0.0.0/0           ISIS-L1 15   10         D   10.1.12.2       GigabitEthernet0/0/0
                    ISIS-L1 15   10         D   10.1.13.3       GigabitEthernet0/0/1
192.168.10.0/2      ISIS-L1 15   30         D   10.1.12.2       GigabitEthernet0/0/0
192.168.20.0/2      ISIS-L1 15   30         D   10.1.13.3       GigabitEthernet0/0/1
10.1.24.0/24        ISIS-L1 15   20         D   10.1.12.2       GigabitEthernet0/0/0
10.1.35.0/24        ISIS-L1 15   20         D   10.1.13.3       GigabitEthernet0/0/1
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IS-IS Authentication Types

- To ensure network security, IS-IS authentication encrypts IS-IS packets by adding an authentication field to packets. When a local router receives IS-IS packets from a remote router, the local router discards the packets if the authentication passwords do not match. This protects the local router.
- Based on the types of packets, authentication is classified as follows:
  - Interface authentication: is configured in the interface view to authenticate Level-1 and Level-2 IS-to-IS Hello PDUs (IIHs).
  - Area authentication: is configured in the IS-IS process view to authenticate Level-1 CSNPs, PSNPs, and LSPs.
  - Routing domain authentication: is configured in the IS-IS process view to authenticate Level-2 CSNPS, PSNPs, and LSPs.
- Based on the authentication modes of packets, authentication is classified into the following types:
  - Cleartext authentication: is a simple authentication mode in which passwords are directly added to packets. This mode provides the lowest password security among all the authentication modes listed here.
  - MD5 authentication: uses the MD5 algorithm to encrypt a password before adding the password to the packet, which improves password security.
  - Keychain authentication: further improves network security with configurable key chain that changes with time.
  - HMAC-SHA256 authentication: uses the HMAC-SHA256 algorithm to encrypt a password before adding the password to the packet, which improves password security.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IS-IS Authentication

- Interface authentication
  - Authentication passwords for IIHs are saved on interfaces. The interfaces send authentication packets with the authentication TLV. Interconnected router interfaces must be configured with the same password.
- Area authentication
  - Every router in an IS-IS area must use the same authentication mode and have the same key chain.
- Routing domain authentication
  - Every Level-2 or Level-1-2 router in an IS-IS area must use the same authentication mode and have the same key chain.
  - In area authentication and routing domain authentication, you can configure a router to authenticate LSPs and SNPs separately in the following ways:
    - A router sends LSPs and SNPs carrying the authentication TLV and verifies the authentication information about the received LSPs and SNPs.
    - A router sends LSPs that carry the authentication TLV and verifies the authentication information of the LSPs it receives. The router sends SNPs that carry the authentication TLV and does not verify the authentication information of the SNPs it receives.
    - A router sends LSPs that carry the authentication TLV and verifies the authentication information of the LSPs it receives. The router sends SNPs without the authentication TLV and does not verify the authentication information of the SNPs it receives.
    - A router sends LSPs and SNPs that carry the authentication TLV but does not verify the authentication information of the LSPs and SNPs it receives.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IS-IS Authentication Configuration (1)

1. Configure IS-IS area authentication.

   [Huawei-isis-1] **area-authentication-mode** { { **simple** | **md5** } { **plain** *plain-text* | [**cipher**] *plain-cipher-text* } **keychain** *keychain-name* | **hmac-sha256 key-id** *key-id* } [**snp-packet** { **authentication-avoid** | **send-only** } | **all-send-only**]

   simple indicates that the password is authenticated in plain-text mode. **keychain** specifies the key chain that changes with time. **md5** indicates that the password is encrypted using the HMAC-MD5 algorithm and then used for authentication. **snp-packet** indicates SNP authentication. **authentication-avoid** indicates that the router encapsulates generated LSPs but not SNPs with authentication information and authenticates received LSPs but not SNPs. **send-only** indicates that the router encapsulates generated LSPs and SNPs with authentication information, and authenticates received LSPs but not SNPs. **all-send-only** indicates that the router encapsulates generated LSPs and SNPs with authentication information, but does not authenticate received LSPs and SNPs.

2. Configure IS-IS routing domain authentication.

   [Huawei-isis-1] **domain-authentication-mode** { { **simple** | **md5** } { **plain** *plain-text* | [**cipher**] *plain-cipher-text* } **keychain** *keychain-name* | **hmac-sha256 key-id** *key-id* } [**snp-packet** { **authentication-avoid** | **send-only** } | **all-send-only**]

   The parameter description is the same as that of area authentication.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

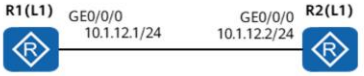
IS-IS Authentication Configuration (2)

3. Configure IS-IS interface authentication.

[Huawei-GigabitEthernet0/0/0] **isis authentication-mode** [**keychain** | **md5** | **simple**] [**level-1** | **level-2**] [**ip** | **osi**] [**send-only**]

level-1 indicates level-1 authentication. **level-2** indicates level-2 authentication. **send-only** indicates that the router only encapsulates authentication information to the sent IIHs, but does not authenticate the received IIHs.

Interface authentication is used as an example.



Configure interface authentication on GE0/0/0 of R1.

```
[R1] interface GigabitEthernet0/0/0
[R1-GigabitEthernet0/0/0] isis authentication-mode simple cipher huawei
```

Interface authentication is not configured on R2, so an authentication error is displayed when IS-IS error information is displayed.

```
[R1] display isis error

Hello packet errors:
Mismatched Max Area Addr : 0          Bad Authentication : 6
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Multiple) Which of the following statements about OSPF and IS-IS are true?( )
   A. Both OSPF and IS-IS are link-state routing protocols.
   B. OSPF supports only IP networks, whereas Integrated IS-IS supports CLNP and IP networks.
   C. OSPF uses area 0 to identify the backbone area.
   D. The IS-IS backbone area consists of Level-2 and Level-1-2 routers.
2. (Multiple) Which of the following statements about Level-1 routers are true?( )
   A. A Level-1 router can establish neighbor relationships only with IS-IS routers in the same area.
   B. A Level-1 router can establish a neighbor relationship with a Level-2 router in the same area.
   C. By default, a Level-1 router can reach the destination outside an area only through the default route.
   D. The LSDB of a Level-1 router contains only Level-1 LSPs.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. ABCD
2. ACD

Summary

- This course describes basic concepts of IS-IS. Each IS-IS router has at least one NET. IS-IS uses a hierarchical architecture. All Level-2 and Level-1-2 routers form a backbone area, and Level-1 routers in the same area form a common area. IS-IS packets are classified into four types: IIH, CSNP, PSNP, and LSP.
- A Level-1 router maintains only the LSDB of the area. A Level-1 route reaches the destination outside the area through the closest Level-1-2 router. To prevent sub-optimal paths, you can configure route leaking on the Level-1-2 router.
- IS-IS supports interface authentication, area authentication, and routing domain authentication.
- Both IS-IS and OSPF are link-state routing protocols and need to maintain LSDBs. As the network scale expands, a large LSDB degrades the overall performance of routers. Therefore, for the two routing protocols, multi-area network design is required for larger-scale networking.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



BGP Basics

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- A network is divided into different autonomous systems (ASs) to facilitate network management. An Exterior Gateway Protocol (EGP) was used to dynamically exchange routing information between ASs. However, the EGP advertises only reachable routes and does not select optimal routes or prevent routing loops. Therefore, the EGP cannot meet network management requirements.
- BGP was designed to replace the EGP and has the following capabilities: BGP selects optimal routes, prevents routing loops, transmits routing information efficiently, and maintains a lot of routes.
- This course describes BGP basics.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe basic concepts of BGP.
  - Describes BGP peer types.
  - Learn how to set up a BGP peer relationship.
  - Learn BGP state machine.
  - Perform basic BGP configurations.

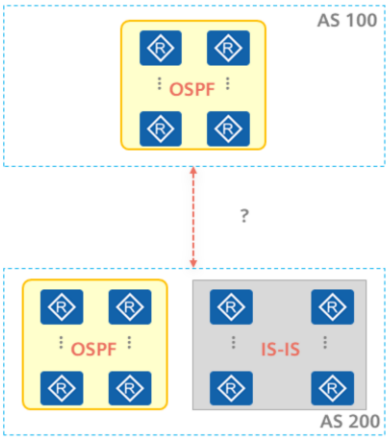
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Introduction to BGP**
2. Basic Concepts of BGP
3. Basic BGP Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

AS

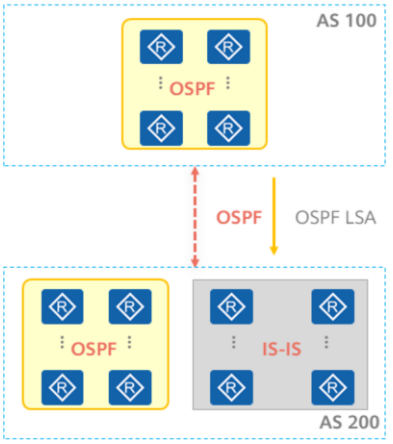


- IGPs, such as OSPF and IS-IS, are widely used on organizational networks. As the network scale expands and the number of routes on the network increases, IGPs cannot manage large-scale networks. As a result, the concept of Autonomous Systems (ASs) emerges.
- An AS is a collection of devices that are managed by the same organization and use the same route selection policy.
- ASs are distinguished by AS numbers. An AS number can be expressed in 16-bit or 32-bit format. The Internet Assigned Numbers Authority (IANA) assigns AS numbers.
- When different ASs need to communicate with each other, which routing protocol should be used to transmit routes between the ASs?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- IANA: an organization under the Internet Architecture Board (IAB). The IANA authorizes the Network Information Center (NIC) and other organizations to assign IP addresses and domain names. In addition, the IANA maintains the protocol identifier database used by the TCP/IP protocol suite, including AS numbers.
- In 16-bit format, AS numbers 64512-65534 are private ones. In 32-bit format, AS numbers 4200000000-4294967294 are private ones.

Using an IGP to Transmit Routes

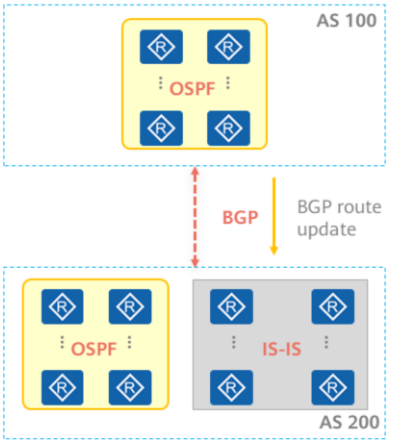


- A direct or logical link (for example, a GRE tunnel) must be established between ASs for peer relationship setup.
- ASs may belong to different organizations or companies and cannot be trusted mutually. Using an IGP may expose network information inside an AS.
- As the network scale expands, the number of routes increases, the routing table size increases, route convergence becomes slow, and device performance consumption increases.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Virtual private network (VPN): is used to build a logically and directly connected network.

Using BGP to Transmit Routes

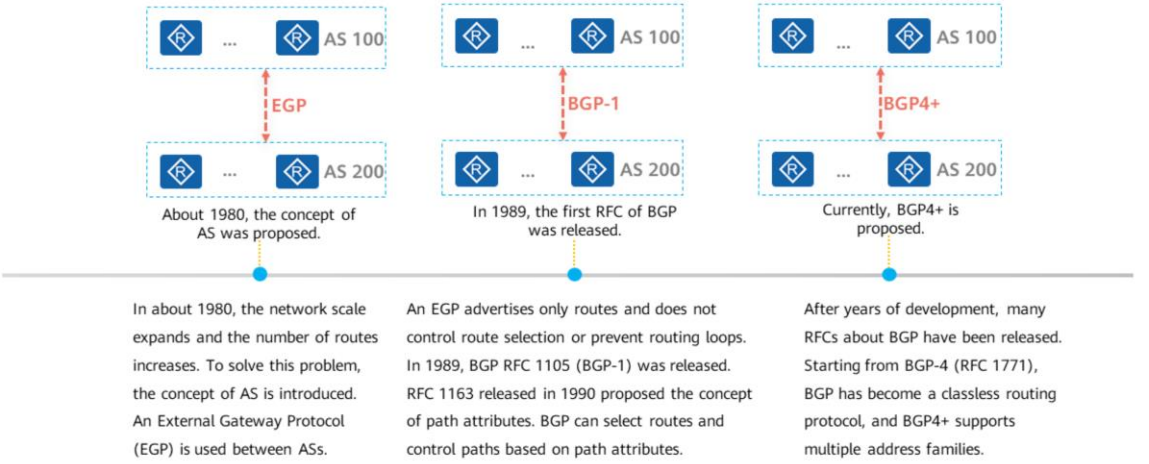


The Border Gateway Protocol (BGP) is specially used between ASs for route transmission. Compared with a conventional IGP, BGP has the following characteristics:

- BGP is based on TCP. A BGP connection can be established as long as a TCP connection can be established.
- BGP transmits only routing information, but does not expose topology information in an AS.
- BGP routes are updated only upon network changes. They are not updated periodically.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BGP Development History



In about 1980, the network scale expands and the number of routes increases. To solve this problem, the concept of AS is introduced. An External Gateway Protocol (EGP) is used between ASs.

An EGP advertises only routes and does not control route selection or prevent routing loops. In 1989, BGP RFC 1105 (BGP-1) was released. RFC 1163 released in 1990 proposed the concept of path attributes. BGP can select routes and control paths based on path attributes.

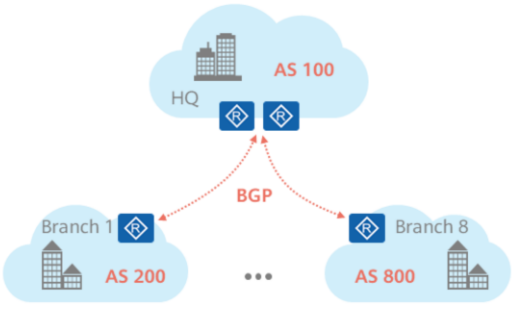
After years of development, many RFCs about BGP have been released. Starting from BGP-4 (RFC 1771), BGP has become a classless routing protocol, and BGP4+ supports multiple address families.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The latest RFC for BGP-4 is RFC 4271. Compared with RFC 1771, RFC 4271 further describes some details, such as events, state machine, and BGP route decision-making process.

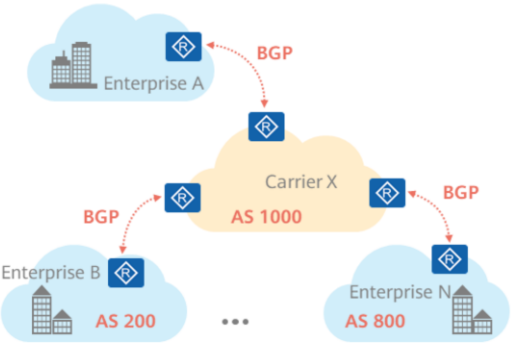
BGP Application in Enterprises

Communication within an Enterprise



Branches of a large enterprise use BGP to exchange routes. Different branches belong to different BGP ASs.

Communication Between Enterprises and a Carrier



The enterprise and carrier can use BGP to exchange routes so that the enterprise network can obtain specific routes to the carrier network and the carrier can obtain routes to the enterprise network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Introduction to BGP
2. **Basic Concepts of BGP**
3. Basic BGP Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

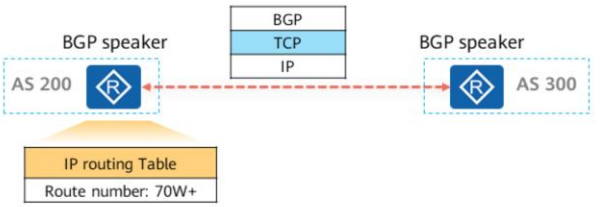


Overview of BGP

- BGP is a path vector protocol that allows devices between ASs to communicate and selects optimal routes. BGP-1 (defined in RFC 1105), BGP-2 (defined in RFC 1163), and BGP-3 (defined in RFC 1267) are three earlier versions of BGP. BGP-4 (defined in RFC 1771) has been used since 1994. Since 2006, unicast IPv4 networks have been using BGP-4 defined in RFC 4271, and other networks (such as IPv6 networks) have been using Multiprotocol BGP (MP-BGP) defined in RFC 4760.
- BGP has the following characteristics:
  - BGP uses TCP (port 179) as the transport layer protocol and triggers route updates instead of periodic route updates.
  - BGP can carry a large amount of routing information and support large-scale networks.
  - BGP provides various routing policies to flexibly select routes and instruct peers to advertise routes based on routing policies.
  - BGP supports MPLS/VPN applications and transmits VPN routes.
  - BGP offers route summarization and route dampening functions to prevent route flapping, enhancing network stability.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

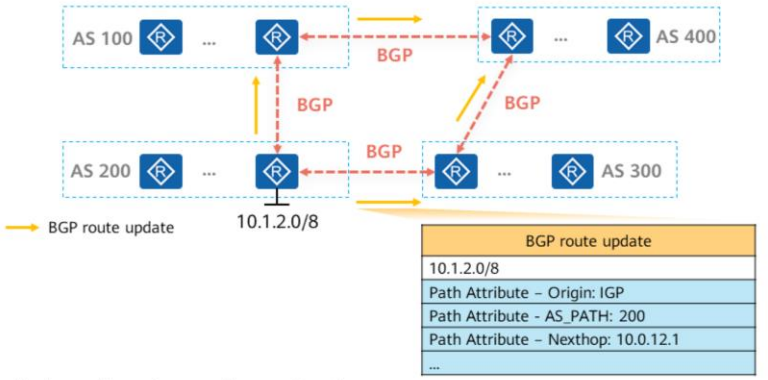
BGP Characteristics (1)



- BGP uses TCP as the transport layer protocol and TCP port number is 179. BGP sessions between routers are established based on TCP connections.
- A router that runs BGP is called a BGP speaker or a BGP router.
- Two routers that establish a BGP session are peers of each other, and BGP peers exchange BGP routing tables.
- A BGP router sends only incremental BGP route updates.
- BGP can carry a large number of route prefixes and can be applied to large-scale networks.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

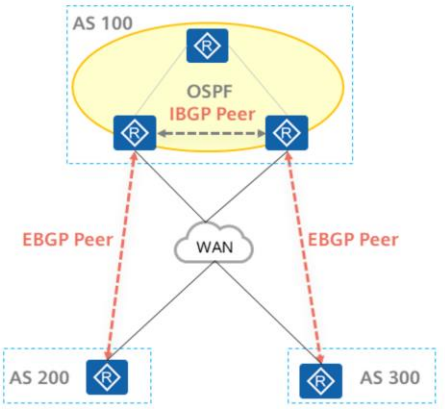
BGP Characteristics (2)



- BGP is also called a path vector routing protocol.
- Each BGP route carries multiple path attributes. Different from IS-IS and OSPF that use costs to select paths, BGP selects paths based on path attributes. Therefore, BGP is easy to operate and allows you to select the most appropriate path control mode in different scenarios.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BGP Peer Relationship

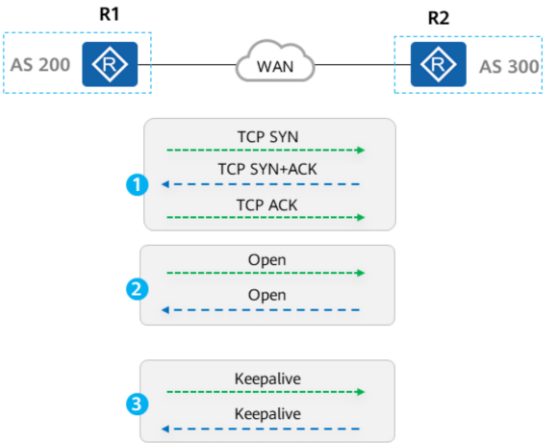


- Different from OSPF and IS-IS, BGP sessions are established based on TCP. The two routers that establish a BGP peer relationship do not need to be directly connected.
- BGP has two types of peer relationships:
  - External BGP (EBGP): BGP peer relationship between BGP routers in different ASs To establish an EBGP peer relationship between two routers, ensure that the following conditions are met:
    - The two routers belong to different ASs (AS numbers).
    - When configuring EBGP, ensure that the IP address of the peer specified in the **peer** command is reachable and the TCP connection can be set up.
  - Internal BGP (IBGP): BGP peer relationship between BGP routers in the same AS.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Establishing a BGP Peer Relationship (1)



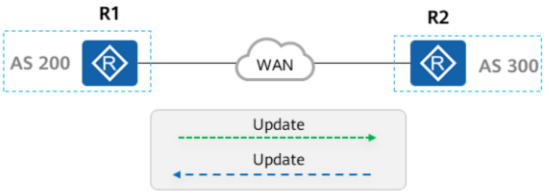
- The router that first starts BGP initiates a TCP connection. As shown in the figure on the left, R1 first starts BGP and uses a random port number to initiate a TCP connection to port 179 of R2.
- After the three-way handshake is complete, R1 and R2 send Open messages carrying parameters to each other to establish a peer relationship. After the parameters are negotiated, R1 and R2 send Keepalive messages to each other. After receiving the Keepalive messages from each other, the two routers establish a peer relationship. In addition, R1 and R2 periodically send Keepalive messages to maintain the connection.
- The Open message carries the following information:
  - My Autonomous System: indicates the AS number.
  - Hold Time: is used to negotiate the time for sending Keepalive messages.
  - BGP Identifier: indicates the router ID of the local router.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Each BGP peer initiates a TCP three-way handshake, so two TCP connections are established. Actually, BGP retains only one TCP connection. After obtaining the BGP identifier of the peer from an Open message, the BGP peer compares the local router ID with the peer router ID. If the local router ID is smaller than the remote router ID, the local router terminates the TCP connection and uses the TCP connection initiated by the remote router to exchange BGP messages.

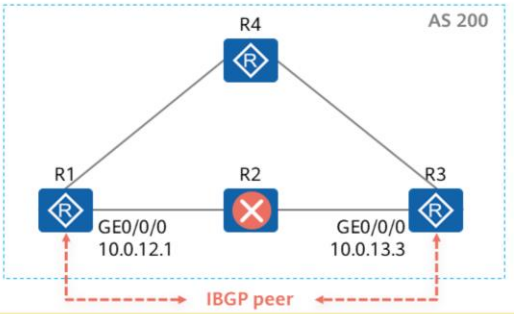
Establishing a BGP Peer Relationship (2)



After a BGP peer relationship is established, a BGP router sends a BGP Update message to advertise routes to the peer.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Source Address of a TCP Connection



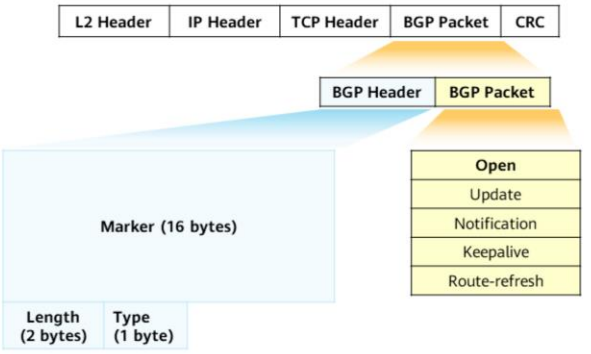
Typically, a network in an AS provides certain redundancy. If R1 and R3 use directly connected interfaces to establish an IBGP peer relationship, the BGP session will be interrupted once the directly connected interface or link fails. However, due to redundant links, the IP connectivity between R1 and R3 is not Down. (R1 and R3 can still communicate with each other through R4.)

- By default, BGP uses the interface that sends BGP messages as the local interface of a TCP connection.
- When deploying IBGP peer relationships, you are advised to use loopback addresses as source addresses of TCP connections. The loopback interface is stable. The reliability can be ensured by using the IGP and redundancy topology in the AS.
- During EBGP peer relationship setup, the IP address of the directly connected interface is often used as the source address. If you use a loopback interface to establish an EBGP peer relationship, pay attention to the EBGP multi-hop problem.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BGP Message Types (1)



There are five types of BGP messages. Different types of messages have the same header.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Different from common IGP protocols, BGP uses TCP as the transport layer protocol and port 179. This enables BGP to establish peer relationships between indirectly connected routers.



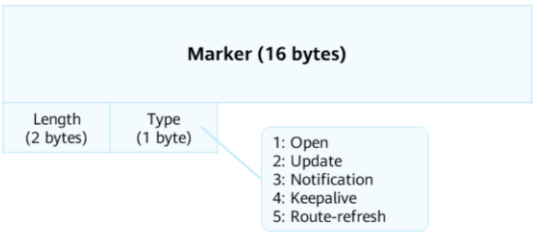
BGP Message Types (2)

Name	Function	Usage Scenario
Open	Negotiate BGP peer parameters and establish peer relationships.	After a BGP TCP connection is established, Open messages are sent.
Update	Send BGP route update.	After a BGP peer relationship is established and routes need to be sent or routes change, Update messages are sent.
Notification	Report error information and terminate the peer relationship.	When detecting an error during BGP running, the local BGP router sends a Notification message to notify the peer of the error.
Keepalive	Indicate that the peer relationship is established and the BGP peer relationship is maintained.	After receiving a Keepalive message from the peer, the BGP router sets the peer relationship status to Established and periodically sends Keepalive messages to maintain the connection.
Route-refresh	Request the peer to retransmit routes if routing policies are changed. Only the BGP devices supporting route-refresh can send and respond to Route-refresh messages.	When the routing policy changes, Route-refresh messages are sent to trigger the peer to re-advertise routes.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BGP Message Format: Header Format



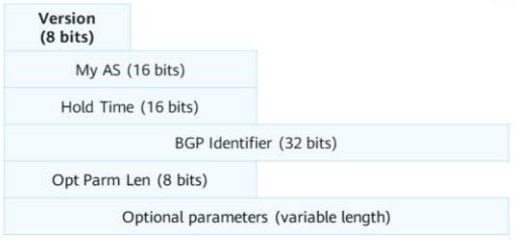
The five types of BGP messages have the same header, as shown in the left part of. The main fields are described as follows:

- Marker: is used to check whether synchronization information of BGP peers is complete and for BGP authentication. It has 16 bytes. If authentication is not used, all bits are set to 1 (all FFs in hexadecimal notation).
- Length: indicates the total length of a BGP message (including the packet header), in bytes. It has 2 bytes.
- Type: indicates the type of BGP messages. It has 1 byte. The value ranges from 1 to 5, indicating Open, Update, Notification, Keepalive, and Route-refresh messages.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BGP Message Format: Open



The Open message is the first message sent after a TCP connection is established. It is used to establish a connection between BGP peers. The figure on the left shows the Open message format. The main fields in the Open message are described as follows:

- Version: indicates the BGP version number. For BGPv4, the value is 4.
- My AS: indicates the local AS number. By comparing AS numbers of the two ends, you can determine whether the peer end and the local end are in the same AS.
- Hold Time: indicates the holding time. BGP peers need to negotiate the holding time and keep it consistent when establishing a peer relationship. If a router does not receive any Keepalive or Update message from its peer within the holding time, the BGP connection is considered as disconnected.
- BGP Identifier: indicates the router ID of a BGP router. The field is in IP address format and identifies a BGP router.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Opt Parm Len: indicates the length of Optional parameters.
- Optional parameters: declares optional capabilities of a BGP router, such as authentication and multi-protocol support.
- In addition to IPv4 unicast routing information, BGP4+ supports multiple network layer protocols, such as IPv6 and multicast. During negotiation, BGP peers negotiate the support for network layer protocols through the Optional parameters field.

BGP Message Format: Update

Unfeasible Routes length (2 bytes)
Withdrawn Routes (N bytes)
Total path attribute length (2 bytes)
Path Attributes (N bytes)
NLRI (N bytes)

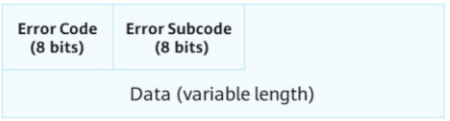
- Update messages are used to transmit routing information between peers. Update messages can be used to advertise and withdraw routes.
- An Update message can advertise a type of routes with the same path attributes, which are placed in Network Layer Reachability Information (NLRI) fields. In addition, the Update message can carry multiple unreachable routes, which are stored in the Withdrawn Routes field.
- The figure on the left shows the message format. The main fields in the message are described as follows:
  - Withdrawn routes: indicates the list of unreachable routes.
  - Path attributes: indicates the list of all path attributes related to NLRI. Each path attribute consists of a Type-Length-Value (TLV).
  - NLRI: indicates the prefix and prefix length of a reachable route.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Unfeasible routes length: indicates the length of the Unfeasible Routes field, in bytes. If the value is 0, the Withdrawn Routes field does not exist.
- Withdrawn Routes Length: indicates the length of the Withdrawn Routes field. If the value is 0, the Withdrawn Routes field is omitted.
- Total path attribute length: indicates the length of the Path Attributes field, in bytes. If the value is 0, the Path Attributes field does not exist.

BGP Message Format: Notification



When BGP detects an error (may occur during or after the establishment of a peer relationship), BGP sends a Notification message to the peer to notify the peer of the error cause. Then the BGP connection is interrupted immediately.

- Error Code and Error Subcode: are used to notify the peer of the specific error type.
- Data: is used to describe the detailed error content. The length is variable.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BGP Message Format: Keepalive



- After receiving a Keepalive message from the peer, the BGP router sets the peer relationship status to Established and periodically sends Keepalive messages to maintain the connection.
- The Keepalive message only contains the packet header without any other fields.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BGP Message Format: Route-refresh

AFI (16 bits)	Res (8 bits)	SAFI (8 bits)

- Route-refresh messages are used to request the peer to resend routing information of a specified address family. In most cases, after the local end modifies a routing policy, the peer sends Update messages again. Then the local end recalculates BGP routes according to the new routing policy.
- The main fields are described as follows:
  - Address Family Identifier (AFI): identifies an address family, for example, IPv4.
  - Res.: is reserved. The 8 bits must be set to 0.
  - Subsequent Address Family Identifier (SAFI): indicates the sub-address family identifier.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

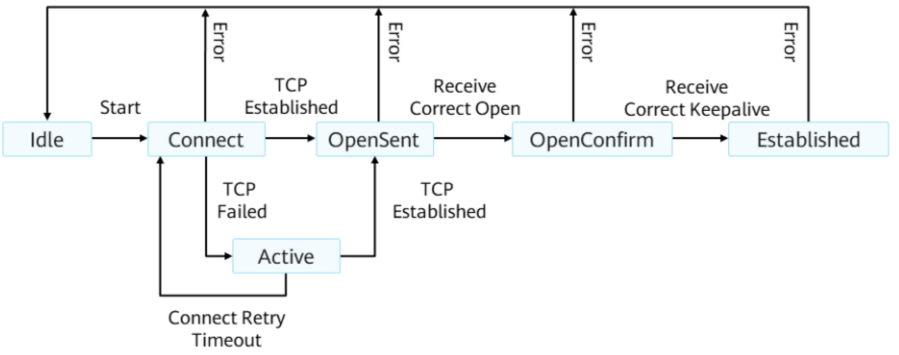
- During Open message negotiation, the two BGP routers negotiate whether they support route-refresh. If they support route-refresh, you can run the **refresh bgp** command to softly reset the BGP connection to refresh a BGP routing table without tearing down any BGP connection
- If a device's peer does not support route-refresh, you can run the **peer keep-all-routes** command to configure the device to retain all routing updates received from the peer so that the device can refresh its routing table without tearing down the BGP connection with the peer.
- By default, the device is not configured to retain all routing updates received from the peer.

BGP State Machine (1)

Peer Status	Usage
Idle	Prepare TCP connections and monitor remote peers. When enabling BGP, prepare sufficient resources.
Connect	A TCP connection is being set up. Authentication is completed during the TCP connection setup. If the TCP connection fails to be established, the local end enters the Active state and attempts to establish a TCP connection repeatedly.
Active	The TCP connection fails to be set up and the local end attempts to establish a TCP connection repeatedly.
OpenSent	After the TCP connection is established, the local end sends an Open packet carrying parameters to negotiate the establishment of the peer.
OpenConfirm	After the parameter and capability negotiation succeeds, the local end sends a Keepalive message and waits for the Keepalive message from the peer end.
Established	The local end has received the Keepalive message from the peer. The capabilities of the two ends are the same after negotiation. The local end starts to use the Update message to advertise routing information.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BGP State Machine (2)

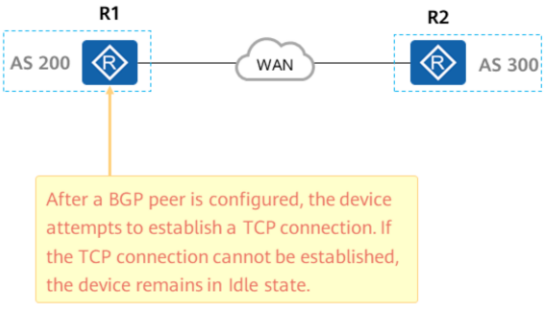


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Initially, BGP is in Idle state. In Idle state, a BGP device refuses BGP connection requests from the peer. The BGP device initiates a TCP connection with its BGP peer and changes its state to Connect only after receiving a Start event from the system.
  - The Start event occurs when an operator configures a BGP process or resets an existing BGP process or when the router software resets a BGP process.
  - If an error occurs in any state, for example, BGP receives a Notification message or a TCP disconnection notification, BGP returns to the Idle state.
- In the Connect state, BGP starts the Connect Retry timer and waits for a TCP connection to be established:
  - If the TCP connection is established, BGP sends an Open message to the peer and transitions to the OpenSent state.
  - If the TCP connection fails to be established, BGP transitions to the Active state.
  - If the BGP device does not receive a response from the peer before the Connect Retry timer expires, the BGP device attempts to establish a TCP connection with another peer and stays in Connect state.
- In Active state, the BGP device keeps trying to establish a TCP connection with the peer.
  - If the TCP connection is established, BGP sends an Open message to the peer, terminates the Connect Retry timer, and transitions to the OpenSent state.
  - If the TCP connection fails to be established, BGP stays in Active state.
  - If the BGP device does not receive a response from the peer before the Connect Retry timer expires, the BGP device returns to the Connect state.

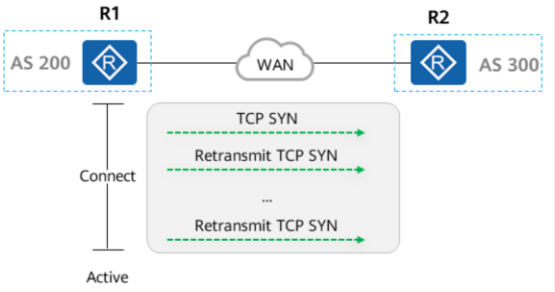
Illustration of BGP State Machine (1)

Idle state



There is no route to the BGP peer (common cause), so the local end remains in Idle state.

Connect, Active



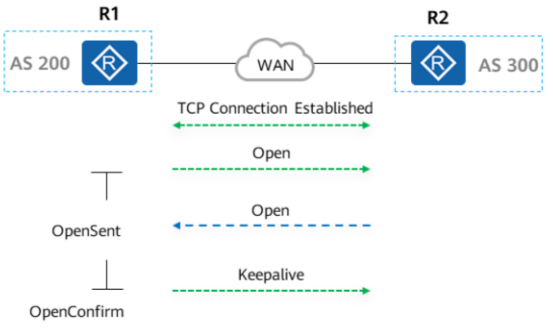
After a BGP peer is configured and the route to the peer address is found, a TCP three-way handshake is initiated. During the three-way handshake, the BGP device is in Connect state. If the TCP connection cannot be established for a long time, the BGP device enters the Active state.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

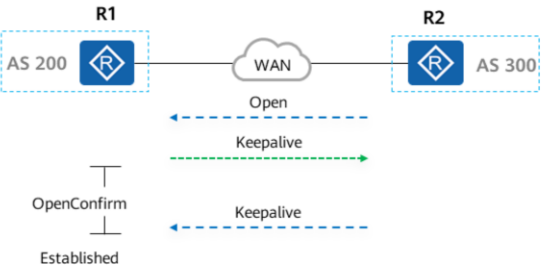
Illustration of BGP State Machine (2)

OpenSent, OpenConfirm



After the TCP three-way handshake is complete, the local end sends an Open message to establish a peer relationship and enters the OpenSent state. After receiving an Open message from the peer end and verifying that the parameters are correct, the local end sends a Keepalive message. Then the local end enters the OpenConfirm state.

Established



After entering the OpenConfirm state and receiving a Keepalive message from its peer, the BGP router enters the Established state. The peer relationship is then established.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BGP Peer Table

```
<R1>display bgp peer
 BGP local router ID : 10.0.1.1
 Local AS number : 100
 Total number of peers : 1                 Peers in established state : 1

  Peer            V     AS  MsgRcvd  MsgSent  OutQ    Up/Down       State       PrefRcv

 10.0.12.2        4    100    25719    25714     0     0428h32m     Established       1
```

You can run the **display bgp peer** command to check the BGP peer table. The parameters in the command output are described as follows:

- Peer: indicates the IP address of the peer.
- V: indicates the BGP version used on the peer.
- AS: indicates the AS number.
- Up/Down: indicates the period of time during which a BGP session keeps the current state.
- State: indicates the status of the peer.
- PrefRcv: indicates the number of route prefixes sent from the peer.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The BGP peer table lists the BGP peer of the local device and the status of the peer.
- MsgRcvd: indicates the number of received messages.
- MsgSent: indicates the number of sent messages.
- OutQ: indicates the message to be sent to the specified peer. The value is always 0.

BGP Routing Table (1)

```
<R1>display bgp routing-table
 BGP Local router ID is 10.0.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
 Total Number of Routes: 2
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn
*>i   10.0.45.0/24       10.0.4.4       0          100       0       ?
* i                      10.0.4.4       0          100       0       ?
```

- Run the **display bgp routing-table** command on the device to check the BGP routing table.
  - Network: indicates the destination network address and subnet mask of the route.
  - NextHop: indicates the IP address of the next hop.
- To check detailed information about a route, run the **display bgp routing-table** *ipv4-address { mask | mask-length}* command. This command displays detailed information about matched BGP routes.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The BGP routing table lists all the BGP routes discovered by the local device. If multiple routes to the same destination exist, all the routes are listed, but only one route is preferred for each destination.

BGP Routing Table (2)

```
<R1>display bgp routing-table 10.0.45.0 24
 BGP local router ID : 10.0.1.1
 Local AS number : 100
 Paths:   2 available, 1 best, 1 select
 BGP routing table entry information of 10.0.45.0/24:
 From: 10.0.2.2 (10.0.2.2)                    #Specify the route source.
 Route Duration: 06h19m44s
 Relay IP Nexthop: 10.0.12.2
 Relay IP Out-Interface: GigabitEthernet0/0/0
 Original nexthop: 10.0.4.4                   #Specify the next-hop IP
address of the route.
 Qos information : 0x0
AS-path Nil, origin incomplete,  MED 0, localpref 100, pref-val 0, valid,
internal, best, select, active, pre 255, IGP cost 2 # Indicate the path
attribute and whether the path is preferred.
 Originator:  10.0.4.4
 Cluster list: 10.0.2.2
 Not advertised to any peer yet
```

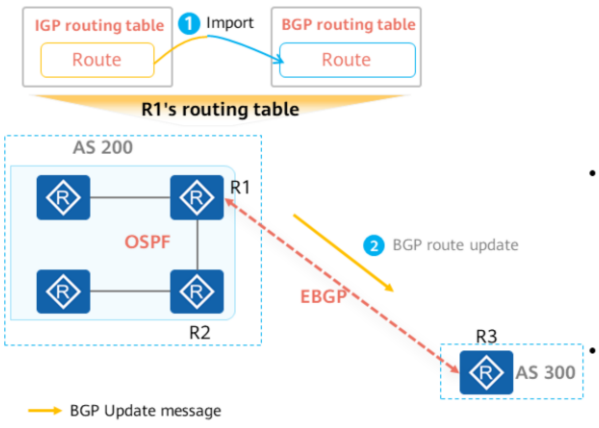
```
BGP routing table entry information of 10.0.45.0/24:
From: 10.0.3.3 (10.0.3.3)
Route Duration: 05h17m56s
Relay IP Nexthop: 10.0.12.2
Relay IP Out-Interface: GigabitEthernet0/0/0
Original nexthop: 10.0.4.4
Qos information : 0x0
AS-path Nil, origin incomplete,  MED 0, localpref 100, pref-val 0, valid,
internal, pre 255, IGP cost 2, not preferred for peer address
Originator:  10.0.4.4
Cluster list: 10.0.3.3
Not advertised to any peer yet
```

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The **display bgp routing-table ipv4-address** { *mask* | *mask-length* } command displays information about a BGP route with a specified IP address/mask length. The information includes the route originator, next-hop address, and route path attributes.

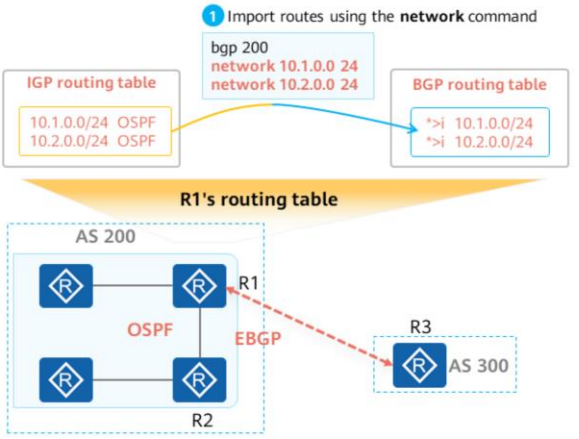
BGP Route Generation



- Different from an IGP, BGP does not discover or calculate routes. Instead, BGP injects routes from the IGP routing table to the BGP routing table and sends Update messages carrying the routes to BGP peers.
- BGP can import routes in either of the following modes:
  - network
  - import-route
- Similar to an IGP, BGP can summarize routes based on existing routes to generate a summarized route.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Importing Routes Using the network Command (1)



Routes are imported by using the **network** command:

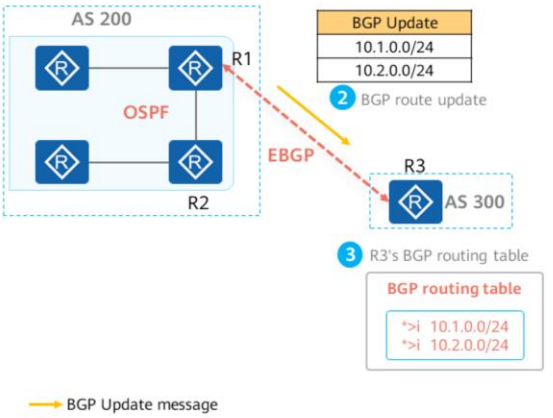
1. The BGP router in AS 200 has learned two routes to 10.1.0.0/24 and 10.2.0.0/24 through OSPF. The two routes are imported to the BGP process through the **network** command and added to the local BGP routing table.
2. The BGP router in AS 200 sends Update messages to advertise routes to the BGP router in AS 300.
3. After receiving the routes, the BGP router in AS 300 adds the two routes to the local BGP routing table.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The routes imported using the **network** command must exist in the IP routing table. Otherwise, the routes cannot be imported to the BGP routing table.

Importing Routes Using the network Command (2)



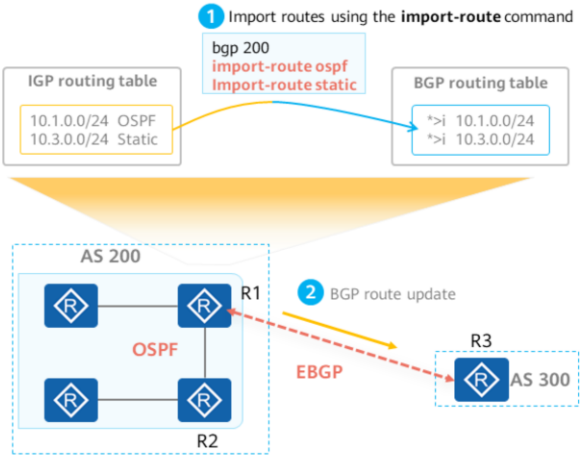
Routes are imported by using the **network** command:

1. The BGP router in AS 200 has learned two routes to 10.1.0.0/24 and 10.2.0.0/24 through OSPF. The two routes are imported to the BGP process through the **network** command and added to the local BGP routing table.
2. The BGP router in AS 200 sends Update messages to advertise routes to the BGP router in AS 300.
3. After receiving the routes, the BGP router in AS 300 adds the two routes to the local BGP routing table.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Importing Routes Using the import-route Command

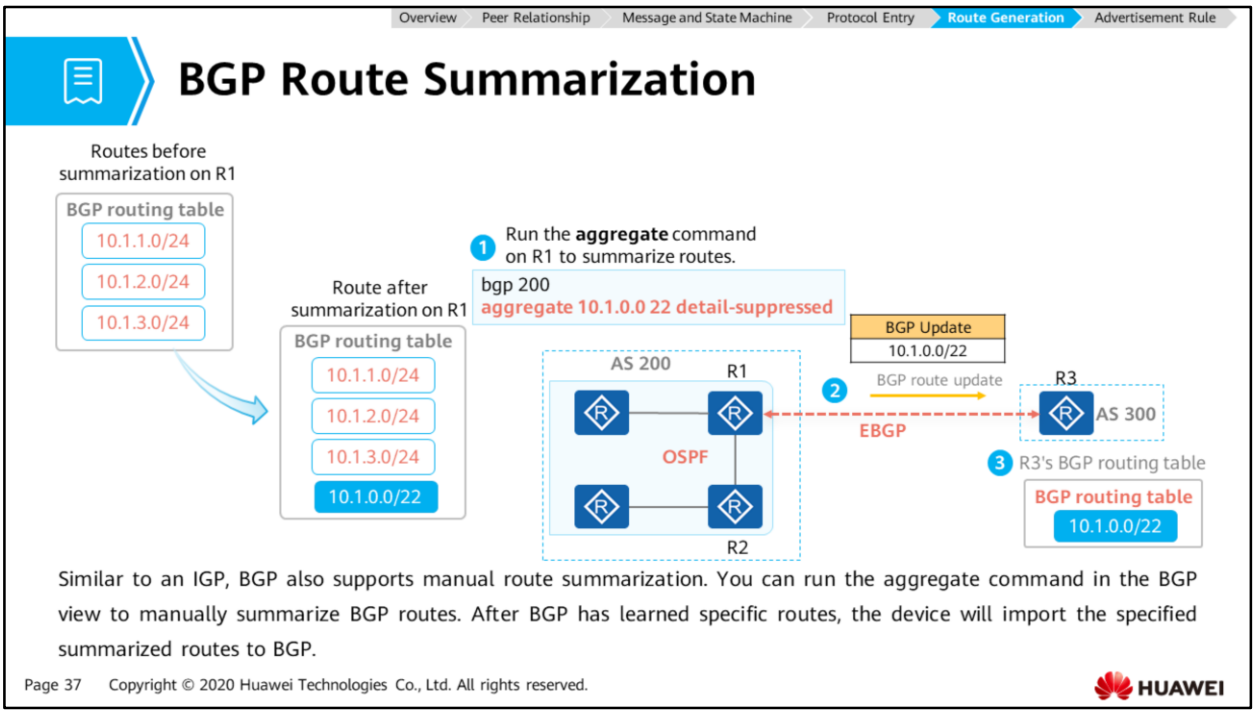


Although the routes imported by using the **network** command are accurate, only one route can be imported into the IP routing table one by one. If a large number of routes are imported, the configuration commands are complex. In this case, you can use the **import-route** command to import the following routes to the BGP routing table:

1. Direct routes
2. Static routes
3. OSPF routes
4. IS-IS routes

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI



- After route summarization is performed, the local BGP routing table contains an additional summarized route in addition to the original specific routes.
- If **detail-suppressed** is specified during route summarization, BGP advertises only the summarized route to the peer, but not the specific routes before route summarization.
- If **detail-suppressed** is configured during route summarization, only BGP route to 10.1.0.0/22 is displayed in R3's routing table, and the specific routes before summarization are not displayed.



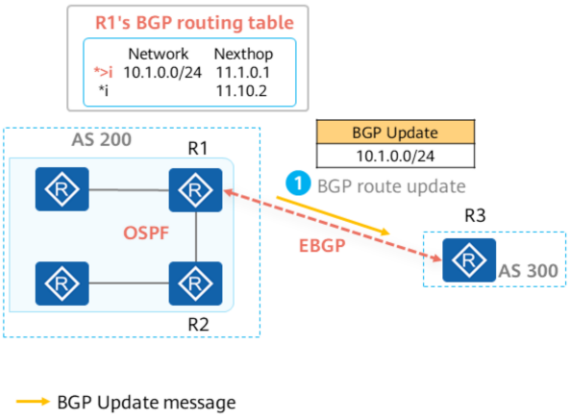
Advertisement Rule

- After BGP generates BGP routes using the network, import-route, or aggregate command, BGP sends Update messages carrying the BGP routes to the peer.
- BGP routes are advertised according to the following rules:
  1. Only the optimal and valid routes are advertised.
  2. The routes obtained from EBGP peers are advertised to all BGP peers.
  3. IBGP split horizon: The routes obtained from IBGP peers are not advertised to IBGP peers.
  4. When a router learns a BGP route (IBGP route) from its IBGP peer, the router cannot use this route or advertise this route to its EBGP peer unless the router learns this route from an IGP. In this situation, IBGP routes and IGP routes need to be synchronized. Route synchronization is used to prevent BGP routing blackholes.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BGP Route Advertisement Rule 1



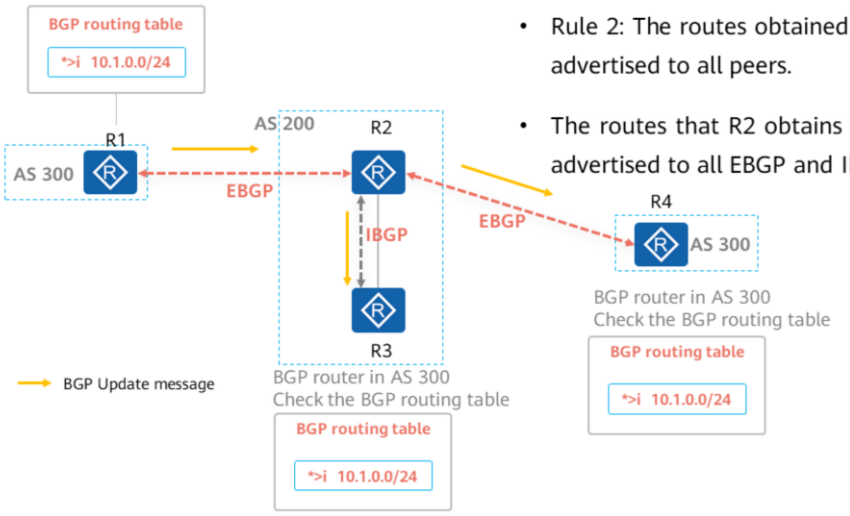
- Rule 1: Only the optimal and valid routes (that is, the next-hop address is reachable) are advertised.
- Run the **display bgp routing-table** command to check the BGP routing table.

```
Total Number of Routes: 2
     Network          NextHop        MED        LocPrf    PrefVal Path/Ogn
*>i  10.1.0.0/24      11.1.0.1       0          100       0       ?
*i                    11.1.0.2       0          100       0       ?
```

- The BGP routing table contains the optimal and valid routes with the following flags:
  - *: indicates the valid route.
  - >: indicates the optimal route.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

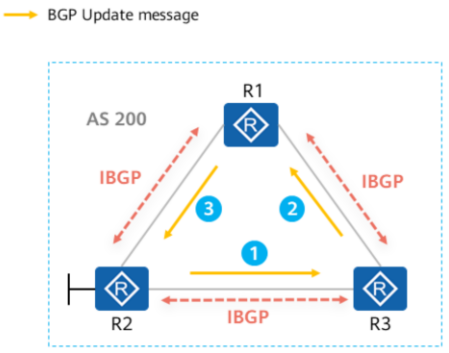
BGP Route Advertisement Rule 2



- Rule 2: The routes obtained from EBGP peers are advertised to all peers.
- The routes that R2 obtains from EBGP peers are advertised to all EBGP and IBGP peers.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BGP Route Advertisement Rule 3 (1)

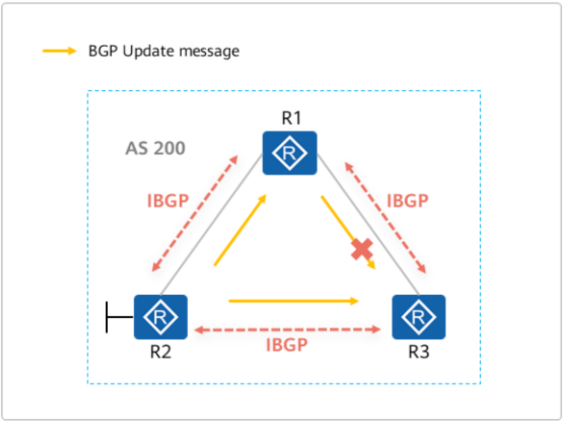


- Rule 3: The BGP routes obtained from an IBGP peer are not advertised to other IBGP peers.
- This is also called IBGP split horizon.
- As shown in the figure, if the routes learned by an IBGP peer are advertised to other IBGP peers, a routing loop occurs:
  - R2 advertises a route to its IBGP peer R3.
  - R3 advertises the received route to its IBGP peer R1.
  - R1 continues to advertise the route to its IBGP peer R2.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BGP Route Advertisement Rule 3 (2)

Full-mesh IBGP connections

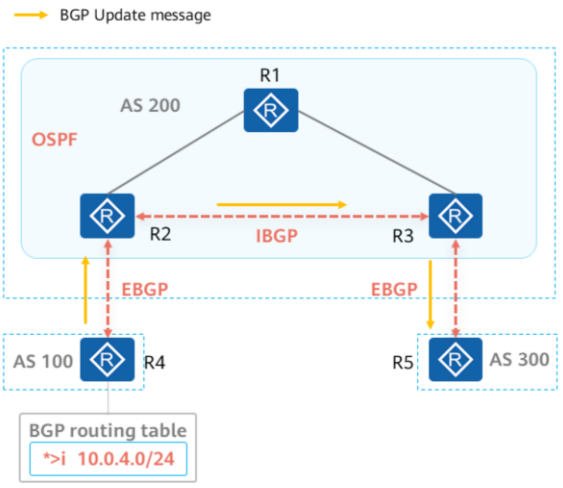


- Rule 3 may bring new problems. As shown in the figure on the left, when R2 sends a route to R1, R1 cannot send the route to R3 because of the limitation of rule 3. As a result, R3 cannot learn the route.
- To solve this problem, you can establish full-mesh IBGP peer relationships in an AS. Here, R2 and R3 establish an IBGP peer relationship so that R2 can advertise routes to R3.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BGP Route Advertisement Rule 4 (1)

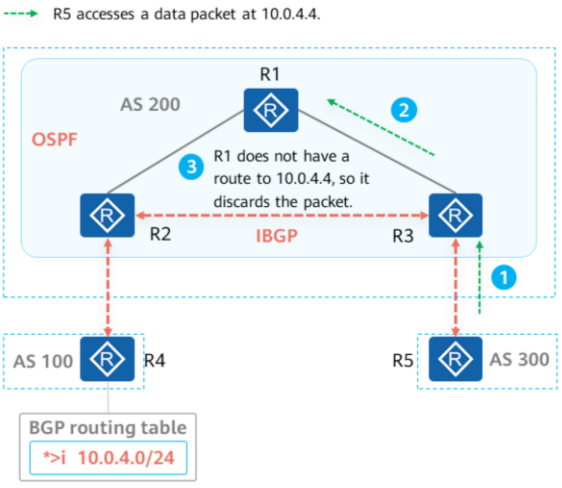


- Rule 4: When a router learns a BGP route (IBGP route) from its IBGP peer, the router cannot use this route or advertise this route to its EBGP peer unless the router learns this route from an IGP. In this situation, IBGP routes and IGP routes need to be synchronized. This rule is also called BGP synchronization rule.
- In the figure:
  1. R4 has a route to 10.0.4.0/24, which is advertised to R2.
  2. R2 advertises the route to its indirectly connected IBGP peer R3.
  3. R3 advertises the route to R5.
  4. R5 initiates an access request to 10.0.4.4.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BGP Route Advertisement Rule 4 (2)



- R5 accesses 10.0.4.4.
  1. R5 searches the routing table and sends the packet to R3.
  2. After receiving the packet, R3 searches the routing table and finds a BGP route with the next hop being R2. R2 is an indirect next hop and needs to recurse routes. The route learned through IGP is recursed to R1. R3 sends the packet to R1.
  3. After receiving the packet, R1 searches the routing table. Because R1 is not a BGP router and does not establish an IBGP peer relationship with R2, R1 does not have the BGP route to 10.0.4.0/24. As a result, R1 does not find the matching route and discards the packet.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

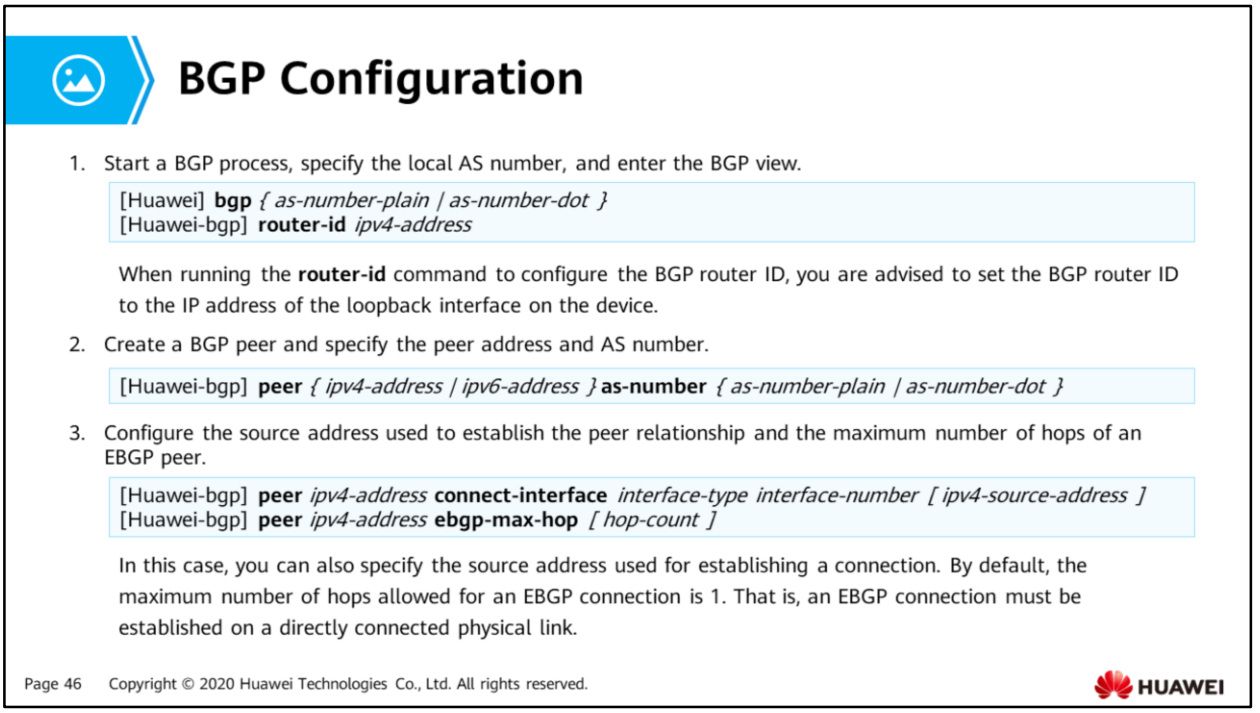
HUAWEI

- The root cause of this problem is that the BGP-incapable router in AS 200 does not have the route learned from BGP. As a result, R1 fails to find the route and discards the packet. BGP synchronization is defined as follows: BGP routes are advertised only when they exist in the IGP routing table. For example, in the figure, when R3 finds that the OSPF routing table does not contain the route to 10.0.4.0/24, R3 does not advertise the route to R5. This prevents subsequent access failures.
- The solutions are as follows:
  - BGP routes are redistributed to an IGP. This mode is seldom used.
  - Fully-mesh IBGP peer relationships are established so that all routers on the network have BGP routes.

Contents

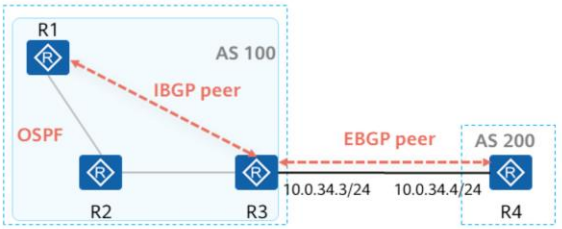
1. Introduction to BGP
2. Basic Concepts of BGP
3. **Basic BGP Configurations**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- If the router ID is not set, BGP selects the router ID in the system view as the router ID. For router ID selection rules in the system view, see the description about the router-id command.

Configuration Example (1)



- The figure shows BGP peer relationships, AS numbers, and device interconnection addresses.
- The IP address of Loopback1 on each device is 10.0.x.x/32, where x is the device ID. All devices use the IP address of Loopback1 as the router ID.
- R1 and R3 establish an IBGP peer relationship using the IP address of Loopback1 as the source address of a TCP connection; R3 and R4 establish an EBGP peer relationship using the IP address of the interconnected interface as the source address.

R1 configuration:

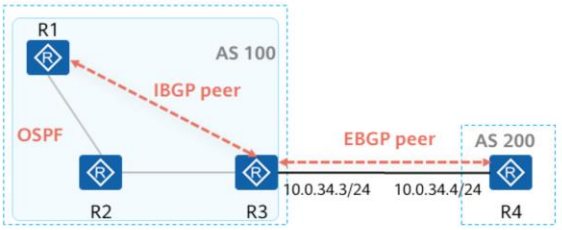
```
[R1] bgp 100
[R1-bgp] router-id 10.0.1.1
[R1-bgp] peer 10.0.3.3 as-number 100
[R1-bgp] peer 10.0.3.3 connect-interface LoopBack1
```

R3 configuration:

```
[R3] bgp 100
[R3-bgp] router-id 10.0.3.3
[R3-bgp] peer 10.0.1.1 as-number 100
[R3-bgp] peer 10.0.1.1 connect-interface LoopBack1
[R3-bgp] peer 10.0.34.4 as-number 200
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (2)



R4 configuration:

[R4] **bgp 200**
[R4-bgp] **router-id 10.0.4.4**
[R4-bgp] **peer 10.0.34.3 as-number 100**

- The figure shows BGP peer relationships, AS numbers, and device interconnection addresses.
- The IP address of Loopback1 on each device is 10.0.x.x/32, where x is the device ID. All devices use the IP address of Loopback1 as the router ID.
- R1 and R3 establish an IBGP peer relationship using the IP address of Loopback1 as the source address of a TCP connection; R3 and R4 establish an EBGP peer relationship using the IP address of the interconnected interface as the source address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (3)

Check the BGP peer relationship on R3.

```
<R3> display bgp peer
 BGP Local router ID : 10.0.3.3
 local AS number : 100
 Total number of peers : 2
 Peers in established state : 2

 Peer        V   AS   MsgRcvd   MsgSent   OutQ   Up/Down      State         PrefRcv

 10.0.1.1    4   100  0         0         0      00:00:07     Established   0
 10.0.34.4   4   200  32        35        0      00:17:49     Established   0
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Essay) What is the TCP destination port number used by BGP?
2. (Essay) What types of BGP peer relationships? What is the basis for classification?
3. (Multiple) Which of the following messages are used to establish a BGP peer relationship and update routes?( )
   A. Route-refresh
   B. Open
   C. Notification
   D. Update

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. 179
2. BGP peer relationships fall into IBGP and EBGP peer relationships, which are classified based on whether the two devices belong to the same AS.
3. B, D

Summary

- This course describes BGP basics, including the background of BGP, AS concept, and BGP characteristics.
- This course describes the process of establishing a BGP peer relationship and the BGP state machine, which helps you understand and memorize the process.
- Different from an IGP, BGP cannot discover or calculate routes by itself. Instead, routes are obtained from advertised routes in the IGP routing table.
- BGP advertises routes based on the four rules, which restrict the transmission of BGP routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



BGP Path Attributes and RRs

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- Each BGP route has multiple path attributes. When a router advertises a BGP route to its peer, the route carries multiple path attributes. These attributes describe route characteristics that affect BGP route selection in some scenarios.
- IBGP split horizon prevents routing loops in an AS, but also brings the following problem: A BGP route can be only transmitted to one next hop in an AS. If IBGP peers establish full-mesh connections, the load on devices increases.
- This course describes BGP path attributes and route reflectors (RRs).

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe common BGP path attributes and their functions.
  - Describe the concept and application scenarios of BGP RRs.
  - Describe the rules and working mechanism of BGP RRs.

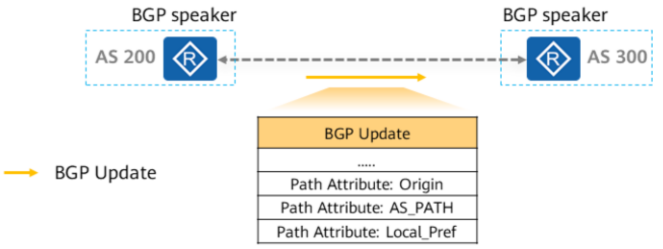
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **BGP Path Attributes**
2. BGP RRs

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

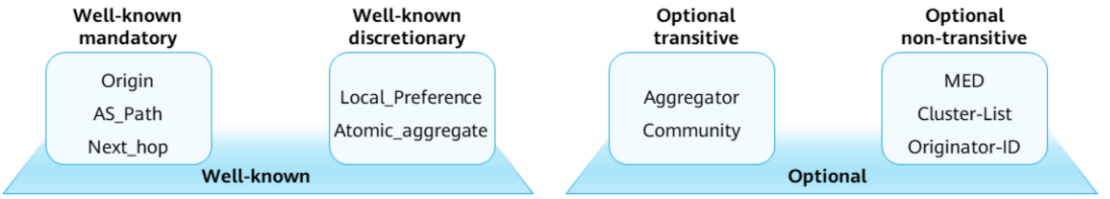
Path Attributes



- Each BGP route has multiple path attributes.
- When a router advertises a BGP route to its peer, it also advertises the path attributes carried in the route.
- BGP path attributes affect route selection.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Path Attribute Classification



- Well-known attributes can be identified by all BGP routers.
- Well-known attributes are classified into the following types:
  - Well-known mandatory: It must be included in each Update message.
  - Well-known discretionary: It may be included in some Update messages.
- Not all optional attributes need to be identified by BGP routers.
- Optional attributes are classified into the following types:
  - Optional transitive: A BGP device may not identify this type of attribute but still accepts and advertises them to peers.
  - Optional non-transitive: A BGP device may not identify this type of attribute, in which case it ignores them without advertising them to peers.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- There are many BGP attributes. This section lists only common BGP attributes.

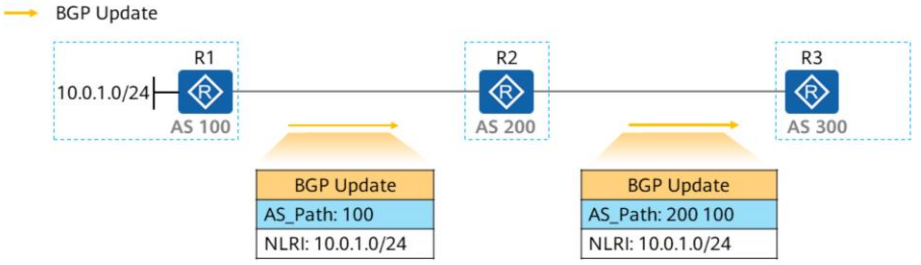
Example of BGP Update Messages

Border Gateway Protocol - UPDATE Message
- Marker: ffffffffffffffffffffffffffffffff
- Length: 53
- Type: UPDATE Message (2)
- Withdrawn Routes Length: 0
- Total Path Attribute Length: 28
- **Path attributes**
  - **Path Attribute - ORIGIN: IGP**
  - **Path Attribute - AS_PATH: 65536 1 2 3**
  - **Path Attribute - NEXT_HOP: 172.16.1.1**
- Network Layer Reachability Information (NLRI)
  - 30.0.0.0/8
    - NLRI prefix length: 8
    - NLRI prefix: 30.0.0.0

Path attribute in a BGP Update message

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

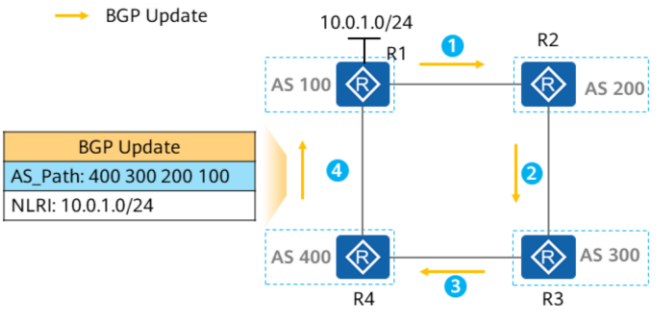
AS_Path



- This attribute is a well-known mandatory attribute. It is a list of AS numbers that the route to the destination network passes through.
- Function: Ensures that routes are transmitted between EBGP peers without loops. In addition, it is one of the attributes used for route selection.
- When a route is advertised to an EBGP peer, the router adds the local AS number to the AS_Path of the route. When a route is advertised to an IBGP peer, the AS_Path remains unchanged.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

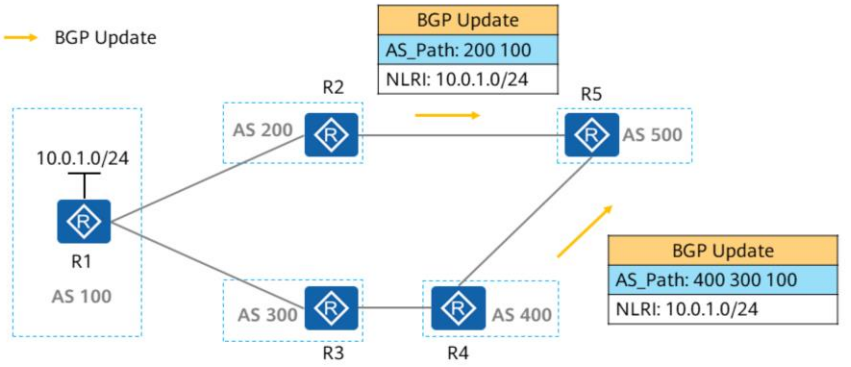
AS_Path Prevents Loops



The AS_Path attribute in the BGP Update message sent by R1 to R4 is 400 300 200 100. R1 does not accept the route because it has its own AS number. This prevents routing loops.

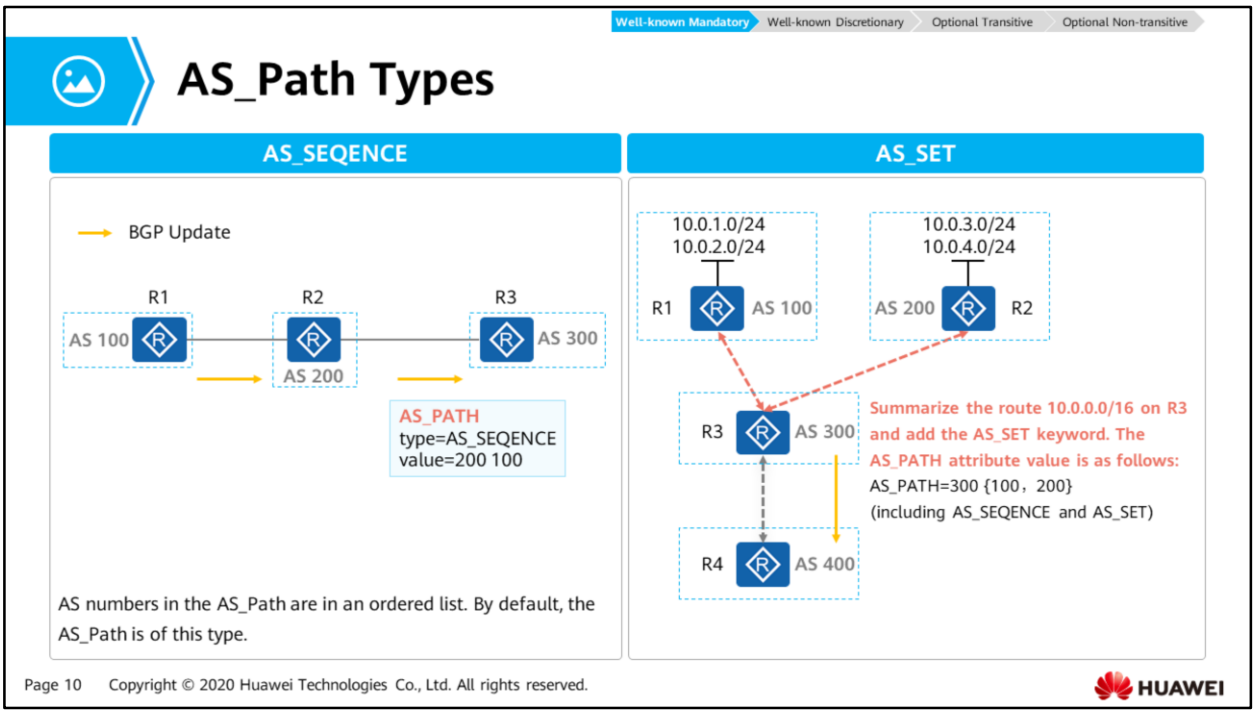
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

AS_Path Affects Route Selection



The AS_Path attribute affects BGP route selection. As shown in the figure, R5 learns BGP routes to the 10.0.1.0/24 network segment from both R2 and R4. If other conditions are the same, R5 preferentially selects the routes advertised by R2. This is because the AS_Path attribute of the route is short, that is, the number of AS numbers is small.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Route summarization reduces the device burden and shields specific routes to reduce the impact of route flapping. After routes are summarized, the AS_Path attribute is lost, which may cause routing loops. Therefore, the AS_Path of the AS_SET type can be used to carry AS numbers before route summarization.
- If the summarized route needs to carry the numbers of ASs through which all specific routes pass so as to prevent routing loops, you can specify the as-set parameter in the **aggregate** command.
- In the example of AS_SET, if routes are summarized in AS 300 and the as-set parameter is configured, the AS_Paths of specific routes are represented by an AS-Set set. The AS numbers in the brackets ({}) are not listed in sequence. The summarized route carries AS numbers to prevent loops.
- In addition to AS_SET and AS_AS_SEQENCE, AS_Path also has two types: AS_Confed_Sequence and AS_Confed_Set, which are used in BGP confederation and are not involved in this course.

Changing the AS_Path

You can use a route-policy to modify the AS_Path attribute of a BGP route in one of the following modes:

additive	overwrite	none overwrite
AS_Path: 100 ↓ apply as-path 300 additive ↓ AS_Path: 300 100	AS_Path: 100 200 300 ↓ apply as-path 400 overwrite ↓ AS_Path: 400	AS_Path: 100 200 300 ↓ apply as-path none overwrite ↓ AS_Path:
Add a new AS_Path value, which is placed preceding original AS_Path value.	Replace the existing AS_Path value with a new value.	Delete the existing AS_Path value.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Origin

Type	Flag	Description
IGP	i	If the route is imported to BGP by the originating BGP router through the **network** command, the Origin attribute of the BGP route is IGP.
EGP	e	If the route is learned through an EGP, the Origin attribute of the BGP route is EGP.
Incomplete	?	If the route is learned in other modes but not the preceding two modes, the Origin attribute is Incomplete. For example, routes are imported to BGP using the **import-route** command.

- This attribute is a well-known mandatory attribute and identifies the origin of a BGP route. As described in the table, the Origin attribute is classified into three types based on the mode in which routes are imported to BGP.
- If multiple routes with different Origin attributes are destined for the same destination and other conditions are the same, routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Origin Attribute in the BGP Routing Table

```
[R2] display bgp routing-table
BGP Local router ID is 10.0.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete


 Total Number of Routes: 4
      Network            NextHop        MED        LocPrf     PrefVal    Path/Ogn

 *>i  10.0.1.0/24        10.0.12.1      0          200        0          i
 * i                     10.0.23.3      0          100        0          i
```

Origin

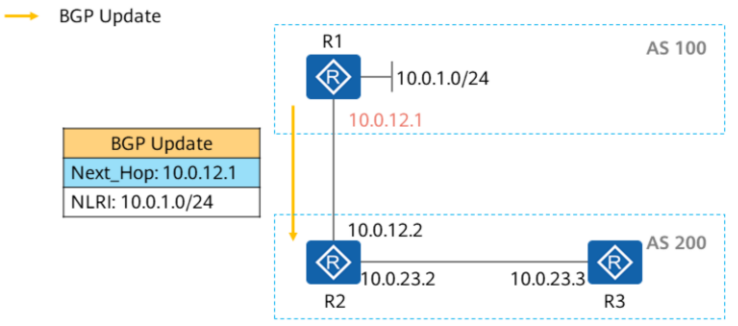
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Next_Hop

- This attribute is a well-known mandatory attribute and is used to specify the next-hop address to the destination network.
- After a router learns a BGP route, it checks the Next_Hop attribute of the BGP route. The Next_Hop attribute (IP address) must be reachable on the local router. If the Next_Hop attribute is unreachable, the BGP route is unavailable.
- The rules for setting the default Next_Hop attribute of BGP routes in different scenarios are as follows:
  - When advertising a BGP route to its EBGP peer, a router sets the Next_Hop attribute of the route to the source IP address of a TCP connection.
  - When receiving a BGP route advertised by an EBGP peer, the router does not change the Next_Hop attribute of the route that is advertised to its IBGP peer.
  - If a router receives a BGP route whose Next_Hop value is on the same network segment as the EBGP peer, the router retains the Next_Hop value of the route and advertises the route to its BGP peer.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

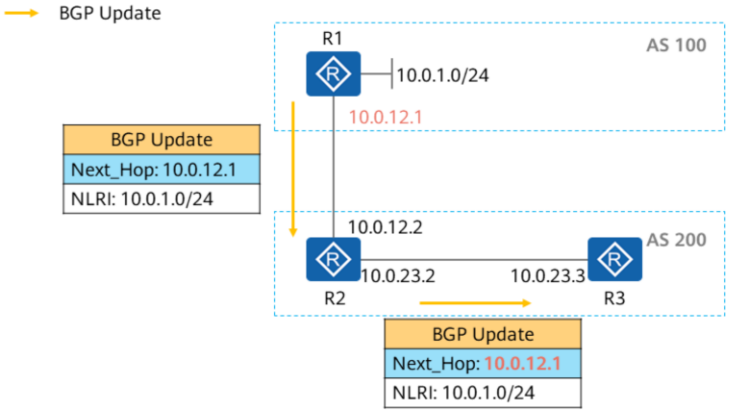
Default Operation for Next_Hop (1)



When advertising a BGP route to its EBGP peer, a router sets the Next_Hop attribute of the route to the source IP address of a TCP connection.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Default Operation for Next_Hop (2)

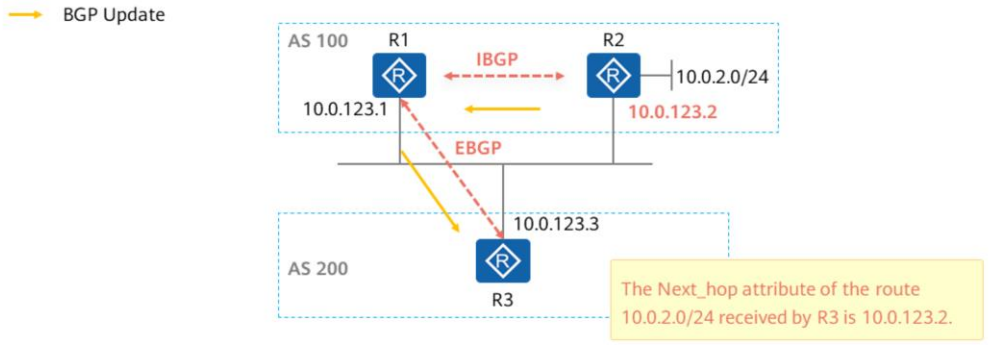


When receiving a BGP route advertised by an EBGP peer, the router does not change the Next_Hop attribute of the route that is advertised to its IBGP peer.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

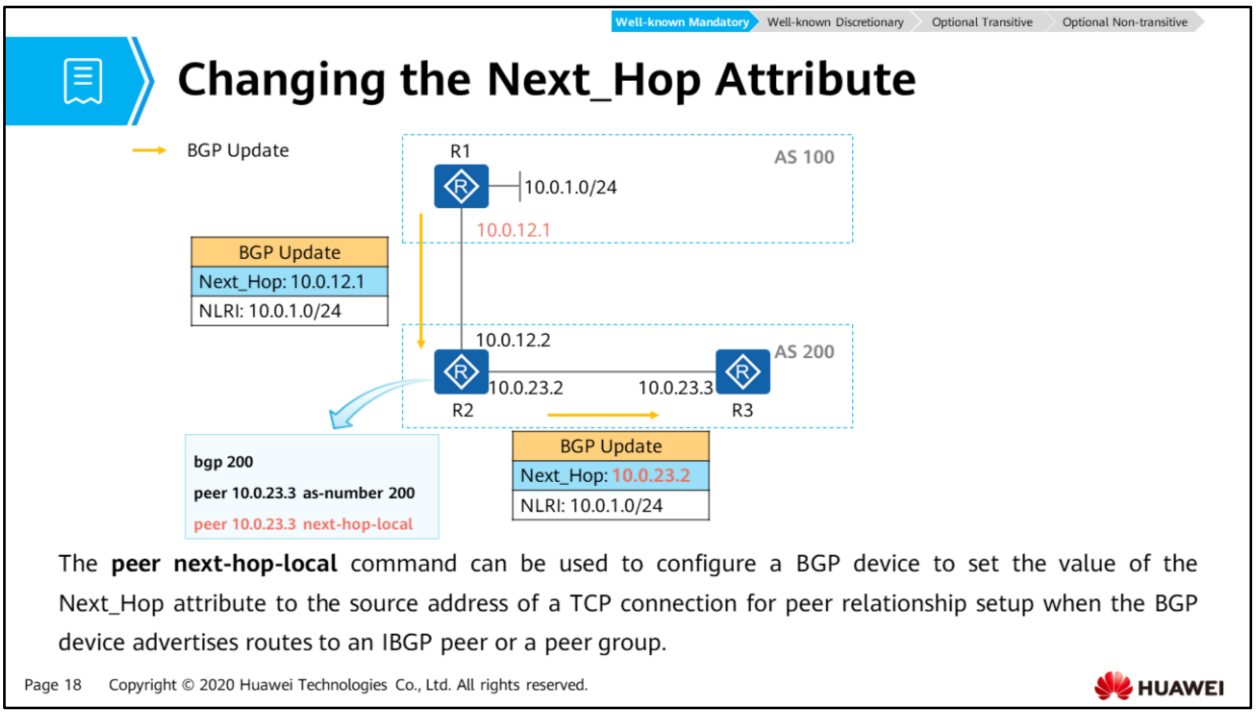
Default Operation for Next_Hop (3)



If a router receives a BGP route whose Next_Hop value is on the same network segment as the EBGP peer, the router retains the Next_Hop value of the route and advertises the route to its BGP peer.

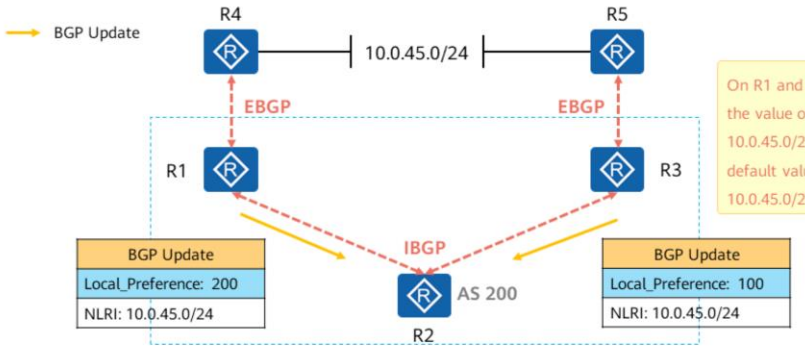
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI



- By default, the Next_Hop attribute of the BGP route 10.0.1.0/24 advertised by R2 to R3 is 10.0.12.1. If R2 does not advertise the route 10.0.12.0/24 to the IGP of AS 200, R3 cannot learn the route to 10.0.12.1. In this case, the next hop of the BGP route 10.0.1.0/24 is unreachable, so the route is considered invalid.

Local_Preference



On R1 and R3, configure a routing policy for R2 so that the value of the Local_Preference attribute in the route 10.0.45.0/24 sent by R1 to R2 is 200 and R3 retains the default value of Local_Preference. In this case, the route 10.0.45.0/24 advertised by R1 is preferred.

- The Local_Preference attribute is a well-known discretionary attribute. It can be used to inform routers in an AS of the preferred path to leave the AS.
- A larger value of the Local_Preference attribute indicates a better BGP route. The default value is 100.
- This attribute can be advertised only to IBGP peers but not EBGP peers.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Checking the Local_Preference Attribute in the BGP Routing Table

```
[R2] display bgp routing-table
BGP Local router ID is 10.0.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete


Total Number of Routes: 4
     Network            NextHop        MED        LocPrf     PrefVal  Path/Ogn

*>i  10.0.45.0/24       10.0.12.1      0          200        0        i
* i                     10.0.23.3      0          100        0        i
```

Local_Preference

The BGP route with the Local_Preference value of 200 takes precedence over the BGP route with the Local_Preference value of 100. In the BGP routing table, the BGP route from 10.0.12.1 is the optimal route.

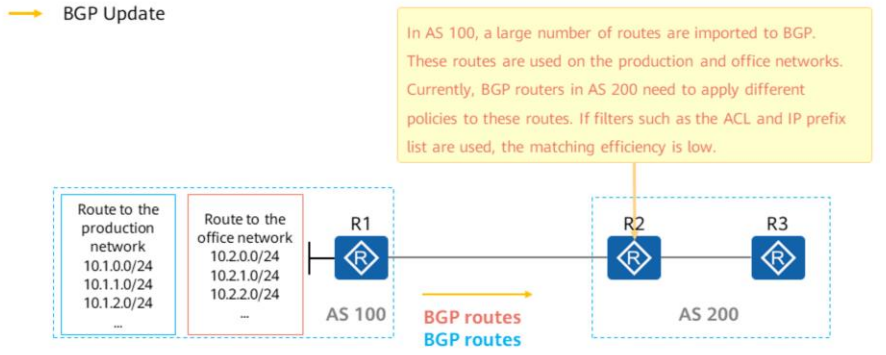
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Precautions for Local_Preference

- The Local_Preference attribute can be transmitted only between IBGP peers (the Local_Preference value is not lost during transmission between IBGP peers unless a policy is configured), but cannot be transmitted between EBGP peers. If the AS_Path attribute of a route received between EBGP peers carries Local_Preference, errors may occur in the BGP process.
- However, you can use the inbound policy on the AS border router to change the Local_Preference value. That is, after a route is received, the local device assigns the Local_Preference value to the route.
- You can run the **bgp default local-preference** command to change the default Local_Preference value. The default value is 100.
- When a router sends a route update to its EBGP peer, the route update cannot carry the Local_Preference attribute. After the peer receives the route, the peer assigns the default Local_Preference value (100) to the route and then advertises the route to its IBGP peer.
- For the routes imported by using the **network** or **import-route** commands, the Local_Preference value is 100 by default and can be transmitted to other IBGP peers in the AS. During the transmission, the Local_Preference value remains unchanged unless it is affected by a routing policy.

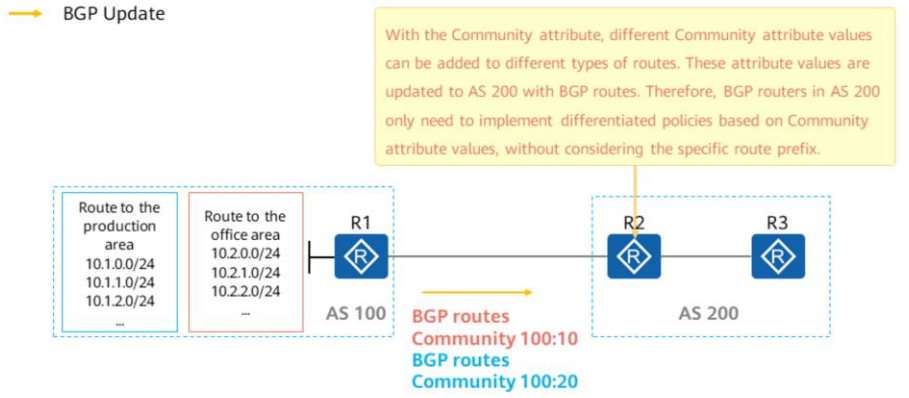
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Background of the Community Attribute (1)



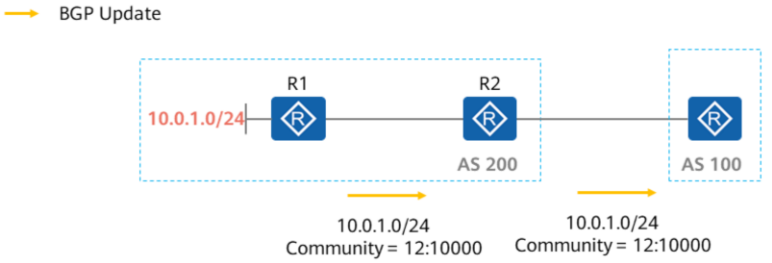
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Background of the Community Attribute (2)



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Community

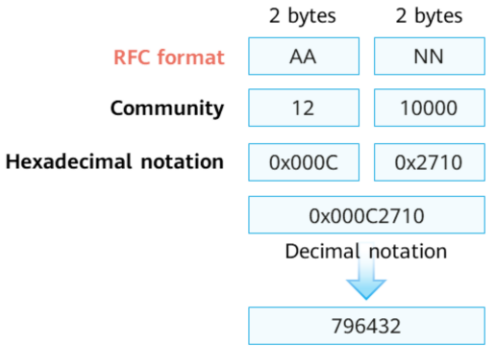


- The Community attribute is an optional transitive attribute. It is a route tag used to simplify the execution of routing policies.
- You can assign a specific Community attribute value to certain routes. Then the routes can be filtered based on the Community attribute value but not the network prefix or mask and the corresponding policy can be executed.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Well-known Mandatory | Well-known Discretionary | Optional Transitive | Optional Non-transitive

Community Attribute Format



The Community attribute has 32 bits, that is, 4 bytes. The value can be in either of the following formats:

- The value is an integer in decimal notation.
- The value is in AA:NN format, in which AA indicates the AS number and NN indicates a user-defined number.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The Community attribute includes self-defined and well-known community attributes.

Well-known Community Attribute

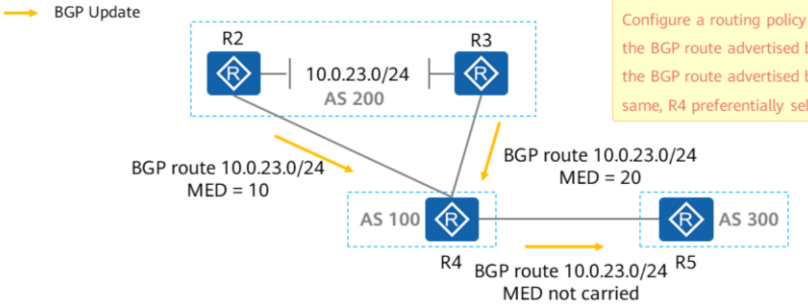
Community Attribute	Value	Description
Internet	0 (0x00000000)	A BGP device can advertise the received route with the Internet attribute to all peers. By default, all routes belong to the Internet community.
No_Advertise	4294967042 (0xFFFFFF02)	A BGP device does not advertise the received route with the No_Advertise attribute to any peer.
No_Export	4294967041 (0xFFFFFF01)	A BGP device does not advertise the received route with the No_Export attribute to devices outside the local AS.
No_Export_Subconfed	4294967043 (0xFFFFFF03)	A BGP device does not advertise the received route with the No_Export_Subconfed attribute to devices outside the local AS or to devices outside the local sub-AS.

RFC 1997 defines several well-known community attributes, as described in the table.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The No_Export_Subconfed community attribute involves the concept of BGP confederation and is not involved in this course.

MED



Configure a routing policy on R2 and R3 so that the MED value of the BGP route advertised by R2 to R4 is 10 and the MED value of the BGP route advertised by R3 is 20. If other conditions are the same, R4 preferentially selects the BGP routes advertised by R2.

- The multi-exit discriminator (MED) is an optional non-transitive attribute. It helps determine the optimal route when traffic enters an AS. That is, when there are multiple ingresses to the local AS, the AS can use the MED attribute to dynamically affect the path that other ASs select.
- A smaller MED value indicates a better BGP route.
- The MED is used for BGP route selection between ASs. After an MED is advertised to an EBGP peer, the route advertised by the peer in the AS carries the MED. When the route is advertised to the EBGP peer again, the route does not carry the MED by default.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



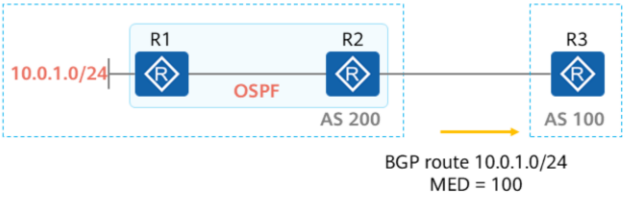
Precautions for the MED

- By default, a router compares the MED values of only the BGP routes from the same neighboring AS. That is, if two routes to the same destination are from different neighboring ASs, the router does not compare the MED values of the two routes.
- A BGP router determines whether to carry the MED attribute in a route advertised to its EBGP peers as follows (assume that no policy is configured on the EBGP peers):
  - If the BGP route is locally originated (imported using the **network** or **import-route** command), the BGP route carries the MED attribute by default and is sent to the EBGP peer.
  - If a BGP route is learned from a BGP peer, the route advertised to an EBGP peer does not carry the MED attribute by default.
  - When a route is transmitted between IBGP peers, the MED value is retained and transmitted with the route. The MED value will not be changed or lost during the transmission unless a policy is deployed.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

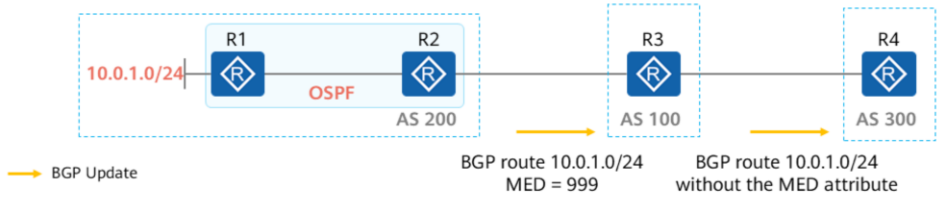
Default Operation for the MED Attribute (1)



- If a router learns a route through an IGP and imports the route to its BGP routing table using the network or import-route command, the MED value of the generated BGP route inherits the IGP metric value of the route. For example, in the preceding figure, R2 learns the route 10.0.1.0/24 through OSPF and the OSPF cost of the route is 100 in the global routing table of R2. When R2 imports the route into its BGP table using the network command, the MED value of the generated BGP route is 100.
- If a router imports a locally direct or static route to BGP using the network or import-route command, the MED of the BGP route is 0 because the cost of the direct or static route is 0.

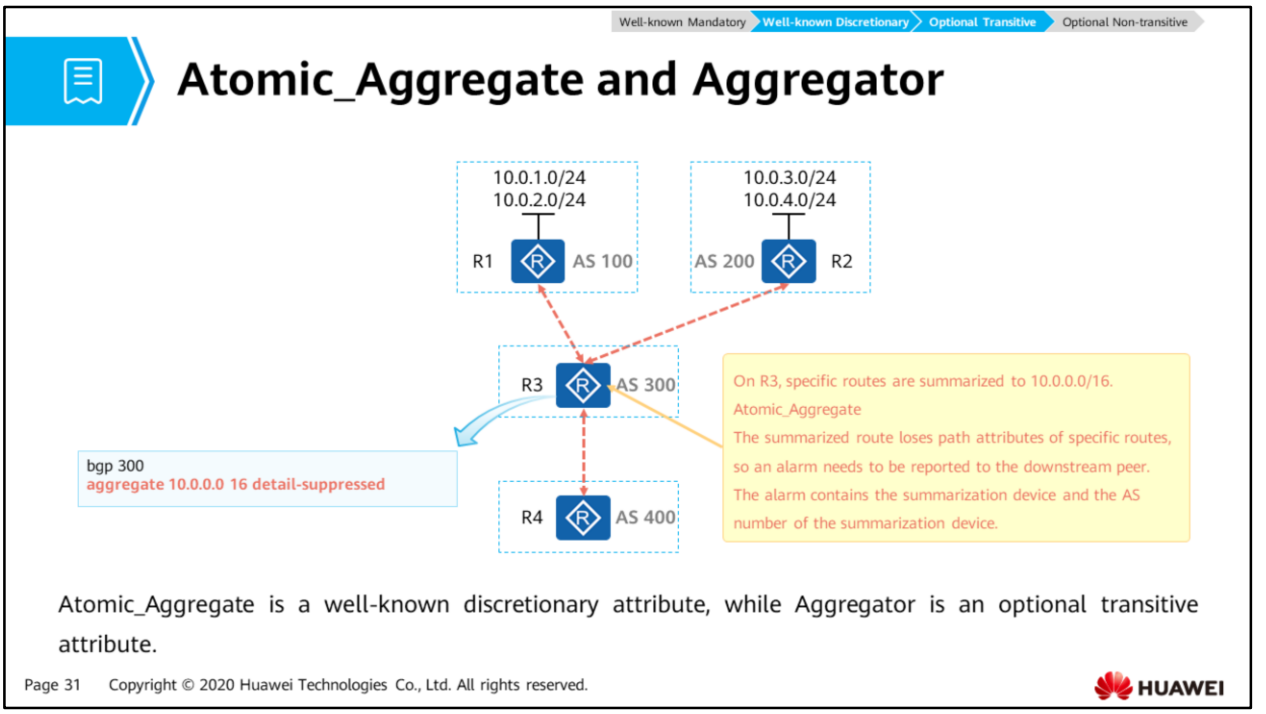
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Default Operation for the MED Attribute (2)



- If a router learns a route from a peer through BGP, the router does not transmit the MED value by default when updating the route to its EBGP peer. That is, the MED attribute is not transmitted across ASs. As shown in the figure, if R3 learns a BGP route with the MED attribute from R2, it advertises the route without the MED attribute to R4 by default.
- You can run the **default med** command to change the **default MED** value. The default med command takes effect only for the local routes imported using the **import-route** command and BGP summarized routes. For example, if **default med 999** is configured on R2, R2 advertises the MED value 999 to R3 with the route imported using the **import-route** command or generated using the **aggregate** command.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- The **aggregate** command on R3 summarizes BGP routes 10.0.1.0/24, 10.0.2.0/24, 10.0.3.0/24, and 10.0.4.0/24 into 10.0.0.0/16, and **detail-suppressed** is specified to prevent specific routes from being advertised. That is, R3 advertises only the summarized BGP route to R4.
- Atomic_Aggregate is a well-known discretionary attribute. It is a warning flag and does not carry any information. When a router receives a BGP route update and finds that the route carries the Atomic_Aggregate attribute, the router knows that the path attribute of the route may be lost. In this case, the router advertises the route with the Atomic_Aggregate attribute to other peers. In addition, the router that receives the route update cannot further redefine the route.
- The Aggregator attribute is an optional transitive attribute. When route summarization is performed, the router that summarizes routes can add the Aggregator attribute to the summarized route and record the local AS number and its router ID in the attribute. The Aggregator attribute is used to identify the AS and BGP router where routes are summarized.

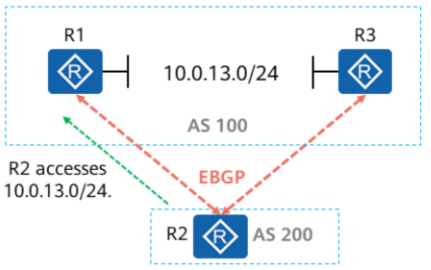
Checking the Summarized Route

```
[R4]display bgp routing-table 10.0.0.0 16
BGP local router ID : 10.0.4.4
 Local AS number : 400
 Paths:   1 available, 1 best, 1 select
 BGP routing table entry information of 10.0.0.0/16:
 From: 10.0.34.3 (10.0.3.3)
 Route Duration: 00h00m21s
 Direct Out-interface: GigabitEthernet0/0/0
 Original nexthop: 10.0.34.3
 Qos information : 0x0
 AS-path 300, origin igp, pref-val 0, valid, external, best, select, active, pre 255
 Aggregator: AS 300, Aggregator ID 10.0.3.3, Atomic-aggregate
 Not advertised to any peer yet
```

In detailed information about BGP routes, the Aggregator attribute records the AS number and router ID of the summarization device. In addition, the Atomic-Aggregate attribute indicates that the route is a summarized route.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferred-Value



Configure a routing policy (import policy) on R2 to set the Preferred-Value value of the route 10.0.13.0/24 advertised by R1 to 300 and the Preferred-Value value of the route 10.0.13.0/24 advertised by R3 to 200. In this way, R2 preferentially selects the route sent from R1 to reach 10.0.13.0/24.

- Preferred-Value is a Huawei-specific attribute and is valid only on the local device. If multiple routes are available to the same destination, the route with the largest PreVal value is selected as the optimal route.
- The value ranges from 0 to 65535. A larger value indicates a higher route priority.
- The Preferred-Value attribute can be configured only on the local router and affects only route selection of the local router. This attribute is not advertised to any BGP peer.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Checking the Preferred-Value in the BGP Routing Table

```
[R2] display bgp routing-table
BGP Local router ID is 10.0.2.2
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete

 Total Number of Routes: 4
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn
 *>   10.0.13.0/24       10.0.12.1      0                    300     100 i
 *                       10.0.23.3      0                    200     100 i
```

Preferred-Value

The BGP route with the Preferred-Value of 300 takes precedence over the BGP route with the Preferred-Value of 200. The BGP route from 10.0.12.1 in the BGP routing table is the optimal route.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

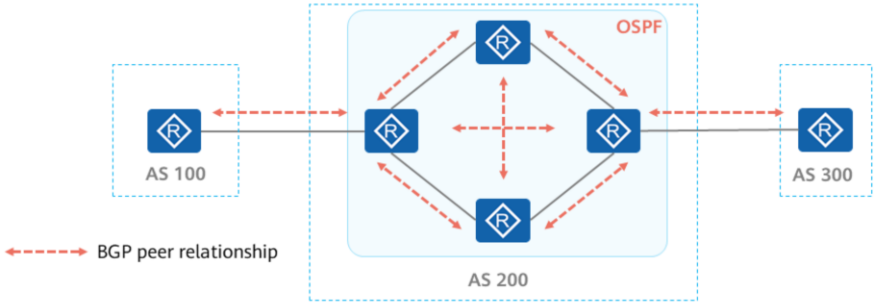
- The Preferred-Value attribute is abbreviated as PrefVal in the routing table.

Contents

1. BGP Path Attributes
2. **BGP RRs**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

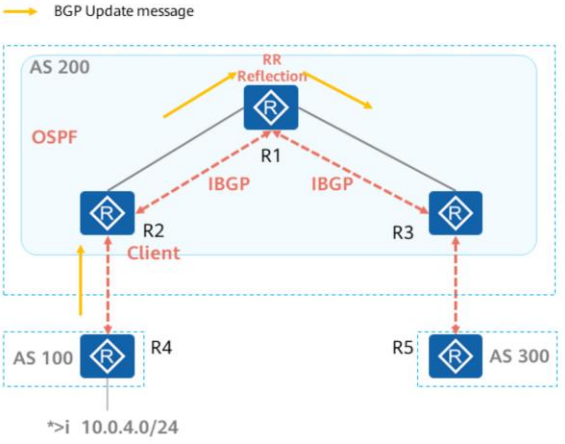
IBGP Issues in the Transit AS



- To ensure that all BGP routers in AS 200 can learn complete BGP routes, IBGP peers in an AS must establish full-mesh connections. However, there are the following disadvantages:
  - Routers need to maintain a large number of TCP and BGP connections, especially when there are a large number of routers.
  - The BGP network in an AS has poor scalability.
- Therefore, route reflector (RR) technology can be used.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

RR



- After the RR is imported, there are two roles:
  - RR
  - Client
- An RR reflects learned routes so that IBGP routes are advertised within an AS without establishing full-mesh IBGP connections.
- When specifying a BGP router as an RR, you need to specify its client. The client does not need to be configured and does not know that an RR exists on the network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Route Reflection Rules

When an RR receives BGP routes:

- If the RR learns an IBGP route from a non-client peer, the RR reflects the route to all clients.
- If the RR learns an IBGP route from its client, the RR reflects the route to all non-clients and all clients except the client that sends the IBGP route.
- If a route is learned from an EBGP peer, the route is advertised to all clients and non-client IBGP peers.

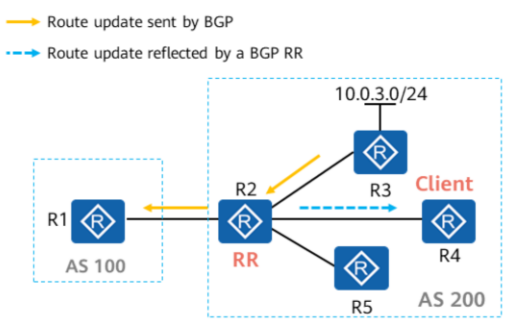
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- An RR reflects only the optimal BGP routes that it uses.

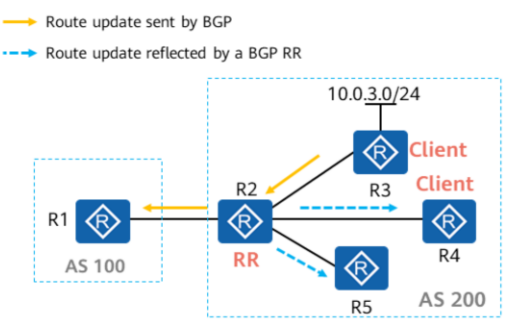
Reflection Rule Example (1)

Rule 1



If the RR learns an IBGP route from a non-client peer, the RR reflects the route to all clients.

Rule 2

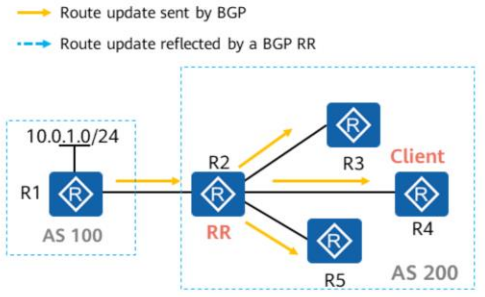


If the RR learns an IBGP route from its client, the RR reflects the route to all non-clients and all clients except the client that sends the IBGP route.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Reflection Rule Example (2)

Rule 3



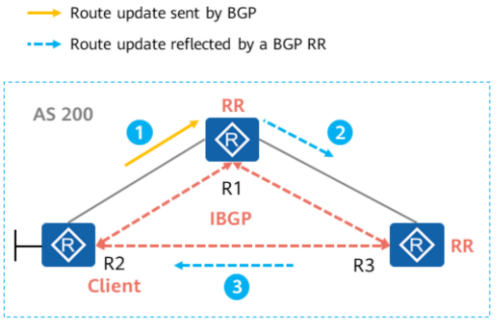
If a route is learned from an EBGP peer, the route is advertised to all clients and non-client IBGP peers.

- Pay attention to the difference between "reflection" and "sending".
- "Sending" refers to traditional BGP route transmission (scenario where the RR does not exist).
- "Reflection" refers to route transmission performed by an RR in compliance with route reflection rules. The RR inserts special path attributes into the reflected routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- When reflecting routes, the RR does not modify the following BGP path attributes: Next_Hop, AS_Path, Local_Preference, and MED. If the RR modifies these attributes, routing loops may occur.

Routing Loop Prevention in RR Scenarios

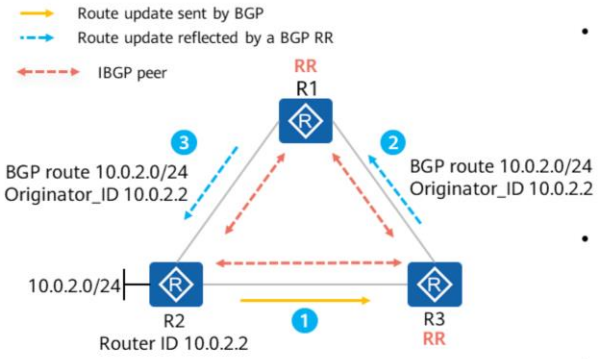


BGP route reflection may cause routing loops.

- RR deployment breaks the split horizon rule, which may cause routing loops. To prevent routing loops, the RR adds two special path attributes to BGP routes:
  - Originator_ID
  - Cluster_List
- Originator_ID and Cluster_List are optional non-transitive attributes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Originator ID



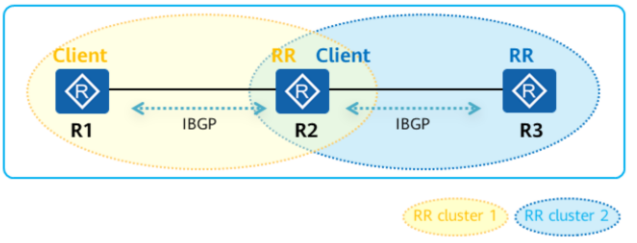
After R3 receives the BGP route 10.0.2.0/24 from R2, it adds the Originator_ID 10.0.2.2 to the route before reflecting the route to R1. After receiving the route, R1 reflects the route with the Originator_ID attribute to R2. After R2 receives the route, it finds that the Originator_ID attribute contains its own router ID and ignores the route update.

- When reflecting a BGP route, an RR adds the Originator_ID attribute to the reflected route. The Originator_ID attribute value is the router ID of the BGP router that advertises the route in the local AS.
- If multiple RRs exist in an AS, the Originator_ID attribute is created by the first RR and is not changed by subsequent RRs (if any).
- When a BGP router receives an IBGP route with the Originator_ID attribute and the Originator_ID attribute value is the same as its router ID, the BGP router ignores the route update.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

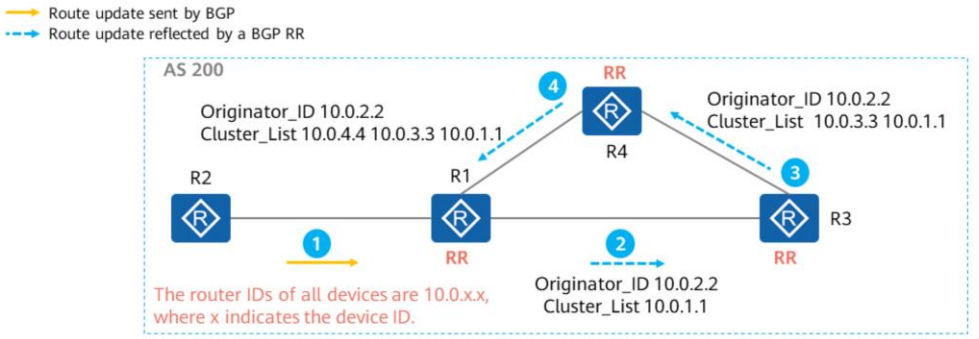
RR Cluster

- Each RR cluster consists of an RR and clients. An AS can have multiple RR clusters, as shown in the following figure.
- Each cluster has a unique cluster ID (Cluster_ID, which is the BGP router ID of an RR by default).
- Before an RR reflects a route, the RR places the local Cluster_ID on top of the Cluster_List.
- When an RR receives a BGP route carrying the Cluster_list attribute and the Cluster_ID of the RR is contained in the Cluster_List attribute, the RR considers that a loop occurs for the route and ignores the route update.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

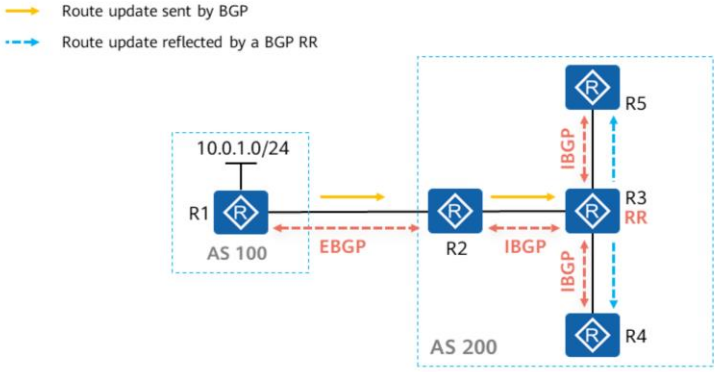
Cluster_List



- A route is sent from R2 to R1. When reflecting the route to R3, R1 adds the Originator_ID and Cluster_List (10.0.1.1) to the route. When R3 reflects the route to R4, the Cluster_List value is 10.0.3.3 10.0.1.1. When R4 reflects the route to R1, the Cluster_List value is 10.0.4.4 10.0.3.3 10.0.1.1.
- When R4 reflects the route to R1, R1 finds that the Cluster_List contains its own Cluster_ID and determines that a loop exists. Therefore, R1 ignores the route update.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Example of RR Application



R1 advertises the route 10.0.1.0/24 to BGP. R2 learns the route from R1 and advertises the route to R3. However, the IBGP route learned by R3 from R2 cannot be advertised to R4 or R5 because of the split horizon rule. Therefore, you can configure R3 as an RR and configure R4 and R5 as clients so that R4 and R5 can learn the BGP route 10.0.1.0/24.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BGP Configuration

1. Specify an RR and its client.

   [Huawei-bgp] **peer** {*group-name* | *ipv4-address* } **reflect-client**

   By default, the RR and its client are not configured.

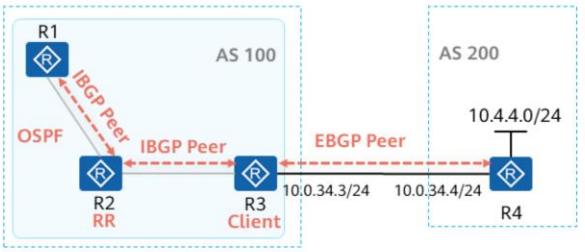
2. Configure the cluster ID of the RR.

   [Huawei-bgp] **reflector cluster-id** *cluster-id*

   By default, each RR uses its router ID as the cluster ID.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (1)



- The IP address of Loopback0 on a device is 10.0.x.x/32, where x is the device ID. All devices use the address of Loopback0 as the BGP router ID.
- R1, R2, and R3 belong to AS 100. OSPF runs in AS 100, and IP addresses of all directly connected interfaces are advertised to OSPF.
- In AS 100, the IP address of the loopback interface is used as the source address for IBGP update. R2 functions as the RR, and R3 functions as the client.
- R4 belongs to AS 200 and uses the IP address of the interconnected interface to establish an EBGP peer relationship with R3. R4 advertises 10.4.4.0/24 to BGP.

R1 configuration:

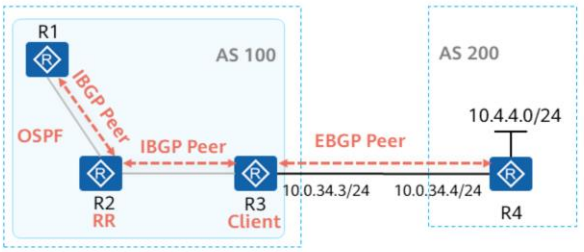
```
[R1] bgp 100
[R1-bgp] router-id 10.0.1.1
[R1-bgp] peer 10.0.2.2  as-number 100
[R1-bgp] peer 10.0.2.2 connect-interface LoopBack0
```

R2 configuration:

```
[R2] bgp 100
[R2-bgp] router-id 10.0.2.2
[R2-bgp] peer 10.0.1.1 as-number 100
[R2-bgp] peer 10.0.1.1 connect-interface LoopBack0
[R2-bgp] peer 10.0.3.3 as-number 100
[R2-bgp] peer 10.0.3.3 connect-interface LoopBack0
[R2-bgp] peer 10.0.3.3 reflect-client
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (2)



- The IP address of Loopback0 on a device is 10.0.x.x/32, where x is the device ID. All devices use the address of Loopback0 as the BGP router ID.
- R1, R2, and R3 belong to AS 100. OSPF runs in AS 100, and IP addresses of all directly connected interfaces are advertised to OSPF.
- In AS 100, the IP address of the loopback interface is used as the source address for IBGP update. R2 functions as the RR, and R3 functions as the client.
- R4 belongs to AS 200 and uses the IP address of the interconnected interface to establish an EBGP peer relationship with R3. R4 advertises 10.4.4.0/24 to BGP.

R3 configuration:

```
[R3] bgp 100
[R3-bgp] router-id 10.0.3.3
[R3-bgp] peer 10.0.2.2  as-number 100
[R3-bgp] peer 10.0.2.2 connect-interface LoopBack0
[R3-bgp] peer 10.0.34.4 as-number 200
```

R4 configuration:

```
[R4] bgp 200
[R4-bgp] router-id 10.0.4.4
[R4-bgp] peer 10.0.34.3 as-number 100
[R4-bgp] network 10.4.4.0 24
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (3)

Check the BGP route 10.4.4.0/24 on R3 and R1.

```
[R3-bgp]display bgp routing-table 10.4.4.0 24
 BGP local router ID : 10.0.3.3
 Local AS number : 100
 Paths:   1 available, 1 best, 1 select
 BGP routing table entry information of 10.4.4.0/24:
 From: 10.0.34.4 (10.0.4.4)
 Route Duration: 00h04m36s
 Direct Out-interface: GigabitEthernet0/0/1
 Original nexthop: 10.0.34.4
 Qos information : 0x0
 AS-path 200, origin igp, MED 0, pref-val 0, valid, external, best, select,
active, pre 255
 Advertised to such 1 peers:
    10.0.2.2
```

```
[R1]display bgp routing-table 10.4.4.0 24
  ..........
 BGP routing table entry information of 10.4.4.0/24:
 From: 10.0.2.2 (10.0.2.2)          #The route is originated from R2.
 Route Duration: 00h00m19s
 Relay IP Nexthop: 10.0.12.2
 Relay IP Out-Interface: GigabitEthernet0/0/0
Original nexthop: 10.0.34.4     # The next-hop address remains
unchanged.
 Qos information : 0x0
 AS-path 200, origin igp, MED 0, localpref 100, pref-val 0, valid, internal,
best, select, active, pre 255, IGP cost 3
Originator: 10.0.3.3               #The route is originated from 10.0.3.3.
Cluster list: 10.0.2.2             #Cluster_ID is the router ID of R2.
 Not advertised to any peer yet
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Single) What is the type of the AS_Path attribute? ( )
   A. Optional transitive
   B. Optional non-transitive
   C. Well-known mandatory
   D. Well-known discretionary
2. (Essay) What is the function of MED? Can the MED be transmitted across ASs?
3. (Essay) Which path attributes are used by an RR to prevent routing loops?

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. C
2. When the local AS has multiple ingresses, the MED can be used to determine the path through which other ASs enter the local AS. The MED is an optional non-transitive attribute and cannot be transmitted across ASs.
3. Originator_ID and Cluster_ID

Summary

- BGP carries various path attributes when advertising routes. These attributes describe route characteristics and affect BGP route selection in some scenarios.
- BGP path attributes are classified into well-known and optional attributes. All BGP routers must identify well-known attribute. Optional attributes do not need to be identified by all BGP routers.
- To solve the problem of IBGP split horizon, you can use full-mesh IBGP connections. However, a large number of IBGP peer relationships need to be maintained. You can deploy RRs to reduce the number of IBGP peer relationships.
- To prevent routing loops during RR deployment , BGP adds two path attributes: Originator_ID and Cluster_ID.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Preferred BGP Route Selection

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- The border gateway protocol (BGP) is a widely deployed routing protocol in the globe. BGP defines multiple path attributes and has various routing policy tools, providing flexible route control and path selection.
- Operations on BGP route attributes may affect route selection and therefore affect network traffic. Therefore, it is important to master Rules for Selecting a Preferred BGP Route.
- This course illustrates Rules for Selecting a Preferred BGP Route.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

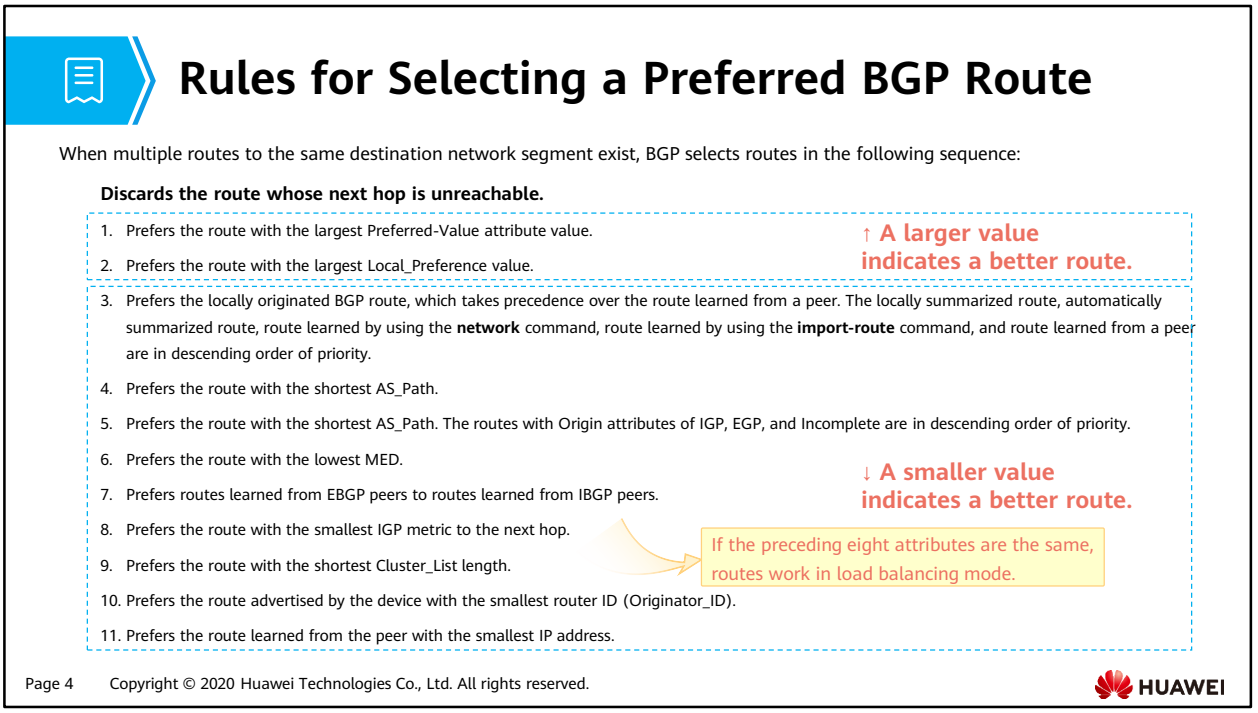
- Upon completion of this course, you will be able to:
  - Describe rules for selecting a preferred BGP route.
  - Learn BGP route control.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

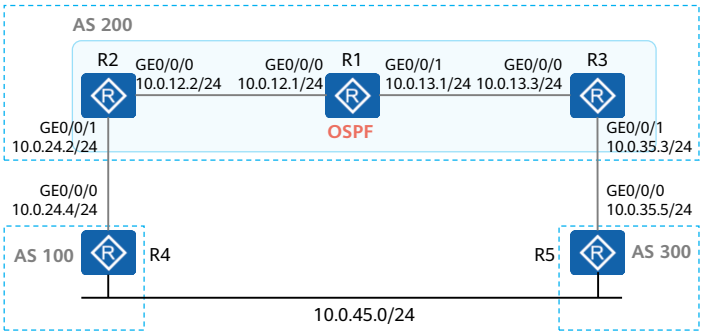
1. Preferred BGP Route Selection

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- The preceding rules are arranged in sequence. BGP selects the optimal route based on the first rule. If the first rule cannot help determine the optimal route, for example, the Preferred-Value attributes of routes are the same, BGP continues to use the next rule. If BGP can determine the optimal route using the current rule, no further action is required.
- This course provides the 12 most important BGP route selection rules. The following describes and verifies the preceding rules one by one.
- Terms such as "the eighth routing rule" may be mentioned in subsequent slides and correspond to the eighth routing rule listed on this page.
- Accumulated Interior Gateway Protocol (AIGP) is used to transmit and accumulate IGP metrics. This attribute is seldom used and is not involved in BGP route selection rules.

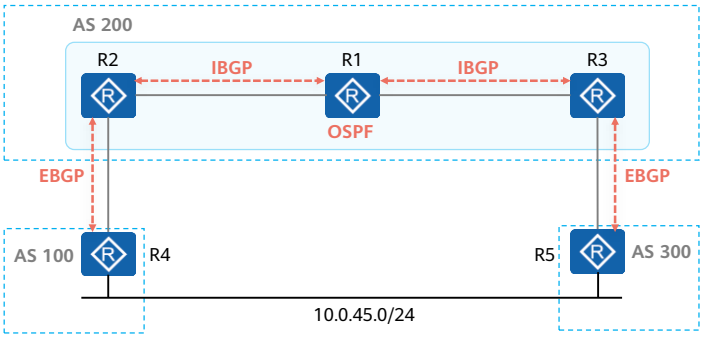
Topology (1)



- The figure shows ASs and interconnection addresses. Loopback0 interfaces are created on all devices, and the IP address is 10.0.x.x (x indicates the device ID). All devices use addresses of Loopback0 interfaces as router IDs.
- OSPF runs in AS 200 and OSPF is enabled on internal interconnection interfaces (excluding the interfaces connected to external AS) and loopback interfaces.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Topology (2)



- Loopback0 interfaces are used for establishing IBGP peer relationships in ASs, and directly connected interfaces are used for establishing EBGP peer relationships between ASs.
- R4 and R5 have the same network segment 10.0.45.0/24. The import-route command can be used to import the direct routes of this network segment to BGP so as to verify BGP route selection.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

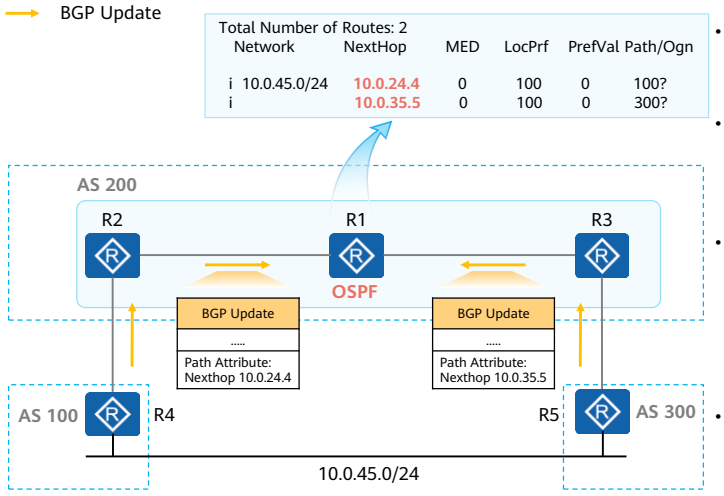
When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the **network** command, route learned by using the **import-route** command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

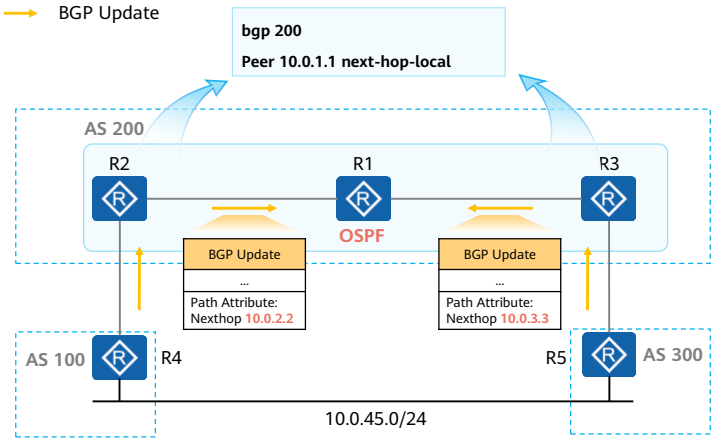
Discarding the Route Whose Next Hop Is Unreachable (1)



- When R4 and R5 advertise the BGP routes 10.0.45.0/24 to AS 200, the Next_Hop attribute values of the routes are 10.0.24.4 and 10.0.35.5.
- R2 and R3 do not modify the Next_Hop attribute when advertising routes to R1. The next hops of the two BGP routes 10.0.45.0/24 learned by R1 are 10.0.24.4 and 10.0.34.5.
- When R1 performs recursive query for next hops of BGP routes, route recursion fails because OSPF is not activated on the interfaces connecting R2 and R3 to external ASs. As a result, the next hop of the BGP route 10.0.45.0/24 on R1 is unreachable.
- Run the **display bgp routing** command on R1 to check the BGP routing table. The command output shows that the BGP route 10.0.45.0/24 is invalid.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Discarding the Route Whose Next Hop Is Unreachable (2)



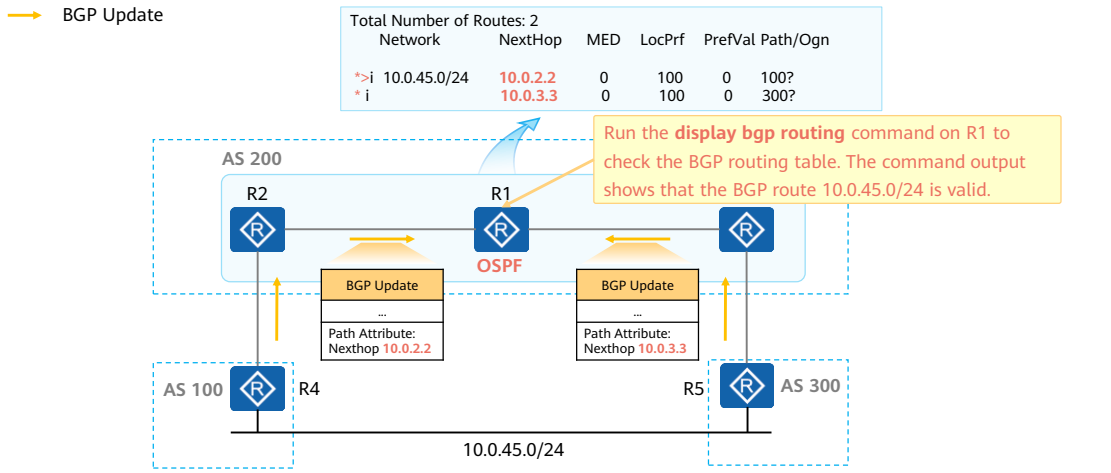
- Run the **next-hop-local** command on R2 and R3 to change the Next_Hop attribute value to the local source address.
- When R2 and R3 advertise BGP routes to R1, the Next_Hop attribute values of the routes are changed to 10.0.2.2 and 10.0.3.3.
- The two next-hop addresses can be successfully recursed on R1, and the next-hop address of the BGP route becomes reachable.

Unless otherwise specified, devices in all subsequent cases use basic configuration. In addition to the basic configuration, R2 and R3 are configured with **peer next-hop-local**.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- By default, the **peer next-hop-local** command is configured on R2 and R3. R1 preferentially selects the BGP route 10.0.45.0/24 advertised by R2.

Discarding the Route Whose Next Hop Is Unreachable (3)



Why is the BGP route with the next hop of 10.0.2.2 the optimal route when next hops of two BGP routes are reachable?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

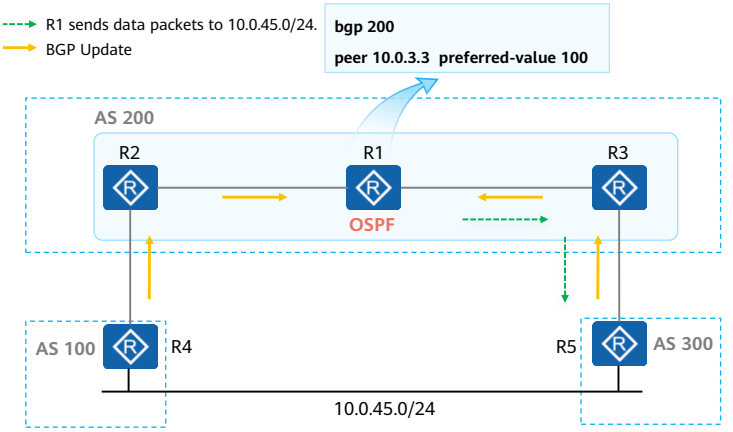
When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. **Prefers the route with the largest Preferred-Value attribute value.**
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the **network** command, route learned by using the **import-route** command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Changing the Preferred-Value Attribute



Run the **preferred-value** command to change the Preferred-Value attribute value of the BGP route advertised by R3 to 100, which takes precedence over the BGP route with the default Preferred-Value attribute value advertised by R2. Then R1 preferentially selects the BGP route 10.0.45.0/24 advertised by R3.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Checking the BGP Routing Table of R1

```
[R1] display bgp routing-table
BGP Local router ID is 10.0.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete

 Total Number of Routes: 4
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn
*>i   10.0.45.0/24       10.0.3.3       0                    100     300 i
* i                      10.0.2.2       0                    0       100 i
```

The BGP route advertised by R3 at 10.0.3.3 has a higher Preferred-Value attribute value (100), so R1 prefers the BGP route 10.0.45.0/24 advertised by R3.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

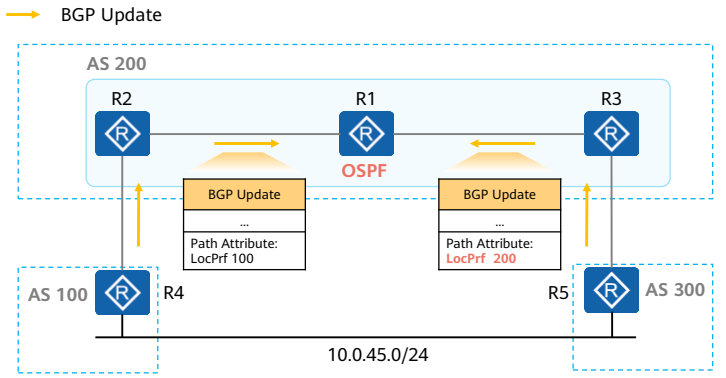
When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. **Prefers the route with the largest Local_Preference value.**
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the **network** command, route learned by using the **import-route** command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Changing the Local_Preference Attribute (1)



Run the following commands on R3.

```
ip ip-prefix local_pref index 10 permit 10.0.45.0 24
#
route-policy local_pref permit node 10
 if-match ip-prefix local_pref
 apply local-preference 200
route-policy local_pref permit node 20
#
bgp 200
peer 10.0.1.1 route-policy local_pref export
```

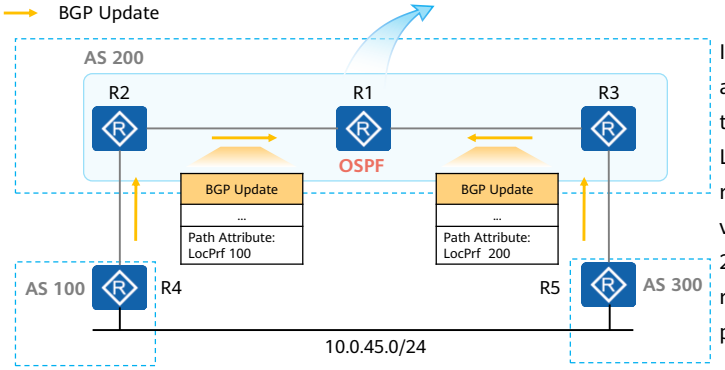
Configure a routing policy on R3 to change the Local_Preference value of the BGP route 10.0.45.0/24 advertised to R1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Changing the Local_Preference Attribute (2)

Total Number of Routes: 2

Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i 10.0.45.0/24	10.0.3.3	0	200	0	300?
* i	10.0.2.2	0	100	0	100?



If the next hops of the routes are reachable and the Preferred-Value attribute values of the routes are the same, R1 compares the Local_Preference attribute values of the routes. The Local_Preference attribute value of the BGP route advertised by R3 is 200, which is greater than that of the BGP route advertised by R2. Therefore, R1 prefers the BGP route advertised by R3.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. **Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the network command, route learned by using the import-route command, and route learned from a peer are in descending order of priority.**
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

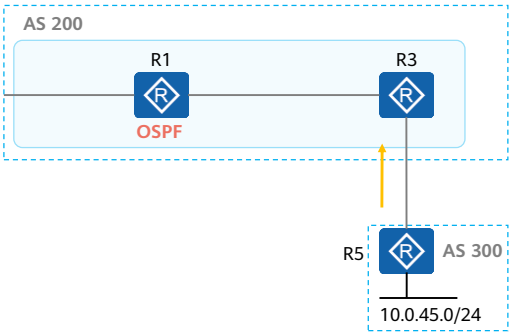
Preferring the Locally Generated Route

- In the case of identical conditions, a locally generated route is preferred, and the route learned from a peer has the secondary priority.
- In addition, locally generated routes may be learned in multiple ways. When the same route is learned in multiple ways, the following routes are in descending order of priority:
  - Summarized route by manually running the **aggregate** command in the BGP view
  - Automatically summarized route by running the **summary automatic** command
  - Route imported using the **network** command
  - Route imported using the **import-route** command

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- According to this rule:
  - The locally generated BGP route takes precedence over the BGP route learned from a peer.
  - The manually summarized route takes precedence over the automatically summarized route.

Manual Route Summarization (1)



To manually summarize routes on R3, configure two static routes pointing to null0 on R3 and import them to BGP.

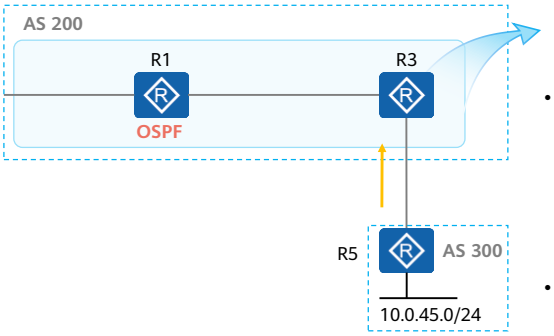
- Run the following commands on R3.

```
ip route-static 10.0.45.0 255.255.255.128 null0
ip route-static 10.0.45.128 255.255.255.128 null0
bgp 200
 aggregate 10.0.45.0 255.255.255.0 detail-suppressed
 import-route static
```

- Configure two static routes on R3, import the static routes to BGP using the **import-route** command, run the **aggregate** command to manually summarize the routes, and specify **detail-suppressed** suppress advertisement of specific routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Manual Route Summarization (2)



	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>	10.0.45.0/24	127.0.0.1			0	?
*		10.0.35.5	0		0	300?
s>	10.0.45.0/25	0.0.0.0	0		0	?
s>	10.0.45.128/25	0.0.0.0	0		0	?

- The BGP routing table of R3 contains two BGP routes 10.0.45.0/24.
  - Locally generated route: Static routes are imported to BGP and are manually summarized.
  - Route advertised by R5 at 10.0.35.5
- The two routes do not have Local_Preference or Preferred-Value attribute values on R3. R3 then compares the source of the two routes and prefers the route that is manually summarized.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The s flag in the BGP routing table indicates that the route is suppressed.

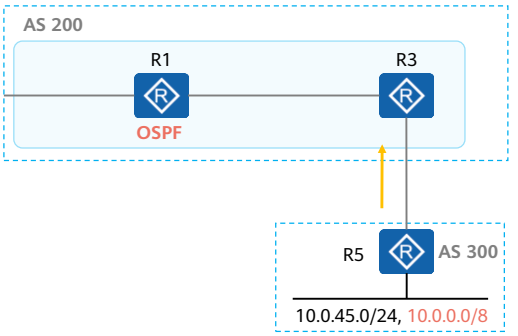
Manual Route Summarization (3)

```
[R3]display bgp routing-table 10.0.45.0 24
 BGP local router ID : 10.0.3.3
 Local AS number : 200
 Paths:   2 available, 1 best, 1 select
 BGP routing table entry information of 10.0.45.0/24:
 Aggregated route.
 Route Duration: 00h00m14s
 Direct Out-interface: NULL0
 Original nexthop: 127.0.0.1
 Qos information : 0x0
 AS-path Nil, origin incomplete, pref-val 0, valid, local, best, select, active,
pre 255
 Aggregator: AS 200, Aggregator ID 10.0.3.3, Atomic-aggregate
 Advertised to such 2 peers:
    10.0.35.5
    10.0.1.1
```

- Run the **display bgp routing-table 10.0.45.0 24** command on R3 to check detailed information about BGP route 10.0.45.0/24. The command output shows that there are two valid routes, and the manually summarized route is better.
- This example verifies that the locally generated BGP route is better than the BGP route learned from a peer.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Automatic Route Summarization (1)



In this case, the configurations on R1, R3, and R5 are irrelevant to the configurations that have been performed in the example of manual summarization.

- Run the following commands on R3.

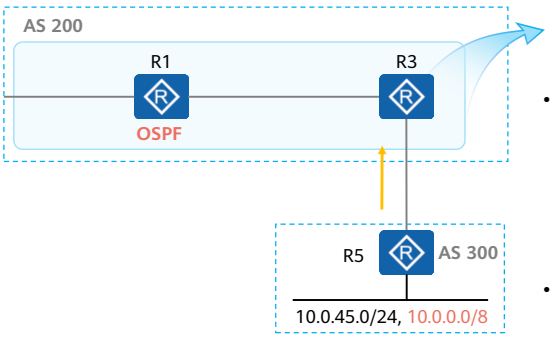
```
ip route-static 10.0.45.0 255.255.255.128 null0
ip route-static 10.0.45.128 255.255.255.128 null0

bgp 200
summary automatic
import-route static
```

- Configure two static routes on R3, import the static routes to BGP using the **import-route** command, and enable automatic summarization. BGP summarizes routes by natural network segment. For example, class A addresses 10.1.1.1/24 and 10.2.1.1/24 on the non-natural network segment are summarized into class A address 10.0.0.0/8 on the natural network segment. In addition, BGP advertises only the summarized route to peers.
- On R3, you can view that the route is summarized to 10.0.0.0/8.
- R5 imports the route 10.0.0.0/8 and advertises it to R3.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Automatic Route Summarization (2)

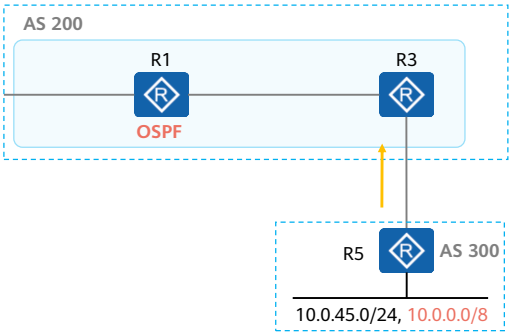


	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>	10.0.0.0	127.0.0.1			0	?
*		10.0.35.5	0		0	300?

- The BGP routing table on R3 contains two BGP routes 10.0.0.0.
  - Locally generated route: Static routes are imported to BGP and automatically aggregated.
  - Route advertised by R5 at 10.0.35.5
- The two routes do not have Local_Preference or Preferred-Value attribute values on R3. R3 then compares the source of the two routes and prefers the route that is automatically summarized.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Automatic Route Summarization (3)



- Perform manual summarization on R3.

```
bgp 200
 aggregate 10.0.0.0 255.0.0.0 detail-suppressed
```

- Check the routing table of R3.

	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>	10.0.0.0	127.0.0.1			0	?
*		127.0.0.1			0	?
*		10.0.35.5	0		0	300?

- The locally generated BGP route is preferred. However, there are two locally generated BGP routes, and the routing entry cannot help determine whether the preferred route is manually or automatically summarized.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Automatic Route Summarization (4)

```
BGP local router ID : 10.0.3.3
Local AS number : 200
Paths:   3 available, 1 best, 1 select
BGP routing table entry information of 10.0.0.0/8:
Aggregated route.
Route Duration: 00h08m17s
Direct Out-interface: NULL0
Original nexthop: 127.0.0.1
Qos information : 0x0
AS-path Nil, origin incomplete, pref-val 0, valid, local, best, select, active, pre 255
Aggregator: AS 200, Aggregator ID 10.0.3.3, Atomic-aggregate
Advertised to such 2 peers:
  10.0.35.5
  10.0.1.1
```

- Run the **display bgp routing-table 10.0.0.0** command on R3 to check detailed information about the BGP route 10.0.0.0/8. The command output shows that there are three valid routes, among which the optimal route is generated by route summarization and has the Atomic-aggregate attribute. The command output shows that the route is manually summarized.
- On R3, the manually summarized route is better than the automatically summarized route.
- This example verifies that the manually summarized route is better than the automatically summarized route.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The route imported using the **network** command is better than the route imported using the **import-route** command. Such an example is not provided.
- The automatically summarized route does not carry the Atomic-aggregate attribute.

Rules for Selecting a Preferred BGP Route

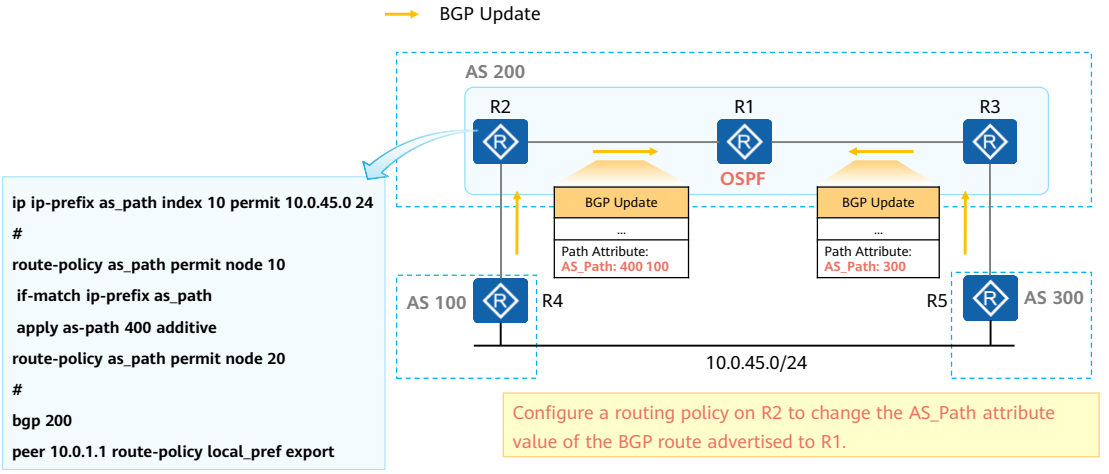
When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the network command, route learned by using the import-route command, and route learned from a peer are in descending order of priority.
4. **Prefers the route with the shortest AS_Path.**
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Shortest AS_Path (1)




```
ip ip-prefix as_path index 10 permit 10.0.45.0 24
#
route-policy as_path permit node 10
 if-match ip-prefix as_path
 apply as-path 400 additive
route-policy as_path permit node 20
#
bgp 200
peer 10.0.1.1 route-policy local_pref export
```

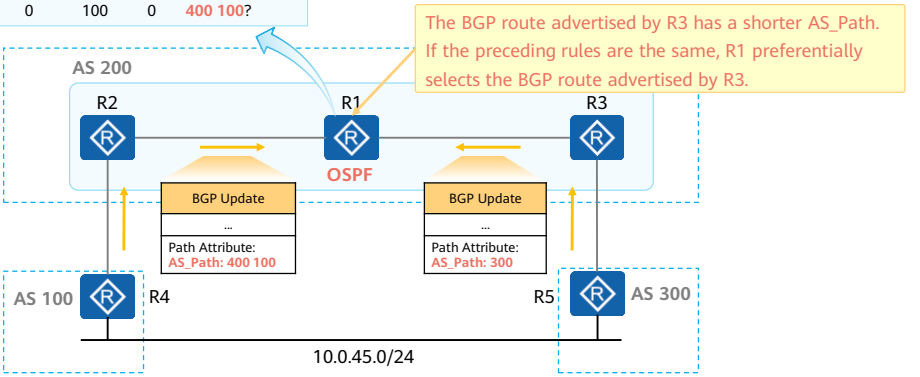
Configure a routing policy on R2 to change the AS_Path attribute value of the BGP route advertised to R1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Shortest AS_Path (2)

```
Total Number of Routes: 2
     Network          NextHop        MED        LocPrf    PrefVal Path/Ogn

*>i  10.0.45.0/24     10.0.3.3       0          100       0       300?
* i                   10.0.2.2       0          100       0       400 100?
```



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

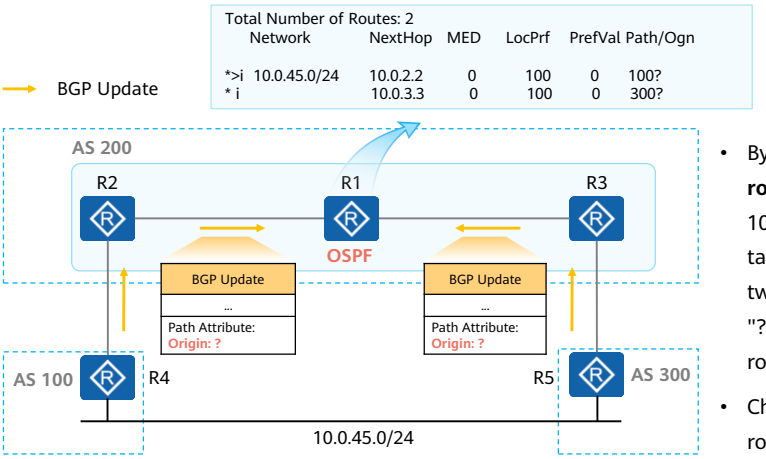
When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the network command, route learned by using the import-route command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. **Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.**
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

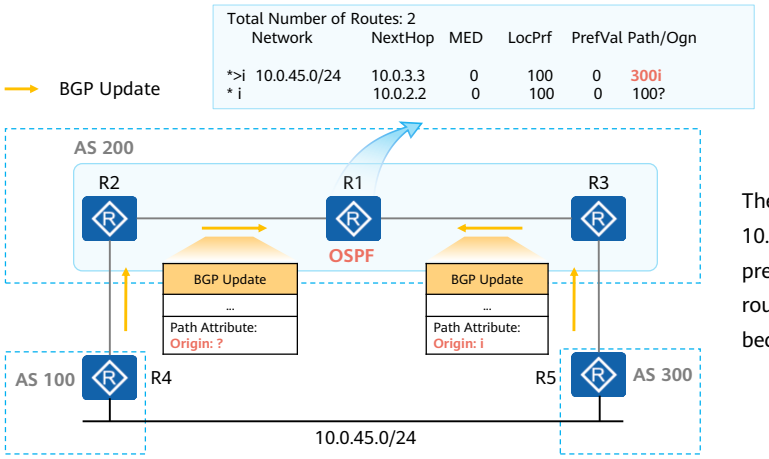
Origin Attribute Verification (1)



- By default, R4 and R5 use the **import-route** command to import the routes 10.0.45.0/24 to BGP. In the BGP routing table of R1, the Origin attributes of the two BGP routes 10.0.45.0/24 are both "?". R1 preferentially selects the BGP route imported by R4.
- Change the command used to import routes to **network** on R5.
- Check the BGP routing table on R1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Origin Attribute Verification (2)



The Origin attribute of the BGP route 10.0.45.0/24 imported by R5 is "i". If the preceding rules are the same, the BGP route with the origin type being "i" becomes the optimal route.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

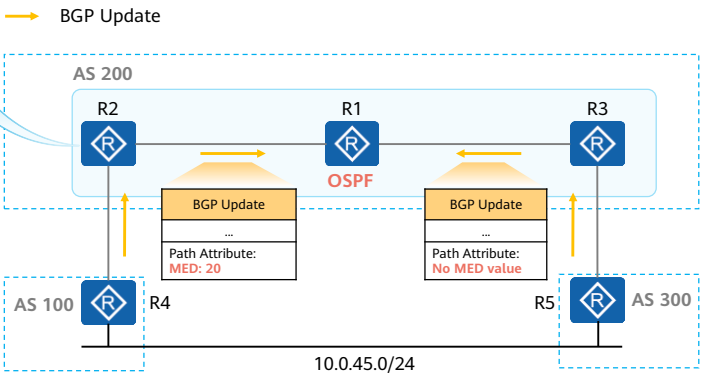
1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the network command, route learned by using the import-route command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. **Prefers the route with the lowest MED.**
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Smallest MED (1)

```
ip ip-prefix med index 10 permit 10.0.45.0 24
#
route-policy med permit node 10
 if-match ip-prefix med
 apply cost 20
route-policy med permit node 20
#
bgp 200
peer 10.0.1.1 route-policy med export
compare-different-as-med
```

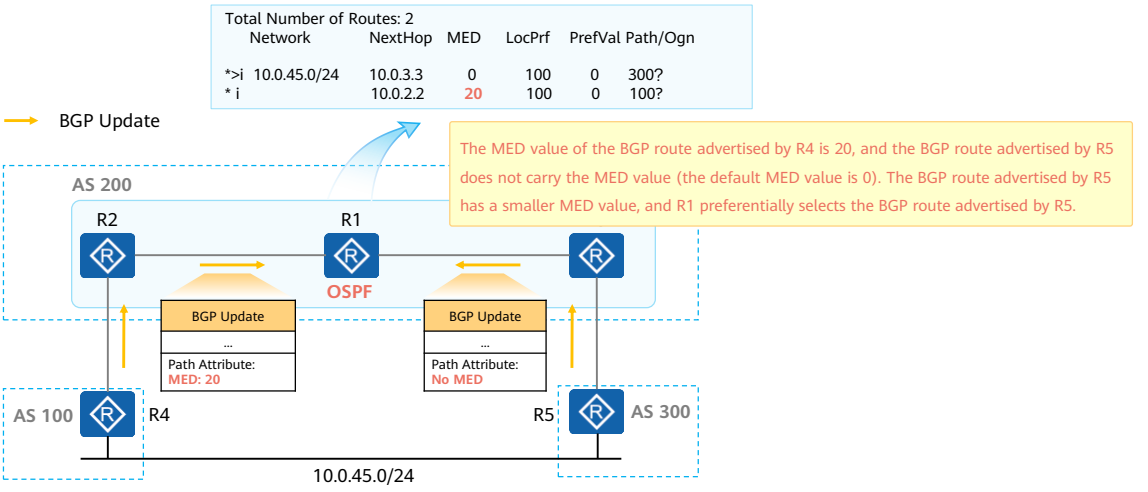
By default, BGP compares the MED values of routes from the same AS and to the same network segment. You can use a command to enable BGP to compare the MED attribute values of the same routes from different ASs.



Configure a routing policy on R2 to change the MED attribute values of the BGP routes advertised to R1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Smallest MED (2)



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

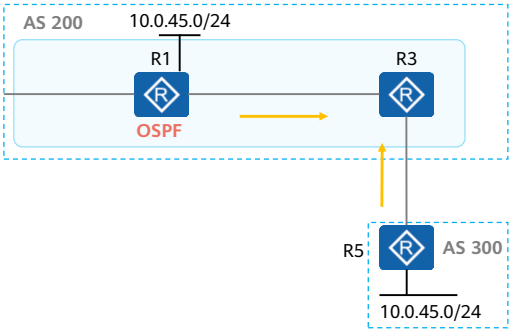
When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the network command, route learned by using the import-route command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. **Prefers routes learned from EBGP peers to routes learned from IBGP peers.**
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring a Route Learned from an EBGP Peer (1)



Create a static route 10.0.45.0/24 pointing to null0 on R1 and advertise the route to BGP. Ensure that the AS_Path attribute values of the BGP routes advertised by R1 and R5 to R3 are the same, configure a routing policy to add the AS_Path attribute to the route advertised by R1 to R3. The AS_Path attribute value is 500.

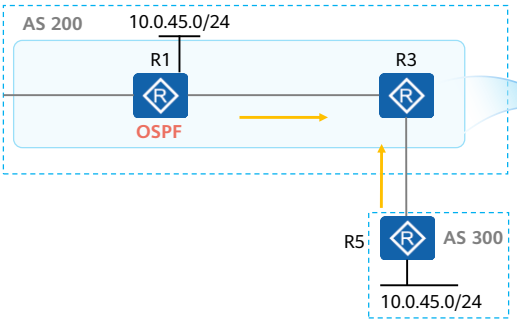
- Run the following commands on R1.

```
ip route-static 10.0.45.0 255.255.255.0 null0
ip ip-prefix ebgp index 10 permit 10.0.45.0 24
#
route-policy ebgp permit node 10
 if-match ip-prefix ebgp
 apply as-path 500 additive
route-policy ebgp permit node 20
#
bgp 200
  import-route static
  peer 10.0.3.3 route-policy ebgp export
```

- R3 will receive the BGP route 10.0.45.0/24 advertised by R1 and R5, and the preceding route selection rules cannot determine the optimal route.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring a Route Learned from an EBGP Peer (2)



	Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>	10.0.45.0/24	10.0.35.5	0		0	300?
* i		10.0.1.1	0	100	0	500?

R5 and R1 are the EBGP peer and IBGP peer, respectively. The BGP routes advertised by EBGP peers take precedence over the BGP routes advertised by IBGP peers. R3 preferentially selects the BGP routes advertised by R5.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring a Route Learned from an EBGP Peer (3)

```
BGP routing table entry information of 10.0.45.0/24:
From: 10.0.1.1 (10.0.1.1)
Route Duration: 00h06m43s
Relay IP Nexthop: 10.0.13.1
Relay IP Out-Interface: GigabitEthernet0/0/0
Original nexthop: 10.0.1.1
Qos information : 0x0
AS-path 500, origin incomplete, MED 0, localpref 100, pref-val 0, valid,
internal, pre 255, IGP cost 1, not preferred for peer type
Not advertised to any peer yet
```

Run the **display bgp routing-table 10.0.45.0 24** command on R3 to check detailed information about BGP routes. The command output is as follows:

not preferred for peer type

The route is not selected because the peer type is not preferred.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the network command, route learned by using the import-route command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. **Prefers the route with the smallest IGP metric to the next hop.**
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IGP Cost

```
BGP local router ID : 10.0.1.1
Local AS number : 200
Paths:   2 available, 1 best, 1 select
BGP routing table entry information of 10.0.45.0/24:
From: 10.0.3.3 (10.0.3.3)
Route Duration: 00h22m35s
Relay IP Nexthop: 10.0.13.3
Relay IP Out-Interface: GigabitEthernet0/0/1
Original nexthop: 10.0.3.3
Qos information : 0x0
AS-path 300, origin incomplete, MED 0, localpref 100, pref-val 0, valid,
internal, best, select, active, pre 255, IGP cost 1
Not advertised to any peer yet
```

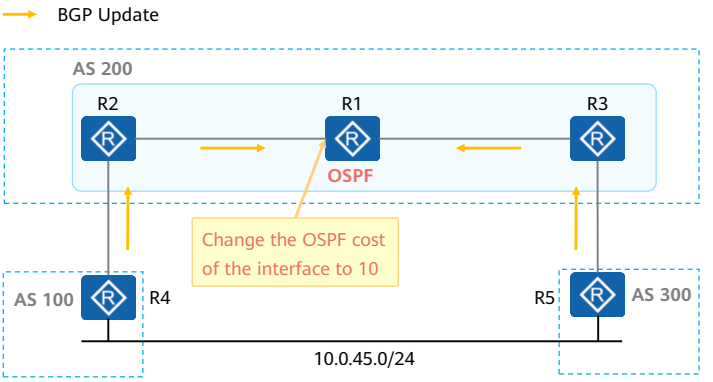
- The IGP cost is displayed in the detailed BGP route information. The IGP cost is the cost of the route to the original next hop in the local IP routing table.

Destination/Mask	Proto	Pre	Cost	NextHop	Interface
10.0.3.3/32	OSPF	10	1	10.0.13.3	GigabitEthernet0/0/1

- If the preceding seven rules cannot determine the optimal BGP route, the IGP cost of the next hop is compared.

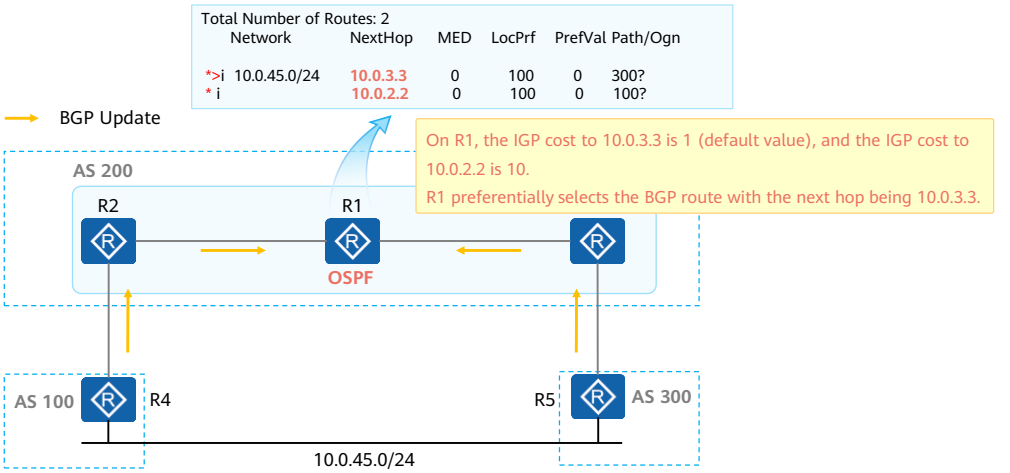
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Smallest IGP Cost (1)



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Preferring the Route with the Smallest IGP Cost (2)



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Smallest IGP Cost (3)

```
BGP routing table entry information of 10.0.45.0/24:
From: 10.0.2.2 (10.0.2.2)
Route Duration: 00h24m07s
Relay IP Nexthop: 10.0.12.2
Relay IP Out-Interface: GigabitEthernet0/0/0
Original nexthop: 10.0.2.2
Qos information : 0x0
AS-path 100, origin incomplete, MED 0, localpref 100, pref-val 0, valid,
internal, pre 255, IGP cost 10, not preferred for IGP cost
Not advertised to any peer yet
```

- Run the **display bgp routing-table 10.0.45.0 24** command on R1 to check detailed information about BGP routes. The command output shows that the IGP cost of the BGP route with the next hop being 10.0.2.2 changes to 10 and the IGP cost of the BGP route with the next hop being 10.0.3.3 is 1 (default value). Therefore, R1 preferentially selects the BGP route with the next hop being 10.0.3.3.
- The following information is displayed in the detailed routing information of R1:

  not preferred for IGP cost

  The route is not selected because of the IGP cost.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Load Balancing Among BGP Routes

- On a large network, there may be multiple valid BGP routes to the same destination. The device will select and add the optimal BGP route to its routing table for traffic forwarding.
- This, however, will result in uneven load balancing of much traffic. Configuring BGP load balancing can enable the device to add these multiple equal-cost BGP routes to its routing table, implementing traffic load balancing and reducing network congestion.
- After BGP load balancing is configured, the device will still select the optimal route among the multiple routes and advertise only this route to its peers.
- After BGP load balancing is enabled on a device, only the BGP routes that meet specified conditions can be used as equal-cost routes for load balancing.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

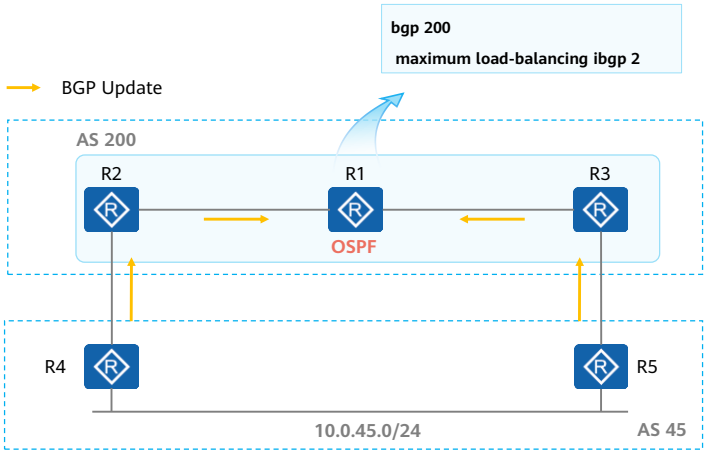
- By default, the device performs load balancing only for routes with the same AS_Path attribute. You can use **load-balancing as-path-ignore** to ignore inconsistency of the AS_Path attribute.
- Before routes to the same destination implement load balancing on a public network, a device determines the type of optimal route. If IBGP routes are optimal, only IBGP routes carry out load balancing. If EBGP routes are optimal, only EBGP routes carry out load balancing. This means that load balancing cannot be implemented using both IBGP and EBGP routes with the same destination address.

Conditions for Load Balancing Among Equal-Cost BGP Routes

- The Preferred-Value attribute values are the same.
- The Local_Preference attribute values are the same.
- All the routes are summarized or non-summarized routes.
- AS_Path attribute values are the same.
- Origin types (IGP, EGP, or incomplete) are the same.
- The MED attribute values are the same.
- All the routes are EBGP or IBGP routes.
- The IGP metric values within an AS are the same.
- AS_Path attribute values are the same.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuring BGP Load Balancing



In the figure, if no routing policy or configuration is performed for the two BGP routes on R1, the first eight rules cannot determine the optimal route. Therefore, you can configure load balancing among IBGP routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Verifying the Configuration of Load Balancing Among BGP Routes

```
[R1]display ip routing-table 10.0.45.0 24
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Table : Public
Summary Count : 2
Destination/Mask    Proto   Pre  Cost      Flags NextHop         Interface

10.0.45.0/24        IBGP    255  0           RD  10.0.2.2        GigabitEthernet0/0/0
                    IBGP    255  0           RD  10.0.3.3        GigabitEthernet0/0/1
```

The equal-cost routes to 10.0.45.0/24 exist in the IP routing table.

```
[R1]display bgp routing-table

 BGP Local router ID is 10.0.1.1
 Status codes: * - valid, > - best, d - damped,
               h - history,  i - internal, s - suppressed, S - Stale
               Origin : i - IGP, e - EGP, ? - incomplete
 Total Number of Routes: 2
      Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

 *>i  10.0.45.0/24       10.0.2.2         0          100          0    45?
 * i                     10.0.3.3         0          100          0    45?
```

There is only one optimal route in the BGP routing table.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

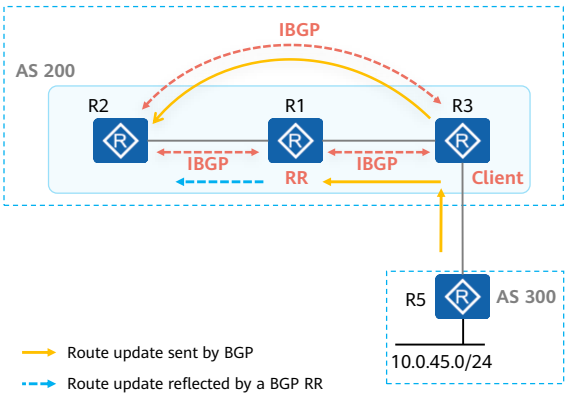
Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the network command, route learned by using the import-route command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. **Prefers the route with the shortest Cluster_List length.**
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. Prefers the route learned from the peer with the smallest IP address.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Preferring the Route with the Shortest Cluster_List Length (1)



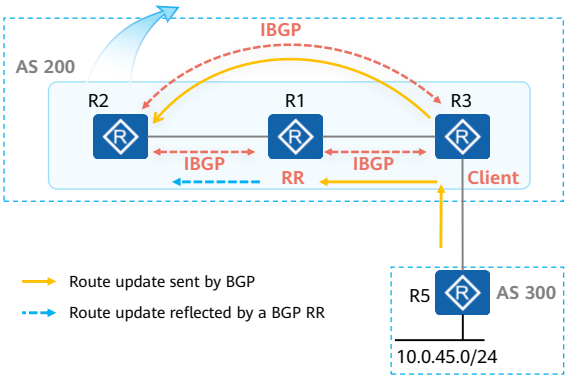
- The following configuration is performed:
  - Configure only R5 to advertise the route 10.0.45.0/24 to BGP.
  - Configure R1 as the RR and R3 as the client of R1.
  - Establish an IBGP peer relationship between R2 and R3 based on loopback interfaces.
- R2 receives the BGP route 10.0.45.0/24 advertised by R3 and the BGP route 10.0.45.0/24 reflected by R1.
- By default, the preceding rules cannot determine the optimal route. In this case, the route is selected based on Cluster_List.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Shortest Cluster_List Length (2)

```
Total Number of Routes: 2
     Network          NextHop    MED    LocPrf    PrefVal Path/Ogn

*>i  10.0.45.0/24     10.0.3.3     0       100          0  300?
* i                   10.0.3.3     0       100          0  300?
```



Based on the BGP routing table, it cannot be determined whether the BGP route reflected by R1 or advertised by R3 is preferred. To check detailed information about BGP routes, run the **display bgp routing 10.0.45.0 24** command.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Shortest Cluster_List Length (3)

```
BGP routing table entry information of 10.0.45.0/24:
From: 10.0.1.1 (10.0.1.1)
Route Duration: 00h03m10s
Relay IP Nexthop: 10.0.12.1
Relay IP Out-Interface: GigabitEthernet0/0/0
Original nexthop: 10.0.3.3
Qos information : 0x0
AS-path 300, origin incomplete, MED 0, localpref 100, pref-val 0, valid,
internal, pre 255, IGP cost 2, not preferred for Cluster List
Originator:  10.0.3.3
Cluster list: 10.0.1.1
Not advertised to any peer yet
```

- The route reflected by R1 is not the optimal route due to the following reason: not preferred for Cluster List
- The BGP route that R3 directly advertises to R2 does not pass through the RR and therefore does not have the Cluster_List attribute. That is, the Cluster_List attribute value of the BGP route advertised by R3 is considered as 0, which is smaller than the Cluster_List attribute value (1) of the BGP route reflected by R1. Therefore, the BGP route advertised by R3 is preferred.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

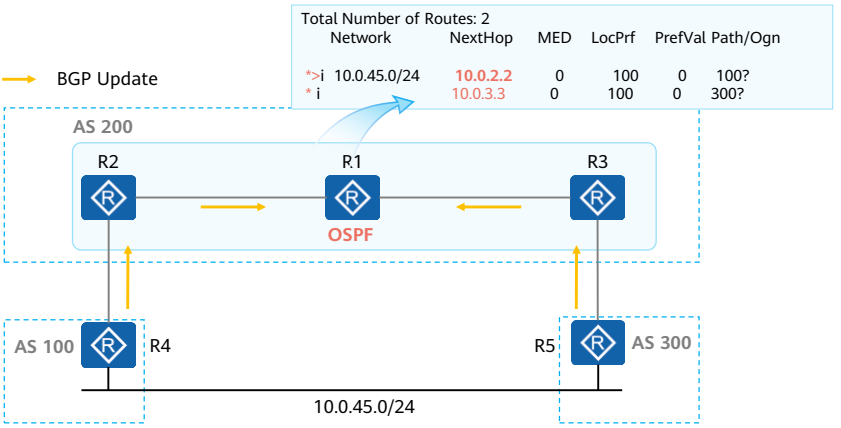
When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the network command, route learned by using the import-route command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. **Prefers the route advertised by the device with the smallest router ID (Originator_ID).**
11. Prefers the route learned from the peer with the smallest IP address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Smallest Router ID (1)



In the preceding topology, R1 receives BGP route 10.0.45.0/24 from both R2 and R3 by default, and the preceding route selection rules cannot determine the optimal route. Therefore, R1 selects the BGP route advertised by the peer with the smallest router ID based on the preceding route selection rules. In this example, the BGP route advertised by R2 is preferred.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

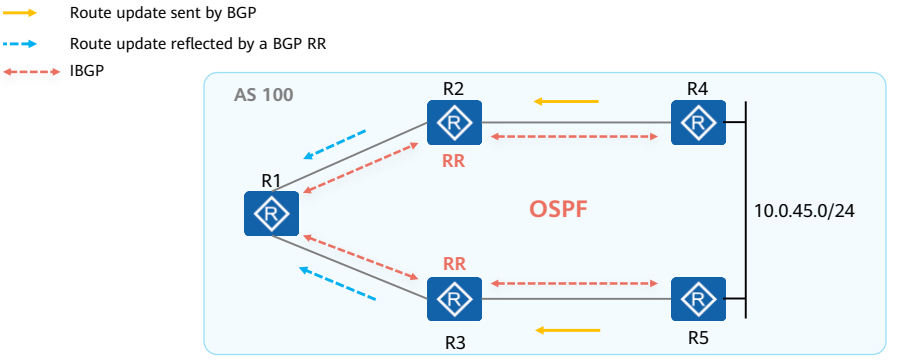
Preferring the Route with the Smallest Router ID (2)

```
BGP routing table entry information of 10.0.45.0/24:
From: 10.0.3.3 (10.0.3.3)
Route Duration: 00h40m15s
Relay IP Nexthop: 10.0.13.3
Relay IP Out-Interface: GigabitEthernet0/0/1
Original nexthop: 10.0.3.3
Qos information : 0x0
AS-path 300, origin incomplete, MED 0, localpref 100, pref-val 0, valid,
internal, pre 255, IGP cost 1, not preferred for router ID
Not advertised to any peer yet
```

Check detailed information about the BGP routing table of R1. The command output shows that the BGP route from 10.0.3.3 is not preferred because of the router ID.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Smallest Originator_ID (1)



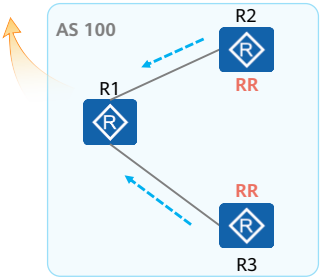
If a BGP route carries the Originator_ID attribute, the router compares the Originator_ID values of the routes and selects the BGP route with the smallest Originator_ID value.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route with the Smallest Originator_ID (2)

Total Number of Routes: 2

Network	NextHop	MED	LocPrf	PrefVal	Path/Ogn
*>i 10.0.45.0/24	10.0.4.4	0	100	0	?
* i	10.0.5.5	0	100	0	?



Route update sent by BGP

Route update reflected by a BGP RR

BGP routing table entry information of 10.0.45.0/24:
From: 10.0.3.3 (10.0.3.3)
Route Duration: 00h33m15s
Relay IP Nexthop: 10.0.13.3
Relay IP Out-Interface: GigabitEthernet0/0/1
Original nexthop: 10.0.5.5
Qos information : 0x0
AS-path Nil, origin incomplete, MED 0, localpref 100, pref-val 0, valid, internal, pre 255, IGP cost 2, **not preferred for router ID**
Originator: 10.0.5.5
Cluster list: 10.0.3.3
Not advertised to any peer yet

The BGP route reflected by R3 is not selected because of the router ID. The router ID here refers to the Originator ID (router ID of the original route advertiser).

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Rules for Selecting a Preferred BGP Route

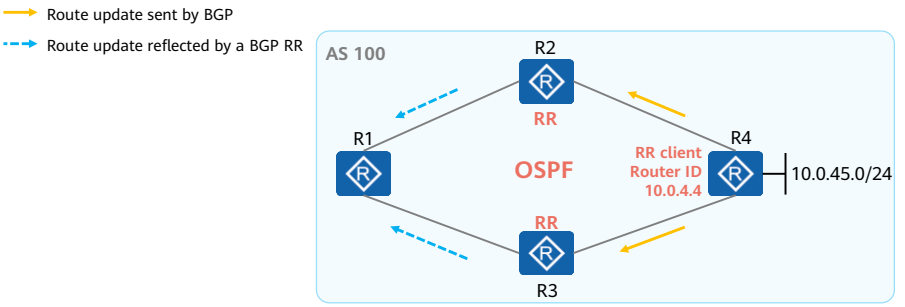
When multiple routes to the same destination network segment exist, BGP selects routes in the following sequence:

Discards the route whose next hop is unreachable.

1. Prefers the route with the largest Preferred-Value attribute value.
2. Prefers the route with the largest Local_Preference value.
3. Prefers the locally originated BGP route, which takes precedence over the route learned from a peer. The locally summarized route, automatically summarized route, route learned by using the network command, route learned by using the import-route command, and route learned from a peer are in descending order of priority.
4. Prefers the route with the shortest AS_Path.
5. Prefers the route with the shortest AS_Path. The routes with Origin attributes of IGP, EGP, and Incomplete are in descending order of priority.
6. Prefers the route with the lowest MED.
7. Prefers routes learned from EBGP peers to routes learned from IBGP peers.
8. Prefers the route with the smallest IGP metric to the next hop.
9. Prefers the route with the shortest Cluster_List length.
10. Prefers the route advertised by the device with the smallest router ID (Originator_ID).
11. **Prefers the route learned from the peer with the smallest IP address.**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route from the Device with the Smallest IP Address (1)



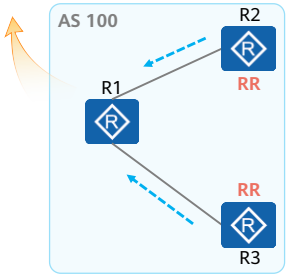
- If the preceding rules cannot determine the optimal route, the route from the device with the smallest IP address is preferred.
- In the preceding topology, R2 and R3 are connected to R4. R4 functions as the RR client and advertises routes to BGP only on R4. In this case, the BGP routes reflected by R2 and R3 have the same Originator ID 10.0.4.4.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preferring the Route from the Device with the Smallest IP Address (2)

```
Total Number of Routes: 2
     Network            NextHop        MED        LocPrf    PrefVal Path/Ogn

*>i  10.0.45.0/24       10.0.4.4       0          100       0       ?
 * i                    10.0.4.4       0          100       0       ?
```



Route update sent by BGP

Route update reflected by a BGP RR

```
BGP routing table entry information of 10.0.45.0/24:
From: 10.0.3.3 (10.0.3.3)
Route Duration: 00h01m07s
Relay IP Nexthop: 10.0.12.2
Relay IP Out-Interface: GigabitEthernet0/0/0
Original nexthop: 10.0.4.4
Qos information : 0x0
AS-path Nil, origin incomplete, MED 0, localpref 100, pref-val 0, valid, intern
al, pre 255, IGP cost 2, not preferred for peer address
Originator:  10.0.4.4
Cluster list: 10.0.3.3
Not advertised to any peer yet
```

The BGP route reflected by R3 is not selected because the peer address is larger. The peer address of the route reflected by R2 is 10.0.2.2, and the peer address of the route reflected by R3 is 10.0.3.3. Therefore, the BGP route reflected by R3 is not selected.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Essay) When a BGP route received from an EBGP peer is advertised to an IBGP peer, how does the EBGP peer change the value of the Next_Hop attribute to its own source address?
2. (TorF) If the preceding three rules are the same, BGP compares the AS_Path attribute values. If the AS_Path attribute values are the same, BGP compares the AS numbers.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. The **peer next-hop-local** command is executed to set the Next-Hop attribute as the source address for peer relationship setup.
2. False

Summary

- BGP selects optimal routes based on path attributes. This allows BGP to select optimal routes based on path attributes in different scenarios.
- BGP defines a set of detailed optimal path selection algorithms, which enable routers to select the optimal path in any complex and highly redundant network environment.
- BGP route selection rules are frequently used in practice and need to be mastered.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



BGP EVPN Basics

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- Standard BGP-4 supports only IPv4 unicast addresses. To support more network layer protocols, Multiprotocol Extensions for BGP-4 (MP-BGP) (RFC 4760) was proposed as an extension to BGP-4 to allow different types of address families to be distributed in BGP at the same time. The address families include IPv4 multicast, IPv6, L3VPN, and Ethernet Virtual Private Network (EVPN) address families.
- With the development and commercial use of software-defined networking (SDN), EVPN plays an important role in various solutions, covering all scenarios, including campus networks, data centers, IP WAN transport networks, and software-defined networking in a wide area network (SD-WAN).
- This course describes the concept of MP-BGP, development history of EVPN, common EVPN route types, and EVPN usage scenarios.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Understand basic MP-BGP concepts.
  - Understand the origin of EVPN.
  - Understand common EVPN route types.
  - Understand typical EVPN usage scenarios.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

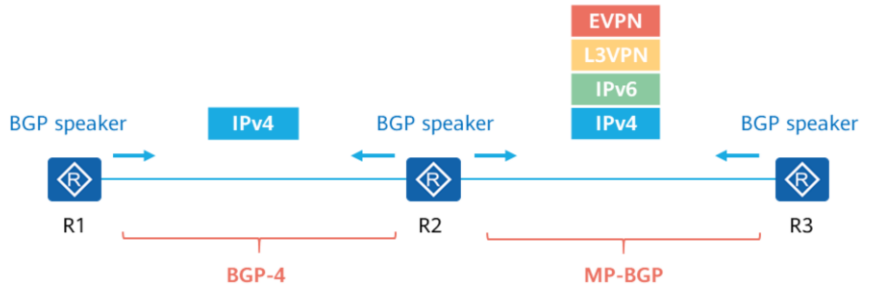
Contents

1. MP-BGP
2. EVPN

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MP-BGP

As defined in RFC 4760, MP-BGP is used to extend BGP-4 to allow BGP to carry multiple network layer protocols, such as IPv6, L3VPN, and EVPN. This extension has good backward compatibility. That is, an MP-BGP-capable router can interact with a router that supports only BGP-4.

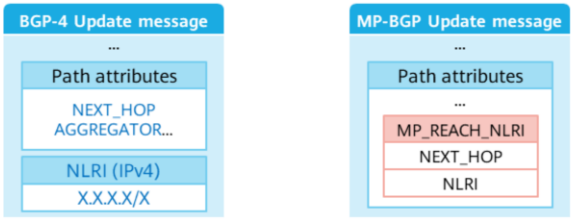


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- https://datatracker.ietf.org/doc/rfc4760/

BGP-4 Extensions

- BGP-4 has three IPv4-specific pieces of information: NEXT_HOP, AGGREGATOR, and IPv4 network layer reachable information (NLRI). To support multiple network layer protocols, BGP-4 has to provide the following abilities:
  - Ability of associating network layer protocols with next-hop information
  - Ability of associating network layer protocols with NLRIs
- The two abilities are collectively referred to as the address family (AF) defined by the Internet digital distribution agency (IANA).
- To implement forward compatibility, MP-BGP adds two new attributes: MP_REACH_NLRI and MP_UNREACH_NLRI, which are used to indicate reachable and unreachable destinations, respectively. The two attributes are optional non-transitive.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- According to BGP-4, NEXT_HOP and AGGREGATOR fields are contained in Path attributes of IPv4, and the IPv4 NLRI carries IPv4 routing entries.
- The Path attributes field is added in MP-BGP. MP_REACH_NLRI is a new field of path attributes. The NEXT_HOP and NLRI fields of the corresponding network layer protocol and the NLRI belong to MP_REACH_NLRI.

MP_REACH_NLRI

- MP_REACH_NLRI is carried in a BGP Update message and provides the following functions:
  - Advertises reachable routes to BGP peers.
  - Advertises the next-hop address of a reachable route to a BGP peer.
- It contains the following fields:

MP_REACH_NLRI Format	Field Description
Address Family Identifier (2 octets)	Network layer protocol. Value 2 indicates IPv6.
Subsequent Address Family Identifier (1 octet)	This field is used together with address family identifier (AFI). Value 1 indicates unicast, and value 2 indicates IPv6 unicast.
Length of Next Hop Network Address (1 octet)	Length of a next-hop IP address.
Network Address of Next Hop (variable)	Next-hop address. The format is determined by the AFI and SAFI.
Reserved (1 octet)	All 0s.
Network Layer Reachability Information (variable)	The length of this field is variable. This field can contain reachable routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In the SAFI field, value 1 indicates unicast, and value 2 indicates multicast. The value is allocated by the IANA. The allocation rules are defined in RFC 2434 (titled "Guidelines for Writing an IANA Considerations Section in RFCs").
- In this section, the AFI of EVPN is 25 (L2VPN) and the SAFI is 70 (EVPN).

MP_UNREACH_NLRI

- MP_UNREACH_NLRI is carried in BGP Update messages to withdraw unreachable routes.
- It contains the following fields:

MP_UNREACH_NLRI Format	Field Description
Address Family Identifier (2 octets)	Network layer protocol. Value 2 indicates IPv6.
Subsequent Address Family Identifier (1 octet)	This field is used together with the AFI. Value 1 indicates unicast, and value 2 indicates IPv6 unicast.
Withdrawn Routes (variable)	The length of this field is variable. This field lists the routes that need to be withdrawn. The format of this field is determined by the AF and SAFI.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The AFI of EVPN is 25 (L2VPN) and the subsequent address family identifier (SAFI) is 70 (EVPN).

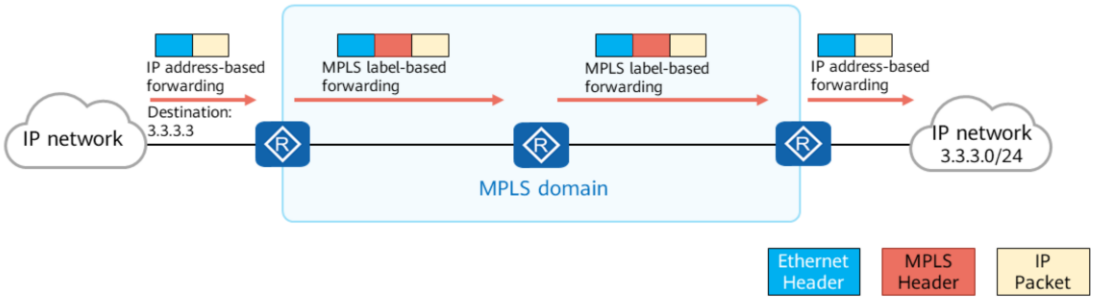
Contents

1. MP-BGP
2. **EVPN**
   - **EVPN Overview**
   - Common EVPN Routes
   - Typical EVPN Usage Scenarios

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MPLS Overview

- Multiprotocol Label Switching (MPLS) is located between the data link layer and the network layer in the TCP/IP protocol stack. An MPLS header is added between the two layers. Packets are forwarded based on the MPLS header. The MPLS header is also called the MPLS label.
- MPLS replaces IP forwarding with label switching to implement label-based rapid forwarding.

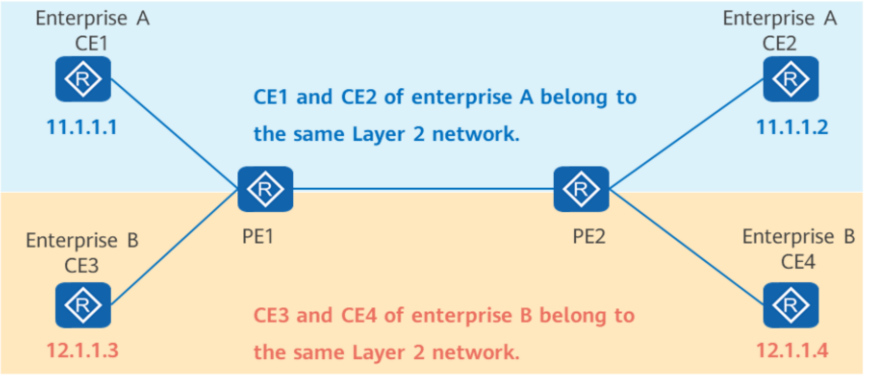


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- MPLS originates from IPv4 and its core technologies can be extended to multiple network protocols, including IPv6, Internet Packet Exchange (IPX), Appletalk, DECnet and Connectionless Network Protocol (CLNP). "Multiprotocol" in MPLS indicates that multiple network protocols are supported.
- MPLS replaces IP forwarding with label switching. A label is a short and fixed-length connection identifier that has only local significance. It is similar to the virtual path identifier (VPI)/virtual channel identifier (VCI) of Asynchronous Transfer Mode (ATM) and the data link connection identifier (DLCI) of Frame Relay.
- MPLS domain: An MPLS domain consists of a series of consecutive network devices that run MPLS.

VPLS Overview

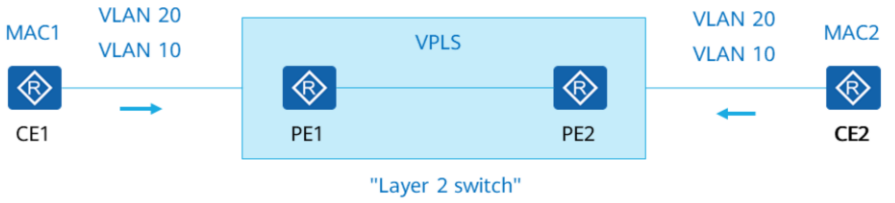
Virtual private LAN service (VPLS) is an Ethernet-based L2VPN technology. VPLS provides services similar to LAN services on an MPLS network and allows users to access the network from different locations.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Traditional L2VPN

- Traditional L2VPN services, such as VPLS, provide Layer 2 connections between remote sites. An L2VPN network is built and functions like a Layer 2 switch to transparently transmit Ethernet packets. In this example, PE1 and PE2 form a VPLS network to transparently transmit VLAN traffic between CE1 and CE2.
- In a traditional L2VPN, remote MAC addresses are learned through ARP broadcast flooding, and therefore, PEs need to carry broadcast traffic. Broadcast consumes a large amount of interface bandwidth, which is a typical issue of traditional L2VPN.

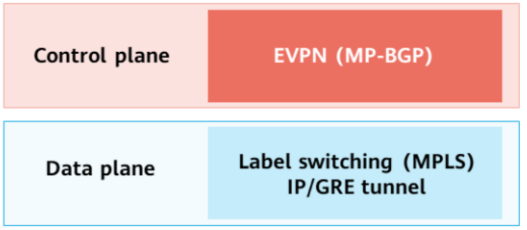


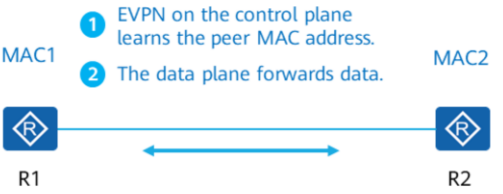
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- VPLS does not support all-active access or load balancing and implements slow fault convergence. For details, see materials of the HCIE-HCIE-Datacom Ethernet VPN and RFC 7209 titled "Requirements for Ethernet VPN (EVPN)."

Emergence of EVPN

- With new technologies and scenarios emerging, VPLS cannot meet the requirements of L2VPN. The industry has reviewed the requirements for Ethernet VPN (RFC 7209) and proposed a new solution, that is, EVPN.
- EVPN was first defined in RFC 7432. EVPN introduces the control plane to better control MAC address learning.
- EVPN uses MP-BGP on the control plane and supports MPLS label switched paths (LSPs) or IP/Generic Routing Encapsulation (GRE) tunneling on the data plane.



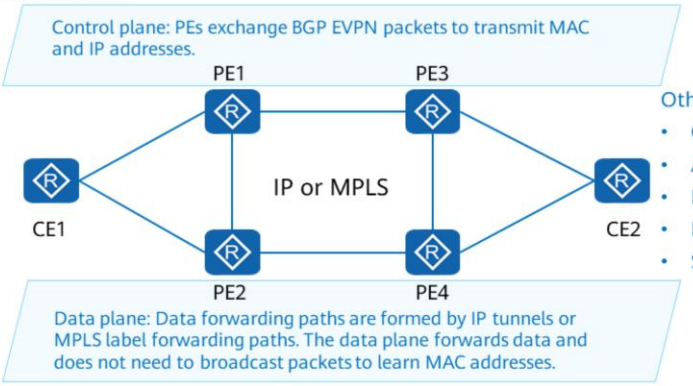


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- https://datatracker.ietf.org/doc/rfc7209/
- https://datatracker.ietf.org/doc/rfc7432/

Advantages of EVPN

- EVPN introduces the control plane to learn MAC and IP addresses to guide data forwarding, implementing forwarding-control separation.
- EVPN resolves typical problems in traditional L2VPNs and offers more benefits, such as active-active, rapid convergence, and simplified O&M.



Other advantages of EVPN:

- CEs can access PEs in all-active mode.
- Automatic discovery of PE members
- Loop prevention
- Broadcast traffic optimization
- Support for Equal-Cost Multi-Path (ECMP)

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- For more details, see the *HCIE-Datacom Ethernet VPN*.

Contents

1. MP-BGP
2. **EVPN**
   - EVPN Overview
   - **Common EVPN Routes**
   - Typical EVPN Usage Scenarios

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

EVPN NLRI

- EVPN defines a new type of BGP NLRI, known as EVPN NLRI, to carry all EVPN routes.
- EVPN NLRI is a new extension to MP-BGP. It is included in MP_REACH_NLRI. For the EVPN NLRI, the AFI is 25 and the SAFI is 70.

MP_REACH_NLRI format	MP_REACH_NLRI of BGP EVPN
Address Family Identifier (2 octets)	AFI: 25
Subsequent Address Family Identifier (1 octet)	SAFI: 70
Length of Next Hop Network Address (1 octet)	Length of a next-hop IP address.
Network Address of Next Hop (variable)	Next-hop IP address in an EVPN route.
Reserved (1 octet)	All 0s.
Network Layer Reachability Information (variable)	EVPN NLRI

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

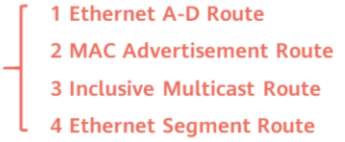
EVPN Route

The EVPN NLRI format uses the Type-Length-Value (TLV) structure, making packets highly flexible and scalable.

- Route Type field: defines different EVPN routes. RFC 7432 defines four types of routes.
- Length field: defines the length of a field.
- Route Type Specific field: contains fields of a particular route type.



EVPN NLRI format



Four common EVPN routes defined in the Route Type field

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The NLRI field in the MP_REACH_NLRI/MP_UNREACH_NLRI attribute contains the EVPN NLRI (encoded as specified above).
- The EVPN NLRI is carried in BGP [RFC4271] using BGP Multiprotocol Extensions [RFC4760] with an Address Family Identifier (AFI) of 25 (L2VPN) and a Subsequent Address Family Identifier (SAFI) of 70 (EVPN). The NLRI field in the MP_REACH_NLRI/MP_UNREACH_NLRI attribute contains the EVPN NLRI (encoded as specified above).
- In order for two BGP speakers to exchange labeled EVPN NLRI, they must use BGP Capabilities Advertisements to ensure that they both are capable of properly processing such NLRI. This is done as specified in [RFC4760], by using capability code 1 (multiprotocol BGP) with an AFI of 25 (L2VPN) and a SAFI of 70 (EVPN).

More Types of EVPN Routes and Their Functions

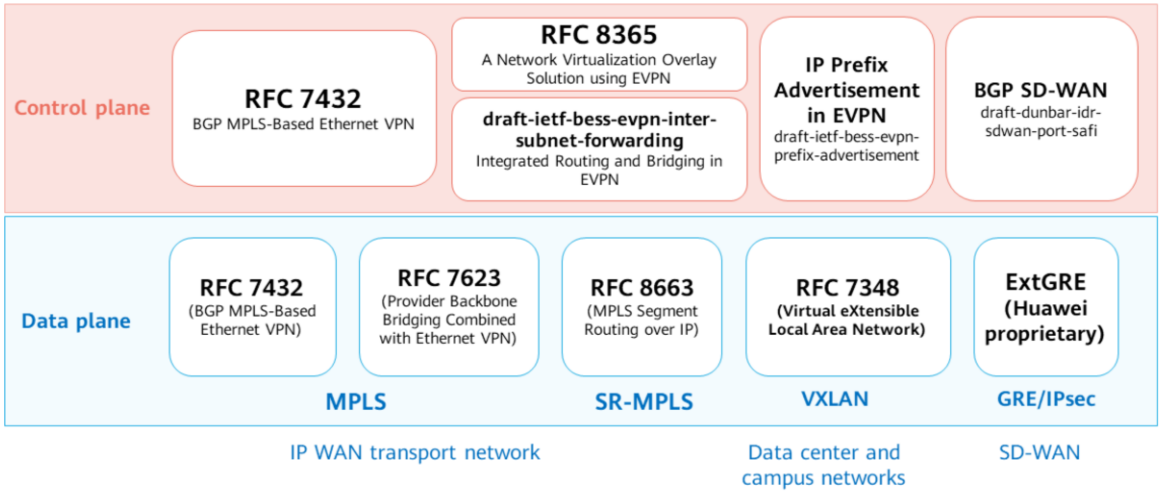
EVPN is not limited to L2VPN applications. With the increase of EVPN route types, more applications, such as L3VPN, are supported.

Type of Route	Function	RFC
(Type 1) Ethernet A-D Route	• Aliasing • MAC address batch withdraw • All-active flag • ESI label advertisement	RFC 7432
(Type 2) MAC/IP Advertisement Route	• MAC address learning notification • MAC/IP binding • MAC mobility	
(Type 3) Inclusive Multicast Route	Automatic discovery of multicast tunnel endpoints and multicast types	
(Type 4) Ethernet Segment Route	Automatic discovery of ES members DF election	
(Type 5) IP Prefix Route	IP prefix advertisement (support for L3VPN)	draft-ietf-bess-evpn-prefix-advertisement

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The Type 5 route (IP prefix route) related standard is in the draft phase, in draft-ietf-bess-evpn-prefix-advertisement.

EVPN Protocol Standards

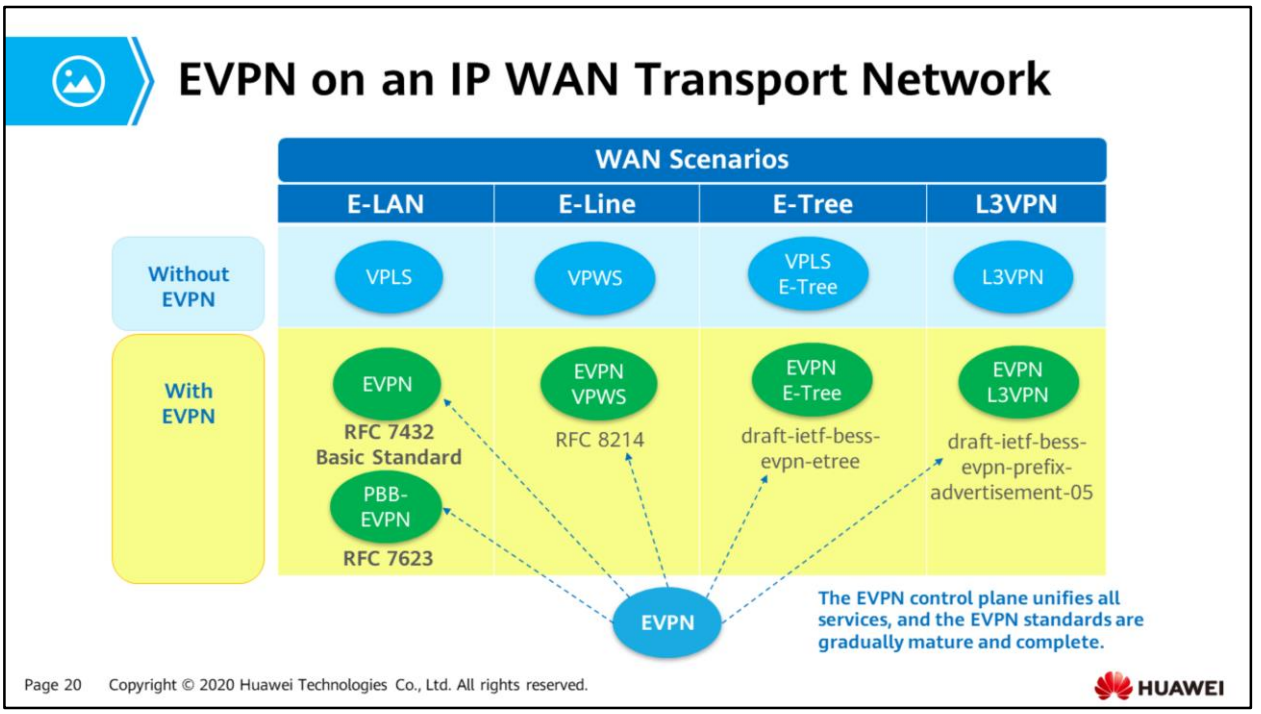


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Contents

1. MP-BGP
2. **EVPN**
   - EVPN Overview
   - Common EVPN Routes
   - **Typical EVPN Usage Scenarios**

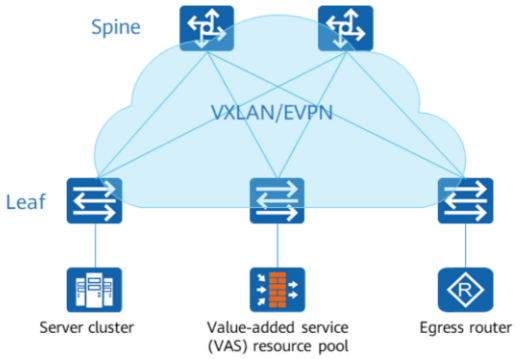
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- E-Line, E-Tree, and E-LAN are three types of Ethernet virtual circuits (EVCs). For details, see metro Ethernet standards at https://wiki.mef.net/display/CESG/E-Line.
- The Metropolitan Ethernet Forum (MEF) defines three types of EVCs: point-to-point EVC, multipoint-to-multipoint EVC, and root-multipoint EVC.
    - E-Line: A point-to-point EVC strictly associates two User-to-Network Interfaces (UNIs).
    - E-LAN: A multipoint-to-multipoint EVC can associate two or more UNIs. Users or carriers can add any UNIs to the EVC or delete some UNIs from the EVC without affecting other UNIs.
    - E-Tree: This EVC is similar to the hub-spoke model in L3VPN. It consists of one or more root UNIs and several leaf UNIs. The root UNI can directly communicate with all UNIs in the EVC, whereas a leaf UNI can only communicate directly with the root UNI in the EVC, and two leaf UNIs cannot communicate with each other directly.

EVPN on a Data Center Network

- The network virtualization overlay (NVO) solution (RFC 8365) of EVPN is used in cloud data centers.
- It is recommended that the data plane use Virtual Extensible LAN (VXLAN) encapsulation and the control plane use EVPN to construct a flexible data center overlay network.

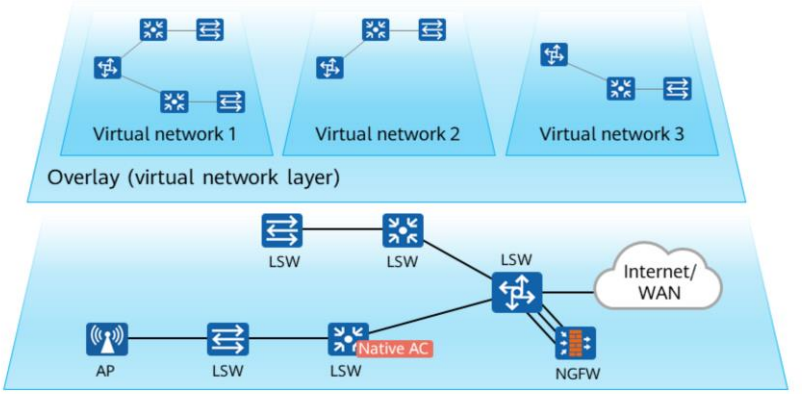


- All services in the data center are carried by the VXLAN overlay network.
- The underlay network consisting of spine and leaf nodes performs high-speed forwarding.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

EVPN Application on a Campus Network

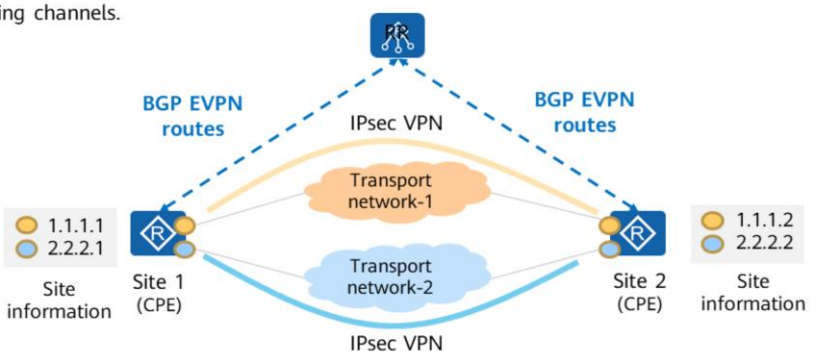
- The campus network virtualization solution is similar to that in the cloud data center. The EVPN NVO solution (RFC 8365) is used.
- VXLAN encapsulation and EVPN are used on different underlying networks to build a flexible overlay network.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

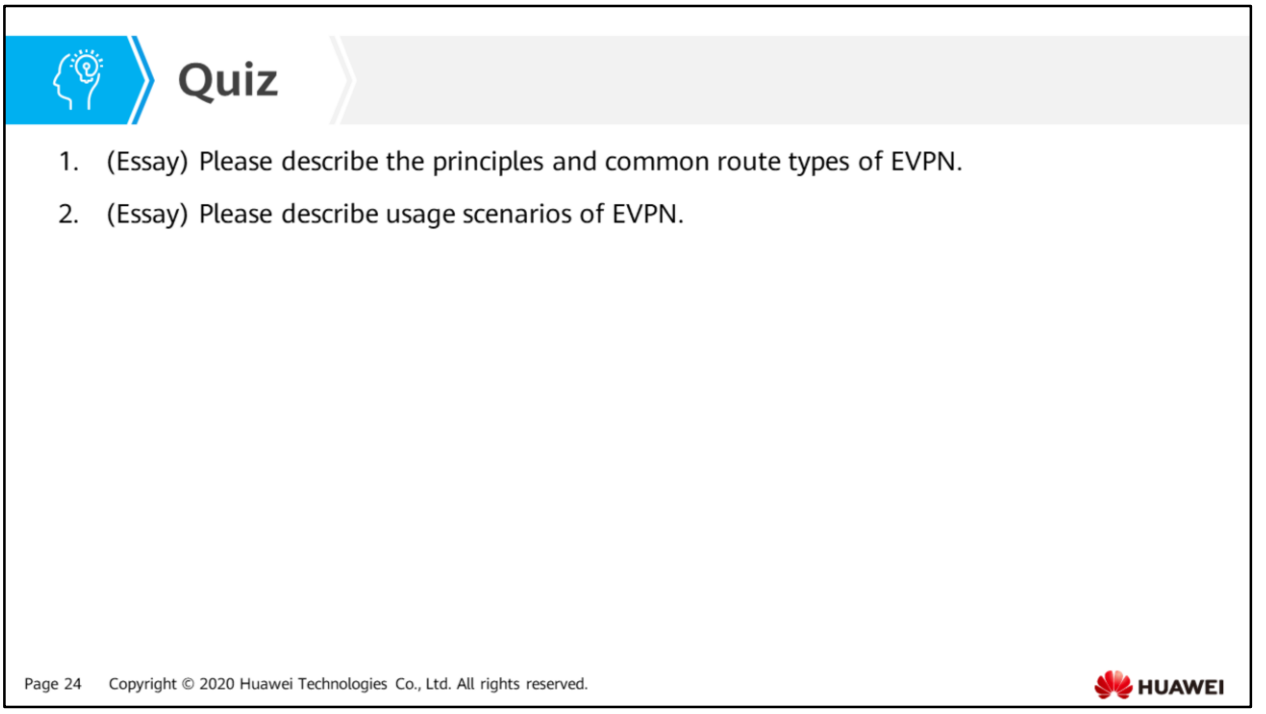
EVPN Application in SD-WAN

- SD-WAN is a next-generation enterprise branch interconnection solution that supports intelligent dynamic traffic steering, Zero Touch Provisioning (ZTP), and visualization.
- In the SD-WAN solution, EVPN is deployed between route reflectors (RRs) and customer-premises equipment (CPE) devices to advertise SD-WAN overlay VPN routes on the control plane. IPsec VPN is used on the data plane to build secure forwarding channels.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Overlay VPN routes include site VPN route prefixes, next-hop route information, and IPsec key pairs required for data encryption of data channels between CPEs. For details, see materials of the SD-WAN course.
- CPE



1. EVPN is an extension to MP-BGP. EVPN provides five major types of routes and is used as the control plane of Layer 2 or Layer 3 tunnels.
2. EVPN can be widely used in all enterprise scenarios, such as SD-WAN, campus networks, data centers, and WANs. In data centers and campus networks, EVPN and VXLAN are used together to construct a service overlay network. In SD-WAN scenarios, EVPN and IPsec are used together to build enterprise branch interconnection networks. On a WAN, EVPN can be used with various underlying tunneling and label technologies, such as MPLS, Segment Routing (SR), VPLS, and virtual private wire service (VPWS).

Summary

- MP-BGP's extension to BGP-4 allows different types of address families, such as IPv4 multicast, IPv6, L3VPN, and EVPN, to be distributed in BGP.
- This course describes EVPN that is used to solve the Ethernet L2VPN problems. With the increase of usage scenarios and protocol extensions, EVPN can be used in various scenarios, including WANs, data centers, campus networks, and SD-WANs.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

More Information

- https://datatracker.ietf.org/doc/rfc4760/
- https://datatracker.ietf.org/doc/rfc7209/
- https://datatracker.ietf.org/doc/rfc7432/
- https://datatracker.ietf.org/doc/rfc8365/
- https://wiki.mef.net/display/CESG/MEF+6.3+-+EVC+Ethernet+Services+Definitions

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Routing Policy and Routing Control

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- On a complex data communication network, the routing policy can be configured to filter routes and set route attributes as needed. Route control affects data traffic forwarding.
- A routing policy is not a single technique or protocol, but a special technical topic or methodology that involves multiple tools and methods.
- This course describes route selection tools and principles, as well as how to configure route-policies.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Use an ACL to filter routes.
  - Use an IP prefix to filter routes.
  - Use a filter-policy to filter routes.
  - Use a route-policy to filter routes and modify route attributes.

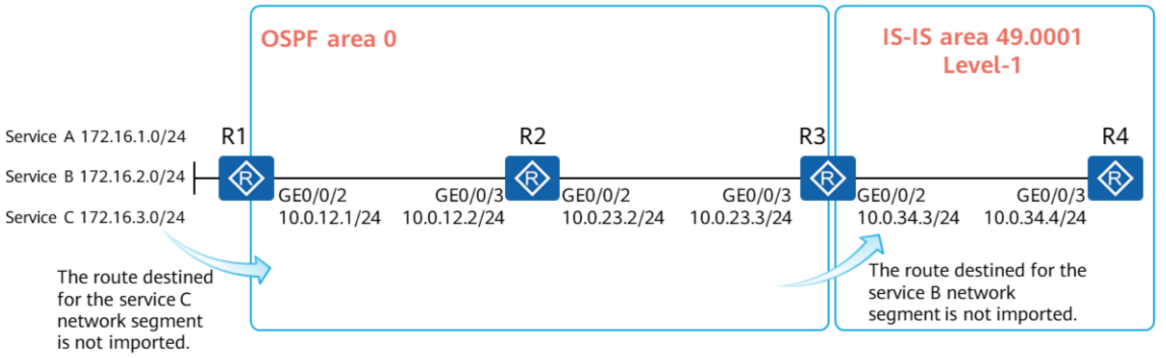
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Routing Control Overview**
2. Route Control Tool
   - Route Matching Tool
   - Route-Policy
3. Route Control Cases

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Technical Background



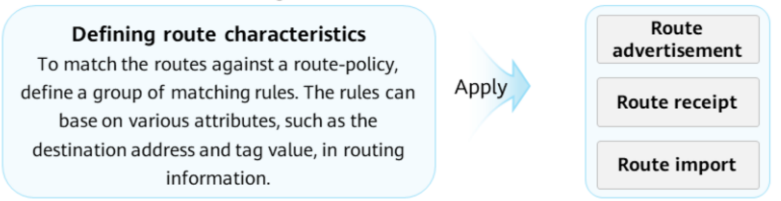
- R1 imports the network segment routes of services A and B to R1, but does not import the route destined for the service C network segment. When R3 imports OSPF routes to the IS-IS routing table, R3 needs to import only the route destined for the service B network segment.
- In this case, a tool is required to control routes to be imported.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Routing Control Overview

Route control can be implemented using a route-policy. The route-policy is flexibly used in the following scenarios:

- Controlling route advertisement: A route-policy is used to filter the routes to be advertised so that a device advertises only the routes that match matching conditions.
- Controlling route receipt: A route-policy is used to filter the routes to be accepted so that a device accepts only the routes that match matching conditions.
- Controlling route import: A route-policy is used to filter the routes to be imported so that a device imports only the routes that match matching conditions.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

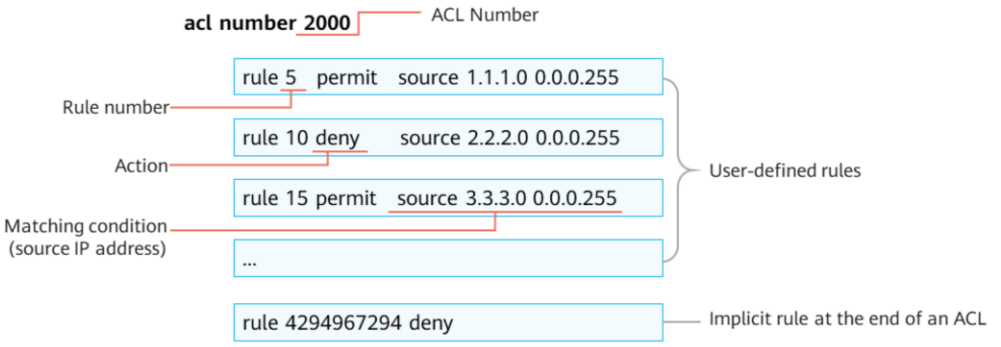
Contents

1. Routing Control Overview
2. **Route Control Tool**
   - **Route Matching Tool**
   - Route-Policy
3. Route Control Cases

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

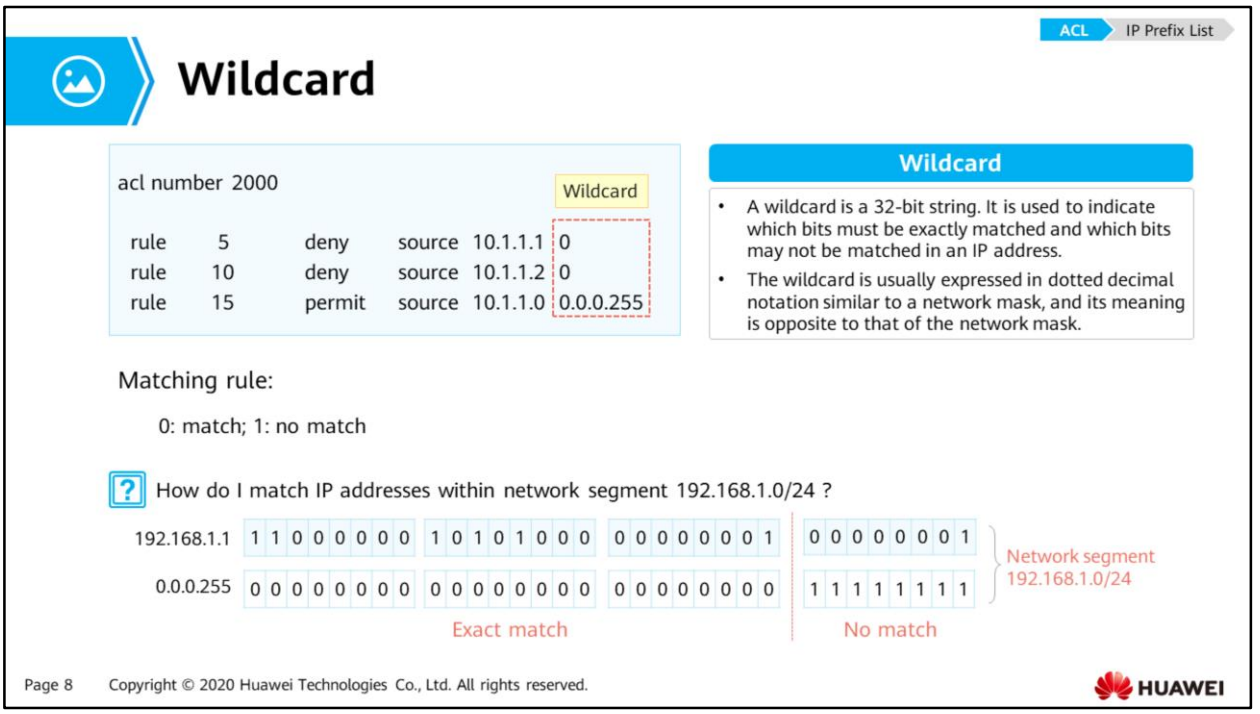
Matching Tool 1: ACL

- An access control list (ACL) is a matching tool that can match and distinguish packets and routes.
- An ACL consists of multiple permit, deny, or both clauses. Each statement is a rule of the ACL. The permit or deny action in each statement is an action bound to the rule.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- An ACL consists of the following elements:
  - ACL number: Each ACL configured on a device is assigned a number, which is called an ACL number and is used to identify the ACL. The ACL number range varies according to the ACL type.
  - Rule: As mentioned above, an ACL is usually consists of multiple permit, deny, or both clauses, and each clause is a rule of the ACL.
  - Rule number: Each rule has a rule number, which identifies an ACL rule. The value can be user-defined or automatically allocated by the system. The number of an ACL rule is an integer ranging from 0 to 4294967294. All ACL rules are numbered in ascending order.
  - Action: "Permit" or "deny" in each rule is an action bound to a rule. ACLs are usually used together with other technologies. The meanings of actions vary according to the scenario.
    - For example, if an ACL is used together with a traffic filtering technology (the ACL is applied to the traffic filtering function), "permit" indicates that traffic is allowed to pass, and "deny" indicates that traffic is rejected.
  - Item to be matched against: The ACL defines abundant items to be matched against. In this example, the source IP address is used. The ACL also supports many other items. For instance, the items can be Layer 2 Ethernet frame header information (such as a source MAC address, destination MAC address, and Ethernet frame protocol type), Layer 3 packet information (such as a destination address and protocol type), or Layer 4 packet information (such as a TCP/UDP port number).
- Question: What does the **rule 5 permit source 1.1.1.0 0.0.0.255** command meanh This will be introduced later.



- For an IP address to be matched against a matching rule, the address is followed by a 32-bit mask. The 32-bit mask is called a wildcard.
- The wildcard is in dotted decimal notation. After it is converted into the binary format, value 0 indicates a "match" and value 1 indicates "no match". 1s or 0s in the wildcard may be discontinuous.
- There are two examples:
  - rule 5: rejects packets with source IP address 10.1.1.1. The all-0 wildcard indicates that each bit must be exactly matched. Therefore, the host IP address 10.1.1.1 matches the rule.
  - rule 15: permits packets whose source IP addresses belong to network segment 10.1.1.0/24. The wildcard is 0.0.0.111111111, and the right-most 8 bits are 1, which indicates that these bits in packets can be ignored. As such, the right-most 8 bits in 10.1.1.xxxxxxxx can be any value, and the network segment 10.1.1.0/24 matches this rule.
- Example: To exactly match the network segment address of 192.168.1.1/24, which wildcard can be used?
  - It can be concluded that network bits must be exactly matched and host bits can be ignored. Therefore, the wildcard is 0.0.0.255.
- Two special wildcards:
  - The all-0 wildcard is used to exactly match a specific IP address.
  - When the all-1 wildcard is used to match 0.0.0.0, it indicates that all IP addresses are matched.

ACL Classification and Basic ACLs

- Classification based on the ACL rule definition mode

Category	Number Range	Rule Definition
Basic ACL	2000–2999	Defines rules based on source IP addresses, fragment information, and validity time ranges.
Advanced ACL	3000–3999	Defines rules based on source IP addresses, destination IP addresses, IP protocol types, ICMP types, TCP source/destination port numbers, UDP source/destination port numbers, and time ranges.
Layer 2 ACL	4000–4999	Defines rules based on the Ethernet frame header information, such as the source MAC address, destination MAC address, and Layer 2 protocol type.
User-defined ACL	5000–5999	Defines rules based on the packet header, offsets, string masks, and user-defined strings.
User ACL	6000–6999	Defines rules based on the source IP address or source user control list (UCL) group of IPv4 packets, destination IP address or destination UCL group, IP protocol types, ICMP types, TCP source/destination port numbers, and UDP source/destination port numbers.

- Basic ACL

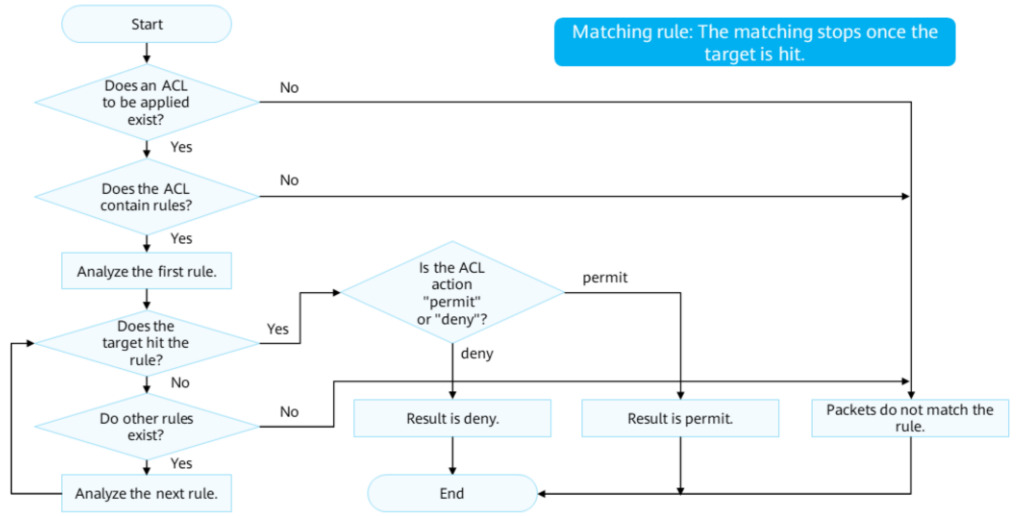



```
acl number 2000
 rule    5     deny     source  10.1.1.1  0
 rule    10    deny     source  10.1.1.2  0
 rule    15    permit   source  10.1.1.0  0.0.0.255
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Only basic ACLs can be used to match routes.

ACL Fundamentals



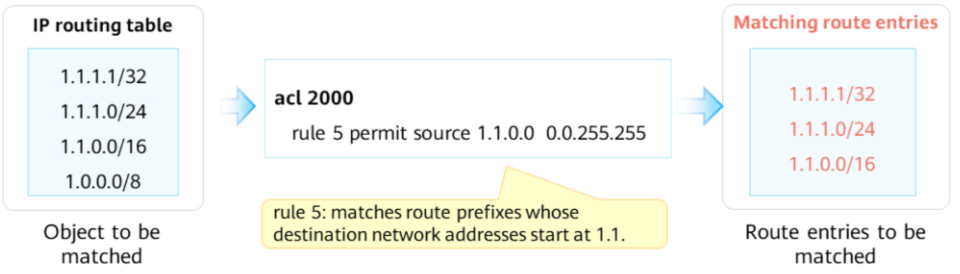
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The ACL matching mechanism is as follows:
  - After a device configured with an ACL receives a packet, the device matches the packet against ACL rules one by one. If the packet does not match an ACL rule, the device attempts to match the packet against a next rule.
  - Once the packet matches a rule, the device performs the action defined in the rule on the packet and no longer matches the packet against other rules.
- Matching process:
- The device checks whether an ACL is configured.
- If no ACL is configured, the device returns the result "negative match."
- If an ACL is configured, the device checks whether the ACL contains rules.
  - If the ACL does not contain rules, the device returns the result "negative match."
  - If the ACL contains rules, the device matches the packets against the rules in ascending order of rule IDs.
    - When the packets match a permit rule, the device stops matching and returns the result "positive match (permit)."
    - When the packets match a deny rule, the device stops matching and returns the result "positive match (deny)."
    - If the packets do not match any rule in the ACL, the device returns the result "negative match."

Matching Order and Result of ACL Rules

Configuration sequence (config mode)

- The system matches packets against ACL rules in ascending order by rule ID. A rule with a smaller ID is earlier to be matched.

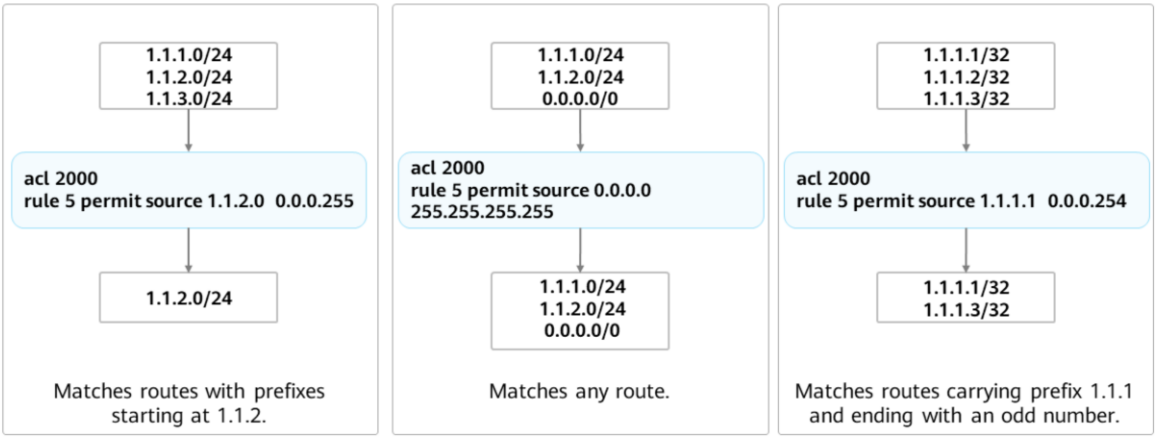


What does "permit" mean?

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- An ACL consists of multiple deny | permit clauses, each of which describes a rule. These rules may repeat or conflict. In this situation, the matching order decides the matching result.
- Huawei devices support two matching orders: automatic order (auto mode) and configured order (config mode). The configured mode is used by default.
  - Automatic order: The system arranges rules according to the precision degree of the rules (depth first principle), and matches packets against the rules in descending order of precision. A rule with the highest precision defines strictest conditions, and has the highest priority. This process is complex, so we will not go into details here. Anyone interested in this can read materials after class.
  - Configured order: The system matches packets against ACL rules in ascending order of rule IDs. That is, the rule with the smallest ID is processed first. This is the matching order we mentioned earlier.
    - If another rule is added, the rule is added to a corresponding position, and packets are still matched against the rules in ascending order by rule ID.
- Note: ACLs are always used together with other technologies. The actual functions of "permit" and "deny" vary with technologies. For example, when an ACL is used together with route filtering, "permit" means that a route is a match, and "deny" means that a route is not a match.

Common Matching Examples



An ACL can match only route prefixes, but not network masks.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Configuration Commands of Basic ACLs

1. Create a basic ACL.

   [Huawei] **acl** [**number**] *acl-number* [**match-order config**]

   Create a numbered basic ACL (2000–2999) and enter the basic ACL view.

   [Huawei] **acl name** *acl-name* { **basic** | *acl-number* } [**match-order config**]

   Create a named basic ACL and enter the basic ACL view.

2. Configure an ACL rule.

   [Huawei-acl-basic-2000] **rule** [*rule-id*] { **deny** | **permit** } [**source** { *source-address source-wildcard* | **any** } | **time-range** *time-name*]

   You can run this command in the basic ACL view to configure rules for the basic ACL.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- Create a basic ACL.
- [Huawei] **acl** [**number**] *acl-number* [**match-order config**]
  - *acl-number*: specifies the number of an ACL.
  - **match-order config**: indicates the matching order of ACL rules. config indicates the configuration order.
- [Huawei] **acl name** *acl-name* { **basic** | *acl-number* } [**match-order config**]
  - *acl-name*: specifies the name of an ACL.
  - **basic**: indicates a basic ACL.
- Configure a basic ACL rule.
- [Huawei-acl-basic-2000] **rule** [*rule-id*] { **deny | permit** } [**source** { *source-address source-wildcard* | **any** } | *time-range time-name*]
  - *rule-id*: specifies the ID of an ACL rule.
  - **deny**: rejects packets that meet the matching conditions.
  - **permit**: permits the packets that meet the matching conditions.

Matching Tool 2: IP Prefix List

- An IP prefix list is a filter that uses the network address and mask length of routes as matching rules. An IP prefix list can be used when a routing protocol advertises or receives routes.
- Different from an ACL, an IP prefix list can match both the IP address prefix length and mask length, which enhances matching accuracy.

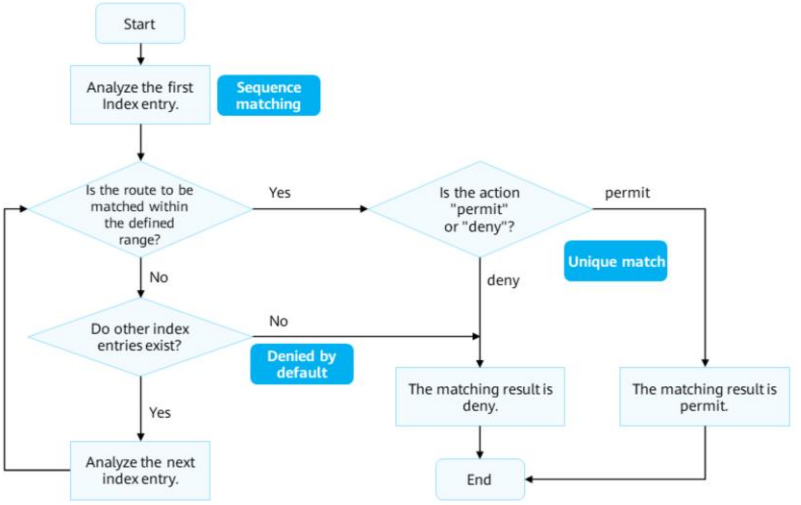
```
[Huawei] ip ip-prefix test index 10 permit 192.168.1.0 22 greater-equal 24 less-equal 26
```

ip-prefix-name | No. | Action | IP network segment and mask | Mask range

1. ***ip-prefix-name***: specifies the name of an IP prefix list.
2. **Sequence number**: indicates the sequence number of a matching entry in the IP prefix list. The IP prefixes are matched in ascending order by sequence number.
3. **Action**: permit or deny, indicating match or mismatch.
4. **IP network segment and mask**: match the network address of a route and exactly match the leftmost N bits of the network address.
5. **Mask length range**: *mask-length* <= *greater-equal-value* <= *less-equal-value* <= 32

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

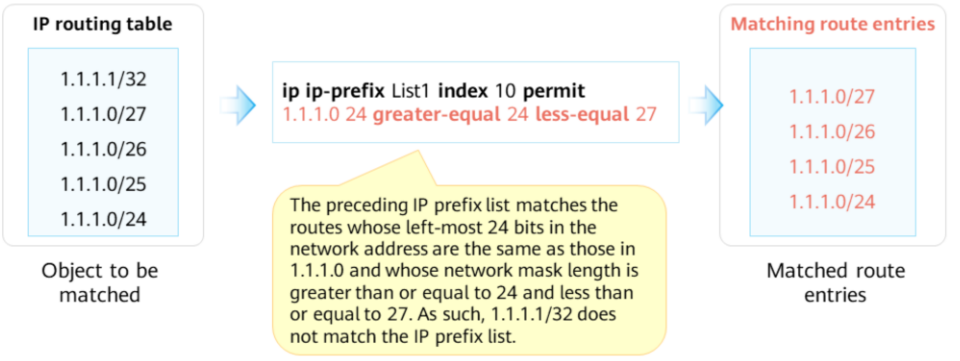
IP-Prefix Fundamentals



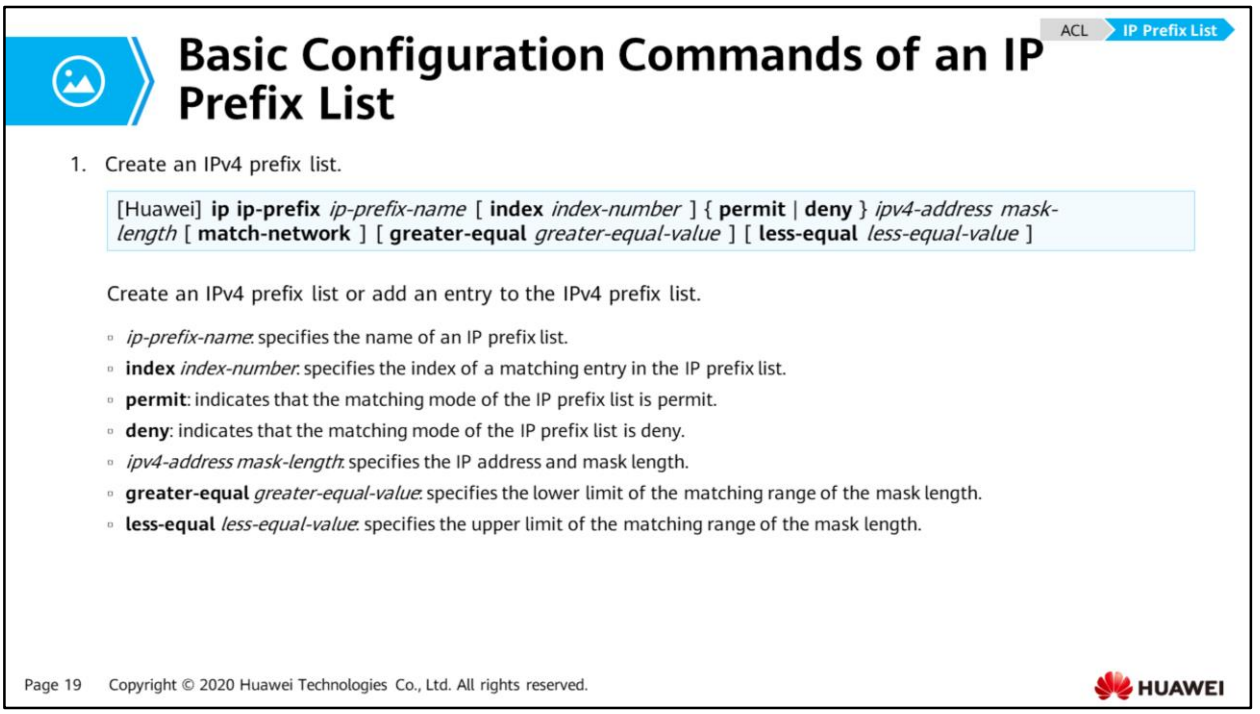
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Example of IP-Prefix Matching

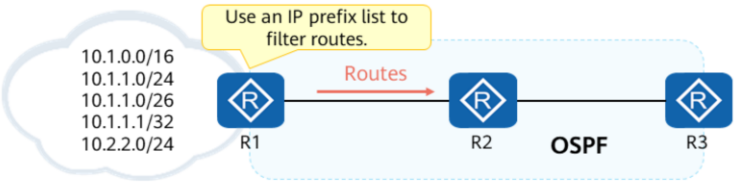


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- *ip-prefix-name*: specifies the name of an IP prefix list. The value is a string of 1 to 169 case-sensitive characters, spaces not supported.
- **index** *index-number*: specifies the index of a matching entry in the IP prefix list. The value is an integer ranging from 0 to 4294967295. By default, the sequence number increases by 10 each time an entry is added and is automatically indexed. If the system automatically assigns an index, the index starts at 10.
- **permit**: indicates that the matching mode of the IP prefix list is permit. In this mode, if the IP address to be filtered is within the defined range, the IP address passes the filtering. If no match is found, the system moves to a next node.
- **deny**: indicates that the matching mode of the IP prefix list is deny. In this mode, if the IP address to be filtered is within the defined range, the IP address fails to pass the filtering and cannot be matched against a next node. If the IP address is out of the range, the system moves to a next node.
- *ipv4-address mask-length*: specifies the IP address and mask length. The mask-length value is an integer ranging from 0 to 32.

IP Prefix Configuration Example (1)



Single-Statement Matching

```
ip ip-prefix aa index 10 permit 10.1.1.0 24
```

- Case 1: The route 10.1.1.0/24 is permitted, and other routes are rejected.

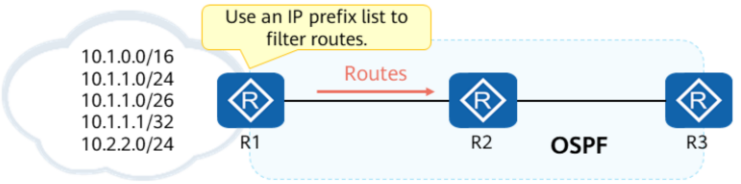
```
ip ip-prefix bb index 10 deny 10.1.1.0 24
```

- Case 2: All routes are rejected.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Case1: This is the single-node exact matching. Only the route of the specified destination address and mask can match the IP prefix. In addition, the matching mode of the node is permit. As such, the route 10.1.1.0/24 matches the node and is permitted, and the other routes are rejected because they fail to match the IP prefix.
- Case 2: This is the single-node exact matching, and the matching mode of the node is deny. As such, the route 10.1.1.0/24 matches the node and is rejected, and other routes are rejected by default because they do not match the IP prefix.

IP Prefix Configuration Example (2)



Multi-Statement Matching

```
ip ip-prefix aa index 10 deny 10.1.1.0 24
ip ip-prefix aa index 20 permit 10.1.1.1 32
```

- Case 1: The route 10.1.1.0/24 is rejected, the route 10.1.1.1/32 is permitted, and other routes are rejected.

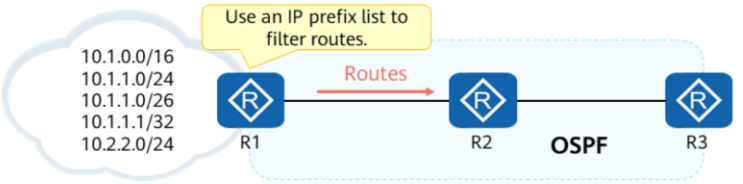
```
ip ip-prefix bb index 10 permit 10.1.1.0 24 greater-equal 26 less-equal 32
```

- Case 2: Routes 10.1.1.0/26 and 10.1.1.1/32 are permitted, and the other routes are rejected.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- Case 1: This is the multi-node exact matching.
  - When the route 10.1.1.0/24 is matched against index 10, the route meets the matching condition but is rejected because the matching mode is deny.
  - The route 10.1.1.1/32 does not match index 10 and continues to be matched against index 20. The matching is successful, and the matching mode of index 20 is permit, indicating that the route is permitted.
  - Other routes are rejected by default because they do not meet the conditions of indexes 10 and 20.
- Case 2: In this case, greater-equal-value is 26, and less-equal-value is 32. The setting must meet the following formula: mask-length <= greater-equal-value <= less-equal-value. Otherwise, the configuration fails.

IP Prefix Configuration Example (3)



Wildcard address matching

```
ip ip-prefix aa index 10 permit 10.0.0.0 8 less-equal 32
```

- Case 1: All routes with the mask length ranging from 8 to 32 bits are permitted.

```
ip ip-prefix bb index 10 deny 10.1.1.0 24 less-equal 32
ip ip-prefix bb index 20 permit 10.1.0.0 16 less-equal 32
```

- Case 2: The route 10.1.0.0/16 is permitted, and other routes are rejected.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- Case 1: In this case, greater-equal-value is 8, less-equal-value is 32, routes whose left-most 8 bits match those in 10.0.0.0 and with the mask length ranging from 8 to 32 bits match the IP prefix.
- Case 2:
  - For index 10, greater-equal-value is 24, and less-equal-value is 32. Routes whose the left-most 24 bits match those in 10.1.1.0 and with the mask length ranging from 24 to 32 bits are rejected.
  - For index 20, greater-equal-value is 0, and less-equal-value is 32. Routes whose left-most 16 bits match those in 10.1.0.0, and with the mask length ranging from 16 to 32 bits match the prefix. In this case, only the route 10.1.0.0/16 is a match, and the other routes are rejected by default.

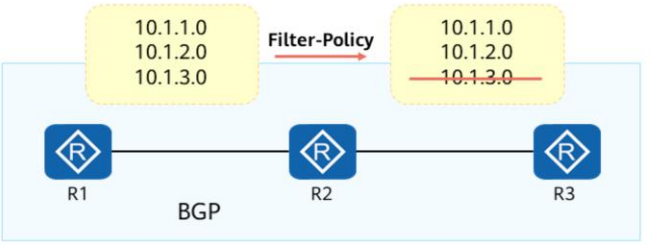
Contents

1. Routing Control Overview
2. **Route Control Tool**
   - Route Matching Tool
   - **Route-Policy**
3. Route Control Cases

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Policy Tool 1: Filter-Policy

- The filter-policy is a common tool for filtering routing information. It can filter routes to be accepted, advertised, and imported. The filter-policy applies to IS-IS, OSPF, and BGP.

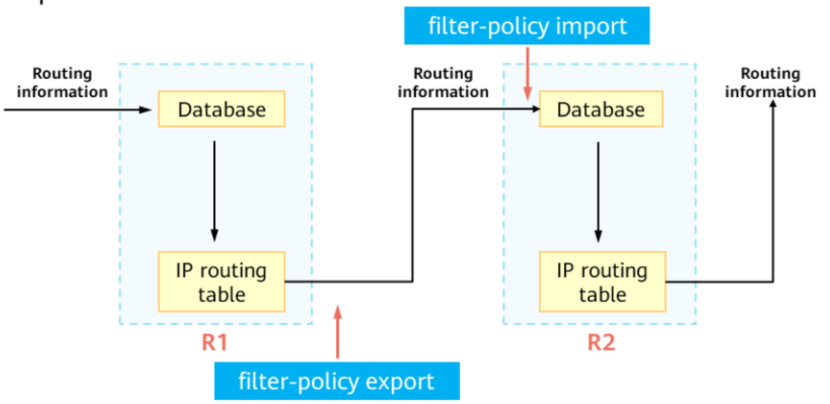


- As shown in the preceding figure, BGP runs between R1, R2, and R3. Routes are transmitted between devices. To filter certain routing information as needed, you can use the filter-policy.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Filter-Policy Application in Distance-Vector Routing Protocols

In distance-vector routing protocols, routing information is transmitted between devices. To filter such information, you can use filter-policies. The following figure shows the locations where the filter-policies take effect in the inbound and outbound directions.

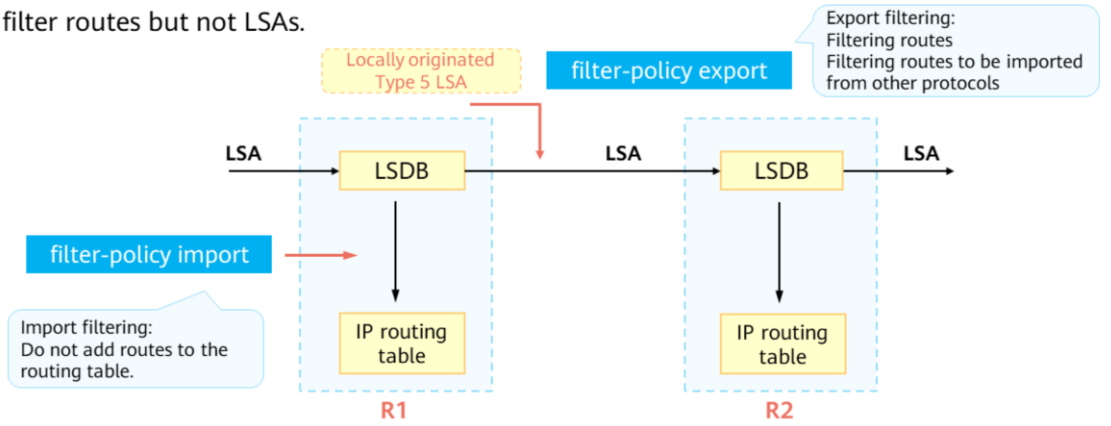


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A distance-vector protocol generates routes based on the routing table. Consequently, the filter affects the routes to be accepted from neighbors and the routes to be advertised to neighbors.
- To filter out the routes from an upstream device to a downstream device, run the **filter-policy export** command on the upstream device or the **filter-policy import** command on the downstream device.

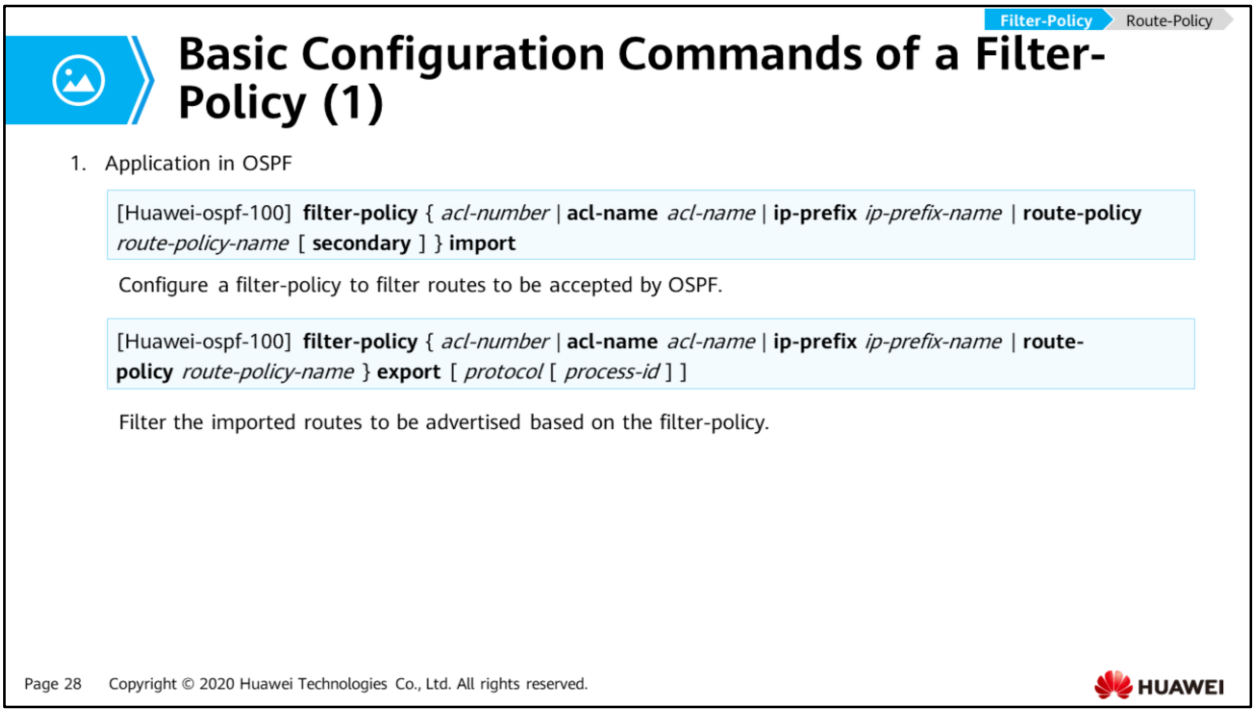
Filter-Policy Application in Link-State Routing Protocols

In a link-state routing protocol, routing devices exchange LSAs, and then calculate entries in the routing table based on LSDB information summarized from the LSAs. A filter-policy, however, can filter routes but not LSAs.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- OSPF stores the flooded LSAs in its LSDB and runs the SPF algorithm to calculate a loop-free SPT with the local device as the root. The filter-policy module filters the routes calculated by OSPF (before the routes are installed into the routing table) but does not filter LSAs.
- The preceding example uses OSPF to show how a filter-policy is applied in a link-state routing protocol.



- Command: [Huawei-ospf-100**] filter-policy** {*acl-number*| **acl-name** *acl-name*| **ip-prefix** *ip-prefix-name*| **route-policy** *route-policy-name* [**secondary**]} **import**
  - *acl-number*: specifies the number of a basic ACL. The value is an integer ranging from 2000 to 2999.
  - **acl-name** *acl-name*: specifies the name of an ACL. The value is a string of 1 to 32 case-sensitive characters, spaces not supported. The value must start with a letter (a to z or A to Z).
  - **ip-prefix** *ip-prefix-name:* specifies the name of an IP prefix list. The value is a string of 1 to 169 case-sensitive characters, spaces not supported. If spaces are used, the string must start and end with double quotation marks (").
  - **route-policy** *route-policy-name:* specifies the name of a route-policy. The value is a string of 1 to 40 case-sensitive characters, spaces not supported. If spaces are used, the string must start and end with double quotation marks (").
  - **secondary**: indicates that the sub-optimal route is selected.
- Command: [Huawei-ospf-100] **filter-policy** { *acl-number*| **acl-name** *acl-name*| **ip-prefix** ip-prefix-name | **route-policy** route-policy-name } **export** [**protocol** [*process-id*]]
  - **protocol** *process-id*: specifies the protocol whose routes are to be filtered. Currently, the value can be direct, isis, bgp, ospf, unr, or static. When RIP, IS-IS, or OSPF is specified as a routing protocol, you can also specify a process ID. The value is an integer ranging from 1 to 65535. The default value is 1.



Basic Configuration Commands of a Filter-Policy (2)

2. Application in IS-IS

[Huawei-isis-1] **filter-policy** { *acl-number* | **acl-name** *acl-name* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **import**

Configure a filter-policy to filter IS-IS routes to be added to the IP routing table.

[Huawei-isis-1] **filter-policy** { *acl-number* | **acl-name** *acl-name* | **ip-prefix** *ip-prefix-name* | **route-policy** *route-policy-name* } **export** [*protocol* [*process-id*]]

Configure a filter-policy for IS-IS to filter imported routes to be advertised.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Configuration Commands of a Filter-Policy (3)

3. Application in BGP

[Huawei-bgp-af-ipv4] **filter-policy** { *acl-number* | **acl-name** *acl-name* | **ip-prefix** *ip-prefix-name* } **import**

Configure a filter-policy to filter routes to be accepted by BGP.

[Huawei-bgp-af-ipv4] **filter-policy** { *acl-number* | **acl-name** *acl-name* | **ip-prefix** *ip-prefix-name* } **export** [*protocol* [*process-id*]]

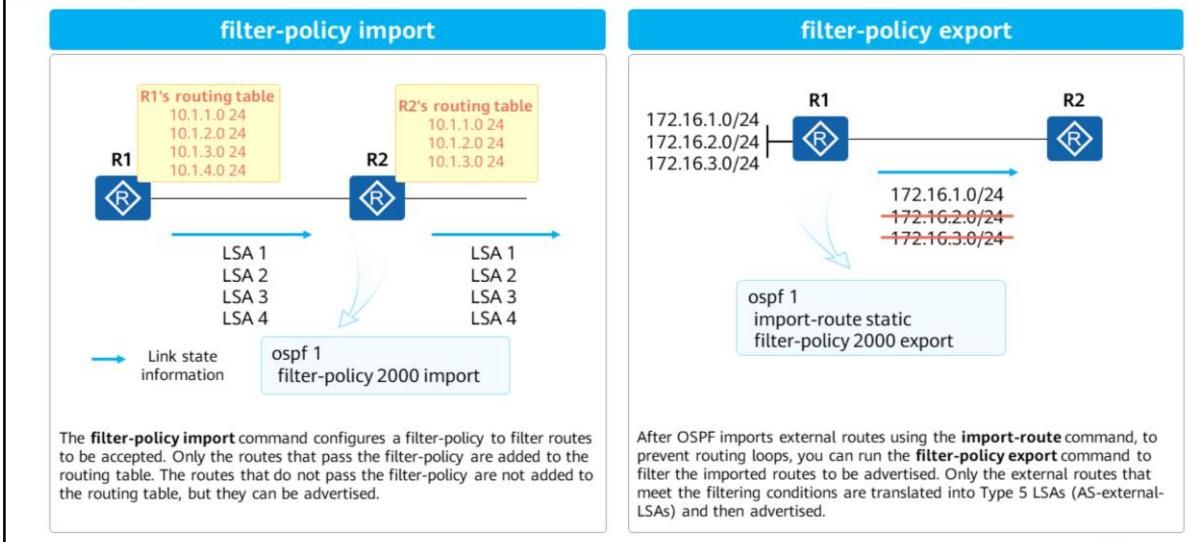
Configure a filter-policy to filter routes to be advertised. Only the routes that pass the filtering can be advertised by BGP.

[Huawei-bgp-af-ipv4] **peer** { *group-name* | *ipv4-address* } **filter-policy** { *acl-number* | **acl-name** *acl-name* } { **import** | **export** }

Configure a filter-policy to filter routes to be advertised to or accepted from a specified peer or peer group.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

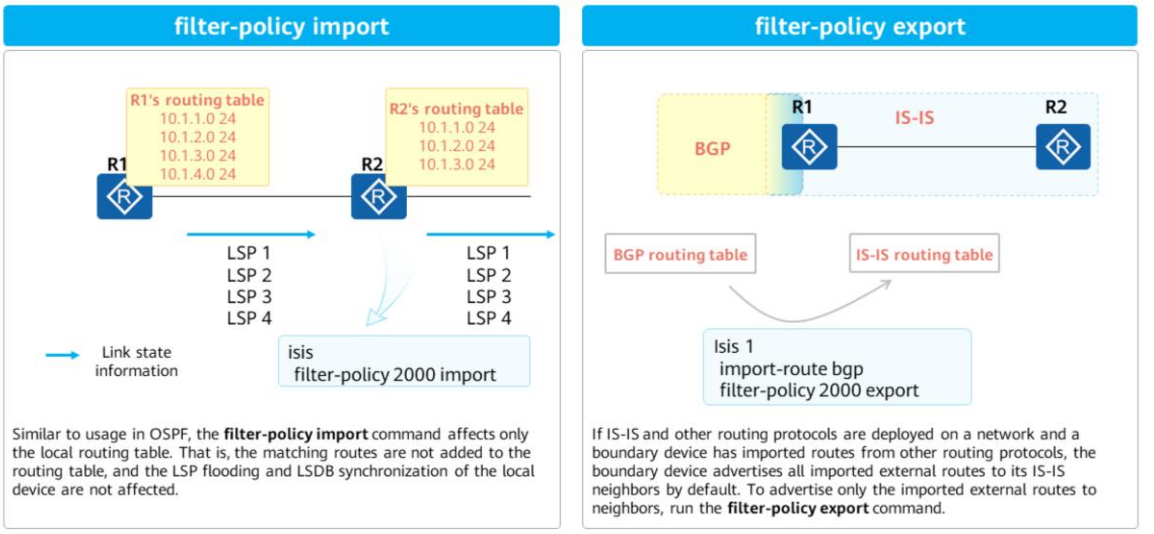
Applying a Filter-Policy to OSPF



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- OSPF routing information is recorded in the LSDB. The **filter-policy import** command is used to filter the routes calculated by OSPF, but not to filter LSAs to be accepted or advertised.
- The **filter-policy export** command allows you to specify a protocol or process ID to filter the routes of a specified protocol or a specified process. If neither protocol nor process-id is specified, OSPF filters all imported routes.

Applying a Filter-Policy to IS-IS



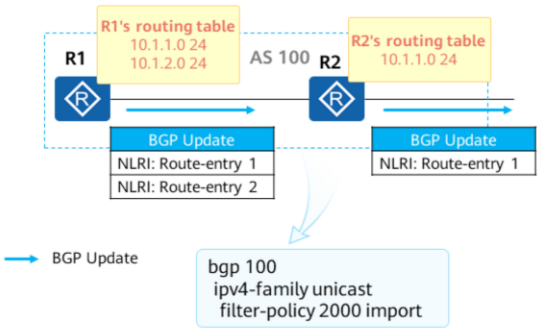
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- IS-IS routing entries can be used to guide IP packet forwarding only after they are successfully delivered to the IP routing table. If an IS-IS routing table has routes destined for a specific network segment but these routes do not need to be added to the IP routing table, run the **filter-policy import** command and use a basic ACL, an IP prefix list, or a route-policy to filter the IS-IS routes to be added to the IP routing table.

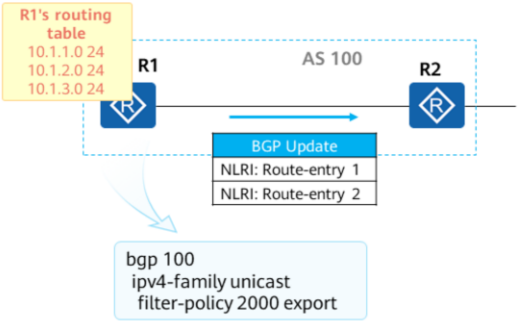
Applying a Filter-Policy to BGP

filter-policy import



The **filter-policy import** command can be used to filter the routes to be accepted by BGP globally and determine whether to add the routes to the BGP routing table.

filter-policy export



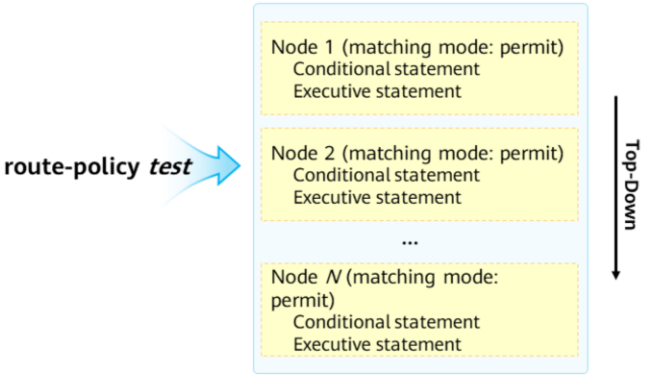
The **filter-policy export** command is used to filter the routes to be advertised. Only the routes that pass the filtering can be added to the local BGP routing table and advertised by BGP.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Policy Tool 2: Route-Policy

- A route-policy is a policy tool used to filter routes and set route attributes for the filtered routes.
- A route-policy consists of one or more nodes. Each node can be a set of conditional statements (matching conditions) and executive statements (actions). These statements are arranged in ascending order by sequence number.



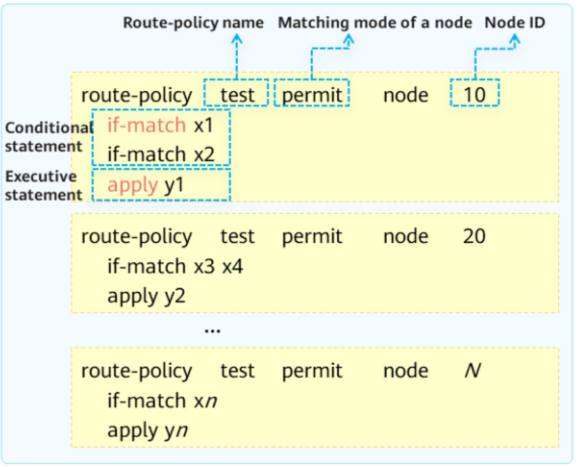
- Each node contains multiple conditional statements. The relationship between multiple conditional statements on a node is AND. The action on the node is executed only when all conditional statements are met.
- The relationship between nodes is OR. The route-policy is executed in ascending order by node ID. A node in a route-policy will not be further matched.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- When a route-policy is used, the node with a smaller ID is matched first. After a route matches a node, the route is not matched against other nodes. If a route fails to match all nodes, the route is filtered out.

Components of a Route-Policy

A route-policy consists of one or more nodes. Each node contains multiple if-match and apply clauses.

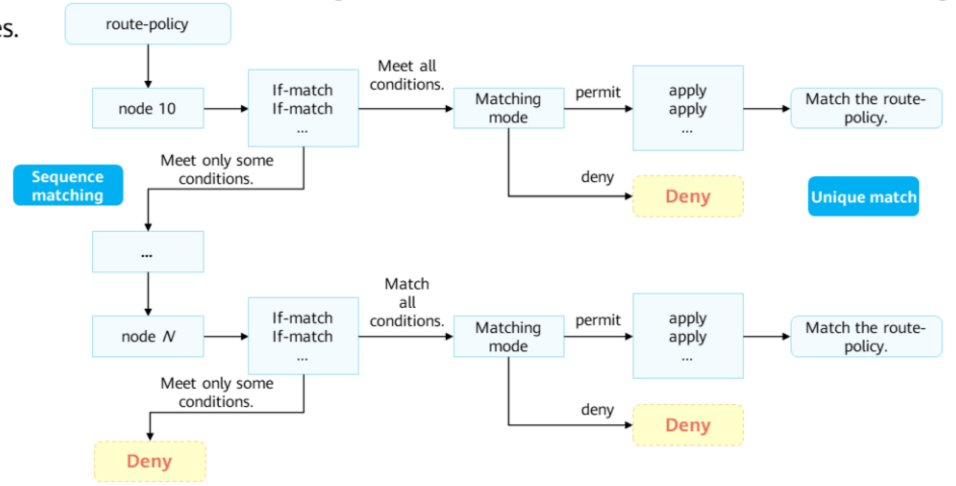


- **permit** or **deny**: The matching mode of a route-policy node is "permit" or "deny".
- *node*: specifies the node ID of a route-policy. The value is an integer ranging from 0 to 65535.
- if-match clause: defines the matching condition of a node.
- apply clause: defines an action to be performed on a matching route.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Matching Order for a Route-Policy

Route-policies use different matching conditions and modes to select routes and change route attributes.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A route-policy contains N (N >= 1) nodes. After routes match the route-policy, the system checks whether the routes match the nodes in ascending order by node ID. The matching condition is defined in the if-match clause.
  - After a route matches all if-match clauses of a node, the system proceeds to select a matching mode and no longer matches the route against other nodes. The matching mode can be "permit" or "deny."
    - **permit:** The route is permitted, and the apply clause of the node is used to set some attributes of the route.
    - **deny**: The route is rejected.
  - If a route fails to match any if-match clause of the node, the route is further matched against the next node. If a route does not match any node, the route is rejected.

Basic Configuration Commands of a Route-Policy (1)

1. Create a route-policy.

```
[Huawei] route-policy route-policy-name { permit | deny } node node
```

Create a route-policy and enter the route-policy view.

2. (Optional) Configure if-match clauses.

```
[Huawei-route-policy] if-match ?
  acl: matches a basic ACL.
  cost: matches the cost of a route.
  interface: matches the outbound interface in a route.
  ip-prefix: matches a prefix list.
  ...
```

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Command: **route-policy** *route-policy-name* { **permit | deny** } **node** *node*
  - **permit**: sets the matching mode of a route-policy node to permit. If a route matches all if-match clauses of a node, the apply clause of the node is executed. Otherwise, the system goes to the next node.
  - **deny**: indicates that the matching mode of the route-policy node is deny. If a route matches all if-match clauses of a node, the route is rejected. Otherwise, the system goes to the next node.
  - **node** *node:* specifies the node ID of a route-policy. When a route-policy is used, the node with the smallest node ID is matched first. After a route matches a node, the route is not matched against other nodes. If a route fails to match any nodes, the route is filtered out. The value is an integer ranging from 0 to 65535.

Basic Configuration Commands of a Route-Policy (2)

3. (Optional) Configure an apply clause.

```
[Huawei-route-policy] apply ?
  cost: sets the cost of a route.
  cost-type {type-1 | type-2}: sets the OSPF cost type.
  ip-address next-hop: sets the next-hop address for an IPv4 route.
  preference: sets the preference of a routing protocol.
  tag: sets the tag field in routing information.
  ...
```

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- The apply clause is used to specify an action for the route-policy and set the attributes of the routes that match the route-policy. If no apply clause is configured for a node, the node only filters routes. If one or more apply clauses are configured, all apply clauses are applied to the routes that match the node.

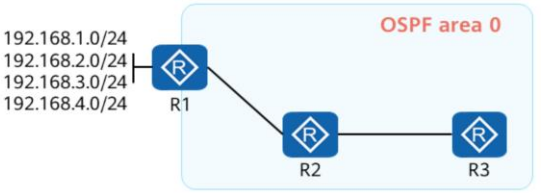
Contents

1. Routing Control Overview
2. Implementation of Route Control
   - Route Matching Tool
   - Route-Policy
3. **Route Control Cases**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Filtering Received Routes



- OSPF runs on R1, R2, and R3. R1 advertises the routes 192.168.1.0/24, 192.168.2.0/24, 192.168.3.0/24, and 192.168.4.0/24 to OSPF.
- It is required that R2 cannot access network segment 192.168.1.0/24 on R1, but R3 can.
- To meet this requirement, you can use a filter-policy to filter the received routes on R2.

The configuration of R2 is as follows:

```
[R2] ip ip-prefix in index 10 permit 192.168.2.0 24
[R2] ip ip-prefix in index 20 permit 192.168.3.0 24
[R2] ip ip-prefix in index 30 permit 192.168.4.0 24

[R2] ospf
[R2-ospf-1] filter-policy ip-prefix in import
```

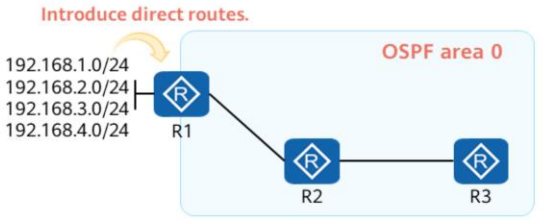
Note: Basic network configurations are not provided.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- When a filter-policy is used to filter routes, only the local routing table is affected. Therefore, R3 can learn the routes to network segment 192.168.1.0/24.

Filtering Routes to Be Advertised



- OSPF runs on R1, R2, and R3. R1 imports the direct routes 192.168.1.0/24, 192.168.2.0/24, 192.168.3.0/24, and 192.168.4.0/24 to OSPF.
- It is required that R2 and R3 learn only the routes to 192.168.1.0/24 and not the routes to the other three network segments.
- To meet this requirement, configure a filter-policy on R1 to filter the imported routes to be advertised.

The configuration of R1 is as follows:

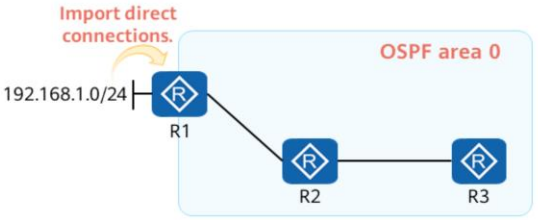
```
[R1] ip ip-prefix out index 10 permit 192.168.1.0 24

[R1] ospf
[R1-ospf-1] import-route direct
[R1-ospf-1] filter-policy ip-prefix out export
```

Note: Basic network configurations are not provided.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Modifying Route Attributes



- OSPF runs on R1, R2, and R3. R1 imports the direct route 192.168.1.0/24 to OSPF.
- It is required that the OSPF route 192.168.1.0/24 learned by R2 and R3 be an external Type 1 route. By default, the route is an external Type 2 route.
- To meet this requirement, you can use a route-policy on R1 to change the type of external routes to external Type 1 when routes are imported.

The configuration of R1 is as follows:

```
[R1] ip ip-prefix external index 10 permit 192.168.1.0 24

[R1] route-policy RP permit node 10
[R1-route-policy] if-match ip-prefix external
[R1-route-policy] apply cost-type type-1
[R1-route-policy] quit

[R1] ospf
[R1-ospf-1] import-route direct route-policy RP
```

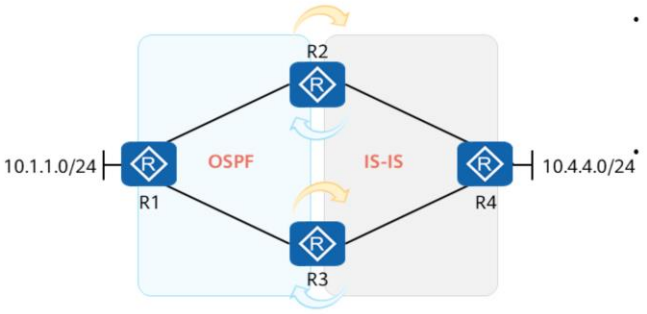
Note: Basic network configurations are not provided.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Dual-Node Bidirectional Route Import

Import OSPF routes to IS-IS.

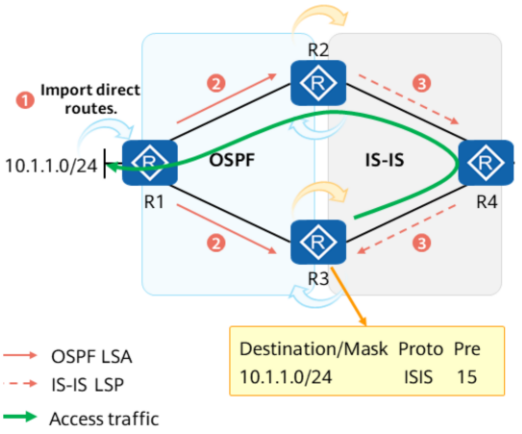
Import IS-IS routes to OSPF.



- Bidirectional route import refers to the process in which boundary routers in two routing domains imports routes into each other.
- If two border routers are on the borders of two routing domains and bidirectional route import is performed on both border routers, this is called dual-node bidirectional route import.
- Dual-node bidirectional route import is a typical routing model. Single-node bidirectional route import lacks redundancy. Once a single-node border router is faulty, communication between two routing domains may fail. Therefore, dual-node bidirectional route import is generally used in large-scale network deployment.
- Although this function enhances network reliability, it may cause problems, such as sub-optimal paths and routing loops.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Suboptimal Path Problem

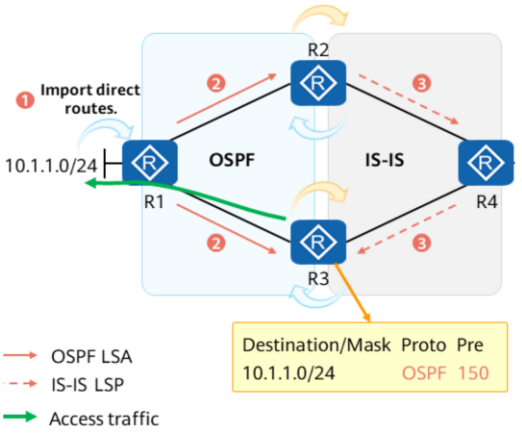


The following uses the direct route 10.1.1.0/24 as an example:

- R1 imports the direct route 10.1.1.0/24 to OSPF.
- R2 and R3 re-advertise bidirectional routes. R2 re-advertises the route 10.1.1.0/24 to IS-IS, and R3 learns the IS-IS route from R4.
- For R3, the IS-IS route with preference 15 is preferred over the OSPF external route with preference 150. Therefore, the IS-IS route from R4 is preferred. The path through which R3 accesses the network segment 10.1.1.0/24 is R3 -> R4 -> R2 -> R1, which is the second optimal path.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solving the Suboptimal Path Problem (1)



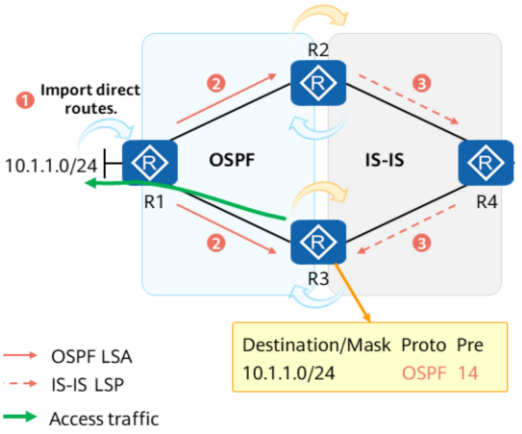
- Solution 1: In the IS-IS process of R3, configure a filter-policy to prevent the route 10.1.1.0/24 sent by R4 from being added to the local routing table.
- Perform the following operations on R3:

```
[R3] acl 2001
[R3-acl-basic-2001] rule 5 deny source 10.1.1.0 0
[R3-acl-basic-2001] rule 10 permit

[R3] isis
[R3-isis-1] filter-policy 2001 import
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solving the Suboptimal Path Problem (2)

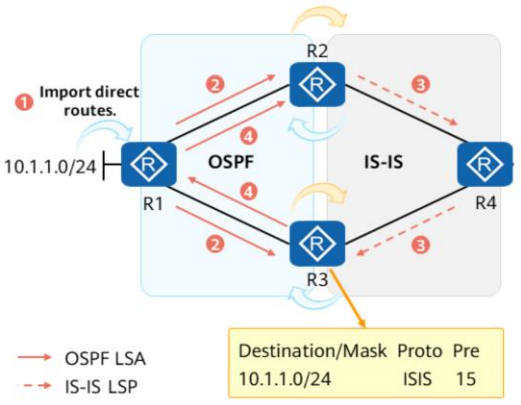


- Solution 2: Configure an ACL on R3 to match the route 10.1.1.0/24, apply the ACL to the route-policy, and set the preference of the route that matches the ACL to 14 (higher than IS-IS). Run the **preference ase** command in the OSPF view to invoke the route-policy to change the preference of external routes.
- Perform the following operations on R3:

```
[R3]acl 2000
[R3-acl-basic-2000] rule permit source 10.1.1.0 0
[R3-acl-basic-2000] quit
[R3]route-policy hcip permit node 10
[R3-route-policy] if-match acl 2000
[R3-route-policy] apply preference 14
[R3-route-policy] quit
[R3]ospf 1
[R3-ospf-1] preference ase route-policy hcip
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Routing Loops



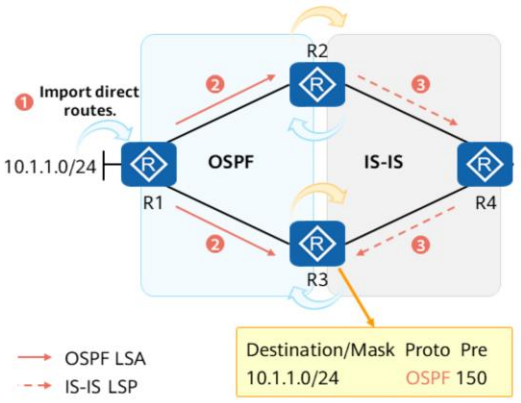
Scenario description:

1. Import the direct route 10.1.1.0/24 to OSPF on R1.
2. Configure OSPF on R1, R2, and R3. The route to the network segment 10.1.1.0/24 is advertised in the entire OSPF area.
3. R2 re-advertises bidirectional routes.
4. Configure IS-IS on R2, R3, and R4. The route to the network segment 10.1.1.0/24 is advertised in the entire IS-IS domain.
5. R3 re-advertises bidirectional routes.
6. The route destined for the network segment 10.1.1.0/24 is advertised to the OSPF area again, forming a routing loop.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Preventing Routing Loops (1)



- Solution 1: Configure a route-policy on R3 to filter out the route destined for 10.1.1.0/24 when OSPF imports IS-IS routes.
- Perform the following operations on R3:

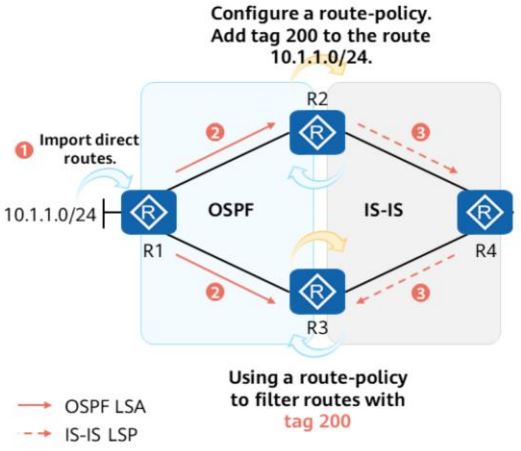
```
[R3] acl 2001
[R3-acl-basic-2001] rule 5 deny source 10.1.1.0 0
[R3-acl-basic-2001] rule 10 permit

[R3] route-policy RP permit node 10
[R3-route-policy] if-match 2001
[R3-route-policy] quit

[R3] ospf
[R3-ospf-1] import-route isis 1 route-policy RP
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preventing Routing Loops (2)

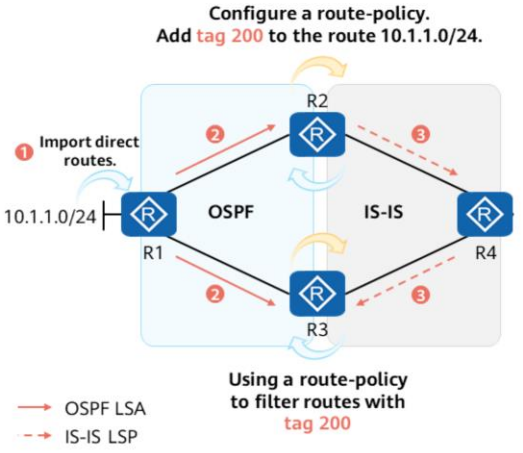


- Solution 2: Use tags to implement selective route import. Add tag 200 to the route 10.1.1.0/24 imported from OSPF to IS-IS on R2, and filter the route with tag 200 when importing the route from IS-IS to OSPF on R3.
- Perform the following operations on R2:

```
[R2]acl 2000
[R2-acl-basic-2000]rule permit source 10.1.1.0 0
[R2-acl-basic-2000]quit
[R2]route-policy hcip permit node 10
[R2-route-policy]if-match acl 2000
[R2-route-policy]apply tag 200
[R2-route-policy]quit
[R2]isis 1
[R2-isis-1]import-route ospf route-policy hcip
```

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Preventing Routing Loops (3)



- Perform the following operations on R3:

```
[R3]route-policy hcip deny node 10
[R3-route-policy]if-match tag 200
[R3-route-policy]quit
[R3]route-policy hcip permit node 20

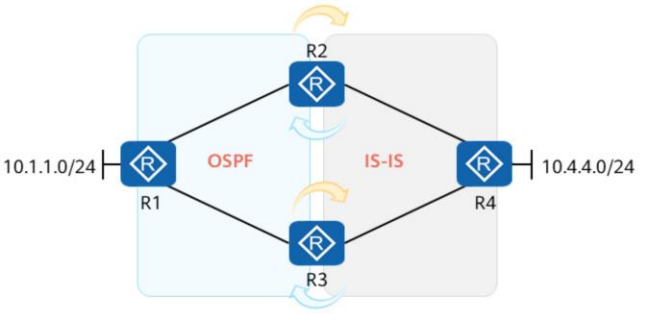
[R3]ospf 1
[R3-ospf-1]import-route isis route-policy hcip
```

Although IP prefix-based route filtering can be used for route re-advertisement, the involved configuration workload is heavy on a large network. Instead of using IP prefixes, setting tags greatly reduces the configuration workload.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Thinking About the Following Scenario



On R1, import the network segment route 10.1.1.0/24 to OSPF; on R4, import the network segment route 10.4.4.0/24 to IS-IS; on R2 and R3, import routes bidirectionally. In this scenario, how can we prevent routing loops by setting tags?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Essay) What are the functions of export filter-policies in OSPF and BGP?
2. (Essay) What is the logical relationship between nodes in a route-policy? What is the logical relationship between multiple conditional statements on a node?

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. In OSPF, the export filter-policy is used to filter the routes to be imported from other routing protocols to OSPF. In BGP, the export filter-policy is used to filter routes to be advertised.
2. The logical relationship between nodes is OR, and the logical relationship between conditional statements is AND.

Summary

- To control the advertisement and receipt of routes, use a tool to obtain routes. The most common tools are ACL and IP prefix list.
- Both filter-policies and route-policies can be used to filter routes to be accepted or advertised. Note that the use of filter-policies in link-state routing protocols cannot filter link state information, but affects only the local routing table.
- Before a device accepts or advertises routes, the device can use a route-policy to flexibly modify route attributes to meet route control requirements.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Traffic Filtering and Forwarding Path Control

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- Traditionally, devices search their routing tables based on packets' destination addresses and then forward the packets. As services develop, in addition to using traditional modes, users want to forward packets and select routes through customized policies.
- To improve network security, users want to control the packets entering the network and isolate the unauthorized packets or the packets that have security risks at the network edge.
- This course discusses the traffic filtering technology, policy-based routing (PBR), as well as how to use the modular QoS command line interface (MQC) to control packet forwarding paths and filter traffic.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe the application scenarios of PBR.
  - Describe the classification of PBR.
  - Describe how to configure PBR in MQC mode.
  - Describe how to filter packets in MQC mode.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

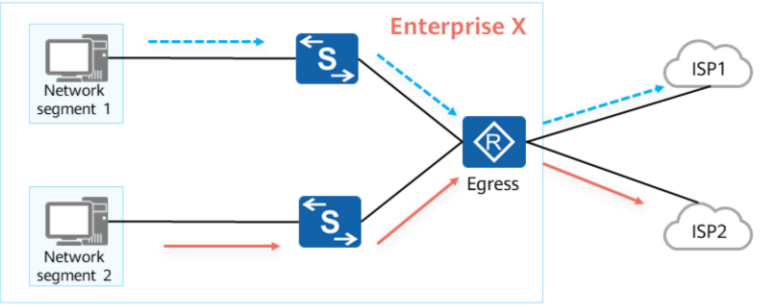
Contents

1. **PBR**
2. MQC
3. Traffic Filtering

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Technical Background

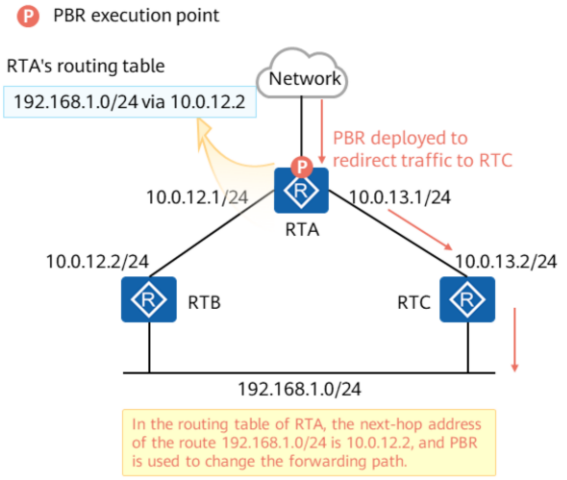
In some scenarios, the traffic of specific users or services needs to be forwarded along a specified path, whereas the traffic of other users or services can still be forwarded based on the routing information base (RIB).



Example: An enterprise connected to two ISPs wants to access the Internet from network segment 1 through ISP1 and access the Internet from network segment 2 and through ISP2. This requirement cannot be met using traditional routing technologies.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

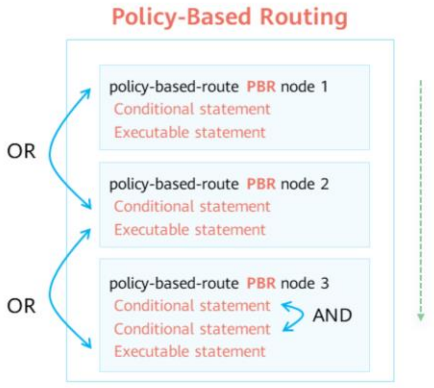
PBR Overview: Basic Concepts



- Policy-based routing (PBR) enables a network device to forward data based on not only the destination IP address of a packet, but also other elements such as the source IP address, source MAC address, destination MAC address, source port number, destination port number, and VLAN ID.
- You can also use an ACL to match specific packets and deploy PBR based on the ACL.
- If PBR is deployed on a device, the matched packets are preferentially forwarded based on a PBR policy. That is, a PBR policy has a higher priority than a traditional IP routing table in packet forwarding.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

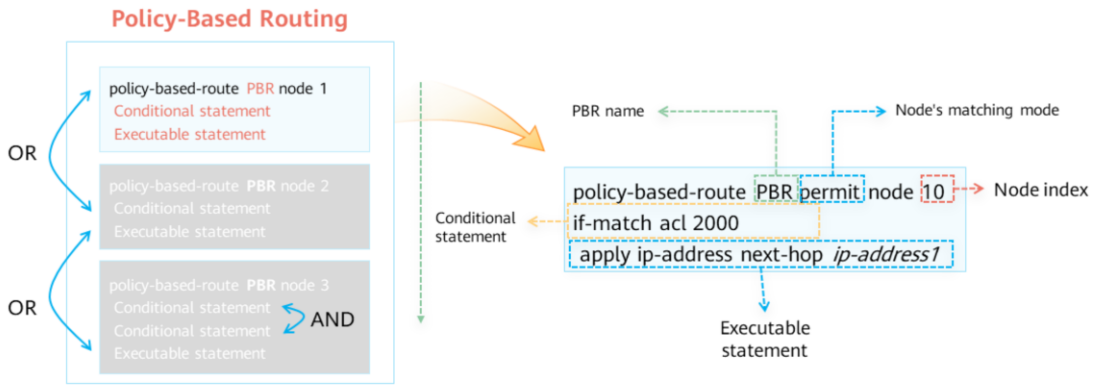
PBR Overview: Structure



- Similar to a route-policy, a PBR policy consists of multiple nodes, each of which consists of matching conditions (conditional statements) and actions (executable statements).
- Each node can contain multiple conditional statements.
- If the relationship between multiple conditional statements in a node is AND, actions in the node can be executed only when all conditional statements are matched.
- If the relationship between multiple conditional statements in a node is OR, PBR is executed based on node IDs in ascending order. The matching is stopped if the conditional statements of a node are matched.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PBR Overview: Command Syntax



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- PBR supports the following node matching modes:
  - permit: indicates that PBR is performed on the packets that meet the matching conditions.
  - deny: indicates that PBR is not performed on the packets that meet the matching conditions.

Differences Between PBR and Route-Policy

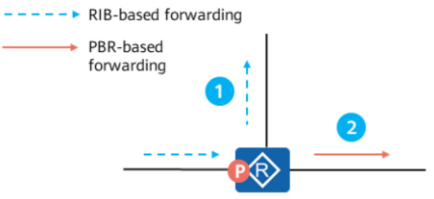
Item	Operation Object	Description
Route-policy	Routing information	A route-policy is a set of methods used to filter routing information and set route attributes. A route-policy sets or controls routes to affect the forwarding paths of data packets.
PBR	Data packet	PBR directly processes data packets, matches concerned packets using multiple methods, and then discards the unwanted packets or forcibly forwards packets along a specified path.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PBR Classification

Interface PBR

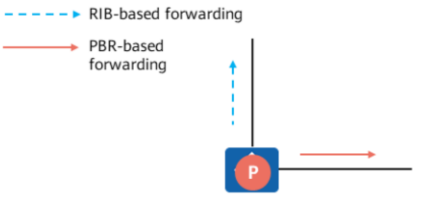
P PBR execution point



- Interface PBR takes effect only for forwarded packets but not locally originated packets.
- Interface PBR is enabled in the interface view and takes effect for incoming packets on the interface. By default, a device forwards packets based on the next hop in the RIB. If interface PBR is configured, the device forwards packets based on the next hop specified by interface PBR.

Local PBR

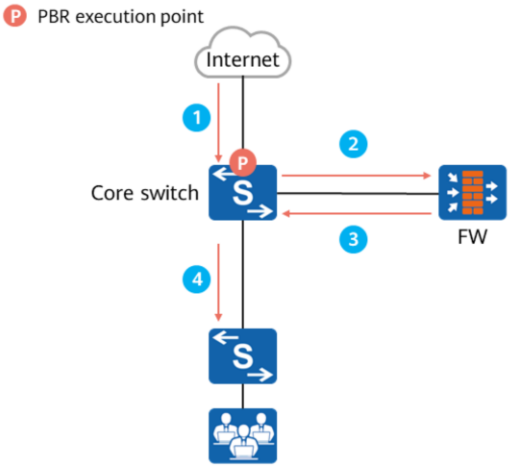
P PBR execution point



- Local PBR takes effect for locally originated traffic, such as the traffic of locally originated ICMP packets.
- Local PBR is enabled globally in the system view.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

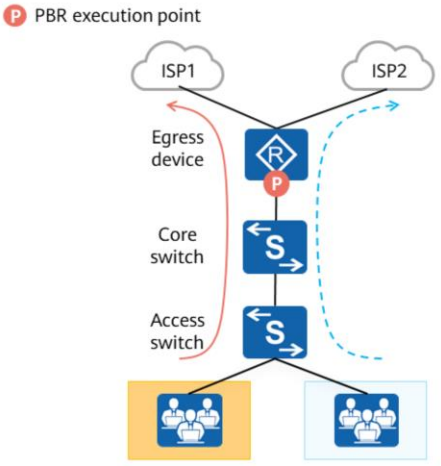
Typical Application Scenarios of PBR (1)



- The intranet firewall is connected to a core switch in off-path mode. To protect the intranet, PBR is deployed on Layer 3 interfaces of the core switch to divert the traffic from external networks to the firewall for security check. After the check is complete, the traffic is sent back to the core switch. The core switch then forwards the traffic to the intranet based on the routing table.
- Traffic diversion means the process of diverting traffic to other devices for security check. PBR is a common traffic diversion tool.

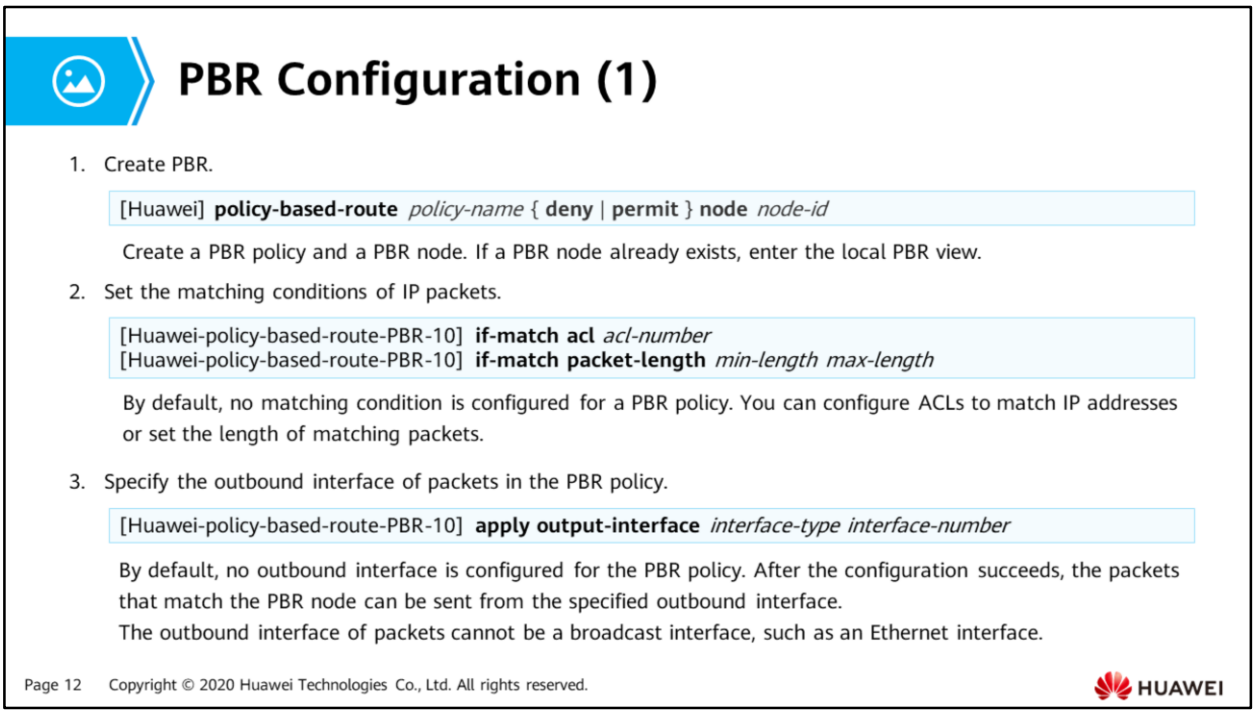
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Typical Application Scenarios of PBR (2)



If an enterprise has multiple egress devices and you want to specify the egress devices for some network segments to access the Internet, configure PBR on intranet interfaces of the egress devices to match the traffic from the intranet and specify different next-hop public address for the egress devices.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- If an ACL rule is set to permit, the device performs the following local PBR actions on the packets matching the ACL rule:
  - When the ACL rule of a PBR node is set to permit, PBR is performed on the packets that meet the matching conditions.
  - When the ACL rule of a PBR node is set to deny, PBR is not performed on the packets that meet the matching conditions, and packets are forwarded based on the destination address through RIB lookup.
- If an ACL is configured with rules, packets that do not match any ACL rule are forwarded according to the destination IP address through RIB lookup.
- If an ACL rule is set to deny or an ACL is not configured with any rule, local PBR that applies the ACL does not take effect, and packets are forwarded according to the destination IP address through RIB lookup.

PBR Configuration (2)

4. Configure the next-hop IP address of packets in a PBR policy.

   [Huawei-policy-based-route-PBR-10] **apply ip-address next-hop** *ip-address1 [ip-address2]*

   You can specify the next hop of packets. If no outbound interface is configured for a PBR node, packets matching PBR node are sent to the specified next hop.

5. Enable PBR globally.

   [Huawei] **ip local policy-based-route** *Policy-name*

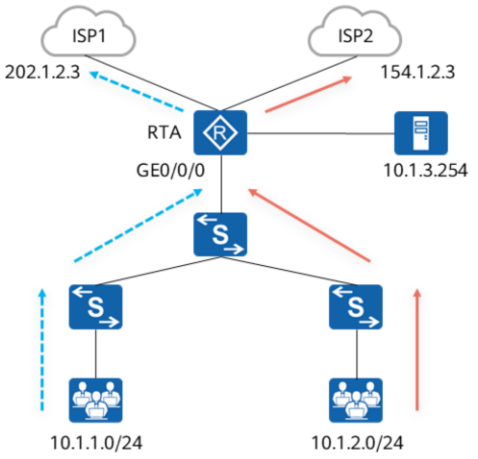
6. Enable PBR on an interface.

   [Huawei-GigabitEthernet0/0/0] **ip policy-based-route** *Policy-name*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In addition to the method described in this slide, interface PBR can also be configured in MQC mode.

Configuration Examples (1)

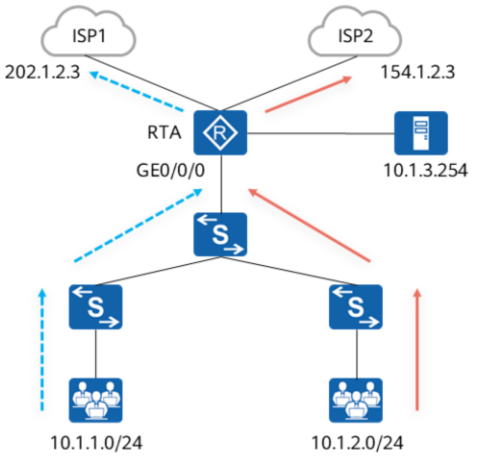


Requirements:

- The intranet has two network segments: network segment 1 (10.1.1.0/24) and network segment 2 (10.1.2.0/24). PBR is configured on GE0/0/0 of RTA so that users on network segment 1 can access the Internet through ISP1 and users on network segment 2 can access the Internet through ISP2.
- RTA is connected to a server in off-path mode. It is required that the PBR deployed on RTA not affect intranet users' access to the server.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Examples (2)



1. Configure ACL 3000. Configure ACL rule 1 to deny the traffic originated from network segment 1 to the server, and configure ACL rule 2 to permit the traffic originated from network segment 1 to the Internet.

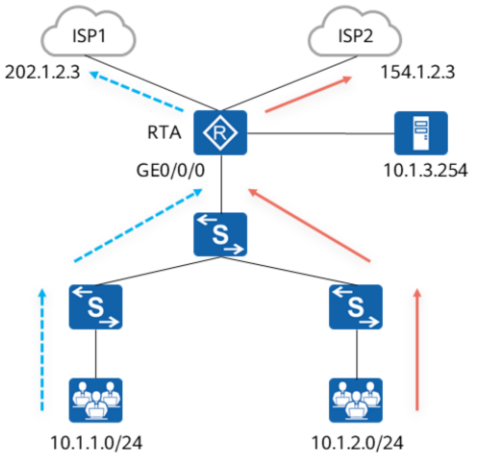
```
[RTA] acl number 3000
[RTA-acl-adv-3000] rule 1 deny ip source 10.1.1.0 0.0.0.255 destination
10.1.3.254 0
[RTA-acl-adv-3000] rule 2 permit ip source 10.1.1.0 0.0.0.255 destination
0.0.0.0 0
```

2. Configure ACL 3001. Configure ACL rule 1 to deny the traffic originated from network segment 2 to the server and ACL rule 2 to permit the traffic originated from network segment 2 to the Internet.

```
[RTA] acl number 3001
[RTA-acl-adv-3001] rule 1 deny ip source 10.1.2.0 0.0.0.255 destination
10.1.3.254 0
[RTA-acl-adv-3001] rule 2 permit ip source 10.1.2.0 0.0.0.255 destination
0.0.0.0 0
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Examples (3)



3. Create a PBR policy named **hcip** and PBR node 10, invoke ACL 3000, and set the next-hop address to 202.1.2.3.

```
[RTA] policy-based-route hcip permit node 10
[RTA-policy-based-route-hcip-10] if-match acl 3000
[RTA-policy-based-route-hcip-10] apply ip-address next-hop 202.1.2.3
```

4. Create PBR node 20 for the PBR policy **hcip**, invoke ACL 3001, and set the next-hop address to 154.1.2.3.

```
[RTA] policy-based-route hcip permit node 20
[RTA-policy-based-route-hcip-20] if-match acl 3001
[RTA-policy-based-route-hcip-20] apply ip-address next-hop 154.1.2.3
```

5. Enable the PBR policy **hcip** on GE0/0/0.

```
[RTA]interface GigabitEthernet 0/0/0
[RTA-GigabitEthernet0/0/0] ip policy-based-route hcip
```

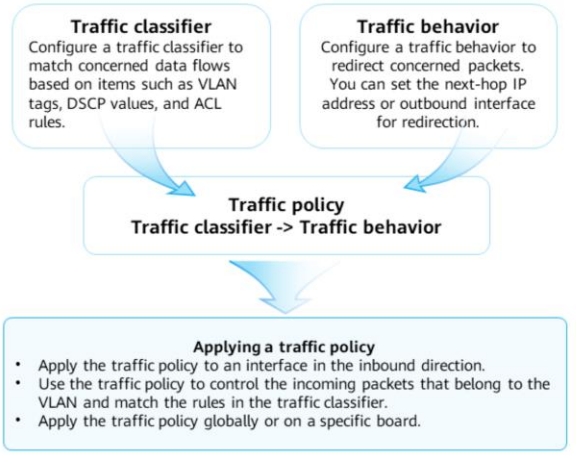

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. PBR
2. **MQC**
3. Traffic Filtering

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

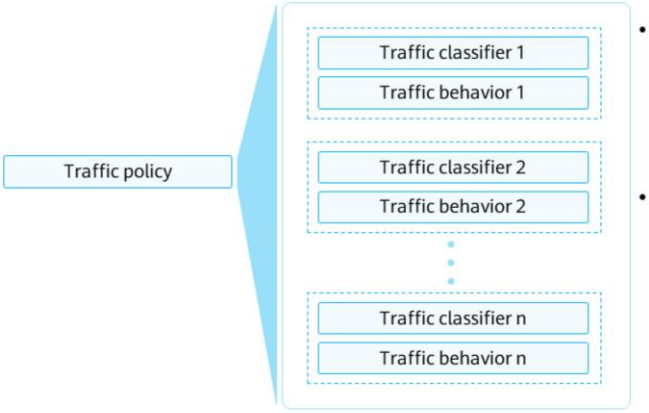
MQC Overview (1)



- Modular QoS command line interface (MQC) groups the data flows that have the same characteristics and provides the same service for a group of data flows. MQC can also provide different services for the data flows of different classes.
- MQC involves three entities: traffic classifier, traffic behavior, and traffic policy.
- MQC traffic behaviors support packet redirection. Therefore, MQC can be used to implement IP unicast policy-based routing.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MQC Overview (2)

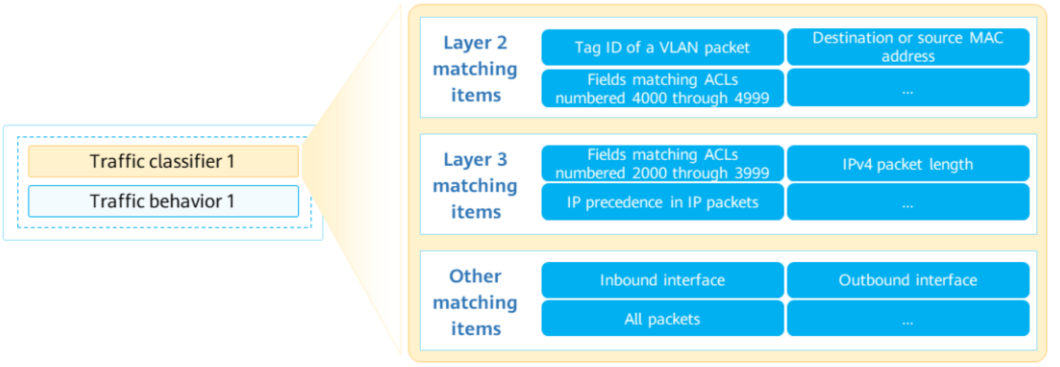


- A traffic policy is bound to a traffic classifier and a traffic behavior so that the device takes the action defined in the traffic behavior on the classified packets.
- A traffic policy can be bound to multiple traffic classifiers and traffic behaviors.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MQC: Traffic Classifier

A traffic classifier defines a group of traffic matching rules to classify packets. This figure illustrates the matching items supported by a traffic classifier.

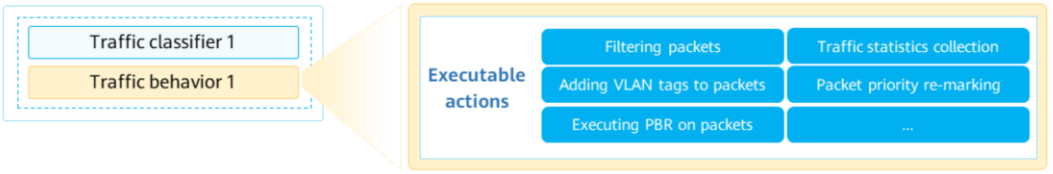


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The relationship between rules in a traffic classifier can be AND or . The default relationship is AND.
  - AND: If a traffic classifier contains ACL rules, packets must match one ACL rule and all non-ACL rules. If a traffic classifier does not contain ACL rules, packets must match all non-ACL rules.
  - OR: If a packet matches a rule in a traffic classifier, the device considers that the packet matches the traffic classifier.

MQC: Traffic Behavior

A traffic behavior defines the actions to be performed, including packet filtering, priority re-marking, redirection, and traffic statistics collection.

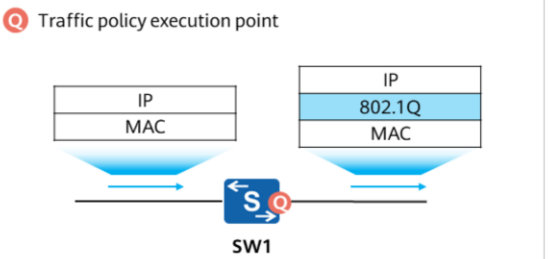


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MQC: Traffic Policy

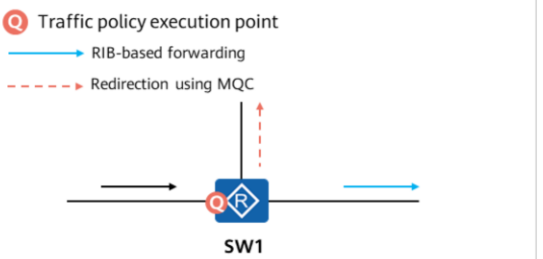
- A traffic policy can be invoked on an interface.
- A traffic policy defines the inbound and outbound directions. A traffic behavior in the traffic policy matches incoming or outgoing packets, and is performed on the matched packets.

Traffic policy invoked in the outbound direction



In the outbound direction, the device applies the traffic policy to add the 802.1Q tag to packets.

Traffic policy invoked in the inbound direction



In the inbound direction, the device applies the traffic policy to redirect packets matching the traffic classifier.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Different from a PBR policy which can be invoked only on Layer 3 interfaces, a traffic policy can be invoked on both Layer 2 and Layer 3 interfaces.

Configuration Procedure

1. Create a traffic classifier.

   [Huawei] **traffic classifier** *classifier-name* **[operator { and | or }]**

   By default, the relationship between rules in a traffic classifier is OR. For details about the matching rules in a traffic classifier, see the corresponding product manual.

2. Create a traffic behavior.

   [Huawei] **traffic behavior** *behavior-name*

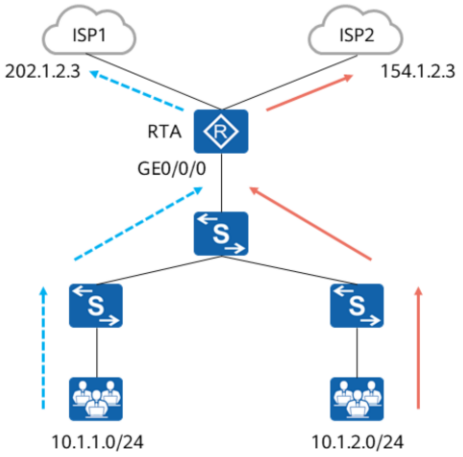
   Define actions in the traffic behavior according to actual conditions. Different actions can be configured in the same traffic behavior so long as they do not conflict. For details about the matching rules in a traffic behavior, see the corresponding product manual.

3. Create a traffic policy. Bind a traffic classifier and a traffic behavior to the traffic policy.

   [Huawei] **traffic policy** *policy-name*
   [Huawei-trafficpolicy-policyname] **classifier** *classifier-name* **behavior** *behavior-name*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PBR Using MQC (1)

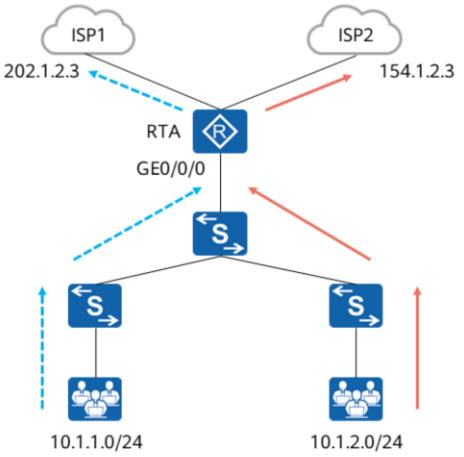


Requirements:

- The intranet has two network segments: network segment 1 (10.1.1.0/24) and network segment 2 (10.1.2.0/24). Configure MQC on RTA to implement PBR, so that users on network segment 1 can access the Internet through ISP1 and users on network segment 2 can access the Internet through ISP2.
- Enable MQC on GE0/0/0 of RTA.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PBR Using MQC (2)



RTA configuration:

1. Configure ACL 3000 to match the traffic originated from network segment 1 to the Internet. Configure ACL 3001 to match the traffic originated from network segment 2 to the Internet.

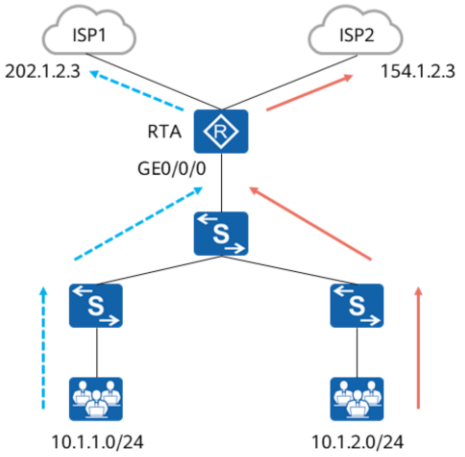
```
[RTA] acl number 3000
[RTA-acl-adv-3000] rule 2 permit ip source 10.1.1.0 0.0.0.255 destination
0.0.0.0 0
[RTA] acl number 3001
[RTA-acl-adv-3001] rule 2 permit ip source 10.1.2.0 0.0.0.255 destination
0.0.0.0 0
```

2. Create traffic classifiers 1 and 2. Bind traffic classifier 1 to ACL 3000, and bind traffic classifier 2 to ACL 3001.

```
[RTA] traffic classifier 1
[RTA-classifier-1] if-match acl 3000
[RTA] traffic classifier 2
[RTA-classifier-2] if-match acl 3001
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PBR Using MQC (3)



3. Create traffic behaviors 1 and 2 to redirect packets to 202.1.2.3 and 154.1.2.3, respectively.

```
[RTA] traffic behavior 1
[RTA-behavior-1] redirect ip-nexthop 202.1.2.3
[RTA] traffic behavior 2
[RTA-behavior-2] redirect ip-nexthop 154.1.2.3
```

4. Create a traffic policy named **Redirect**. Bind traffic classifier 1 to traffic behavior 1, and bind classifier 2 to traffic behavior 2.

```
[RTA] traffic policy Redirect
[RTA-trafficpolicy-Redirect] classifier 1 behavior 1
[RTA-trafficpolicy-Redirect] classifier 2 behavior 2
```

5. Apply the traffic policy **Redirect** in the inbound direction of GE0/0/0.

```
[RTA] interface GigabitEthernet 0/0/0
[RTA-GigabitEthernet0/0/0] traffic-policy Redirect inbound
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

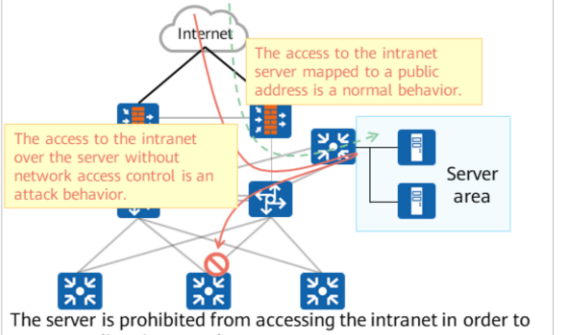
1. PBR
2. MQC
3. **Traffic Filtering**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Requirement Background

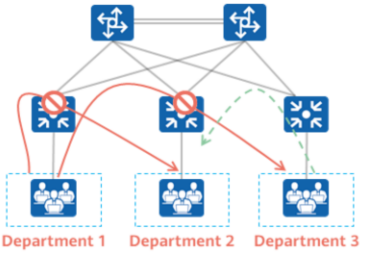
To improve network security, an administrator needs to control the traffic entering the network and discard untrusted packets at the network edge. Untrusted packets are the packets that have security risks and the packets that users are not willing to accept. In addition, to ensure data access security, some departments are not allowed to communicate with each other on an enterprise network.

Discarding Untrusted Packets



The server is prohibited from accessing the intranet in order to prevent redirection attacks.

Restricting Mutual Access Between Departments



Access between departments is restricted based on service requirements.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Traffic Filtering Tools

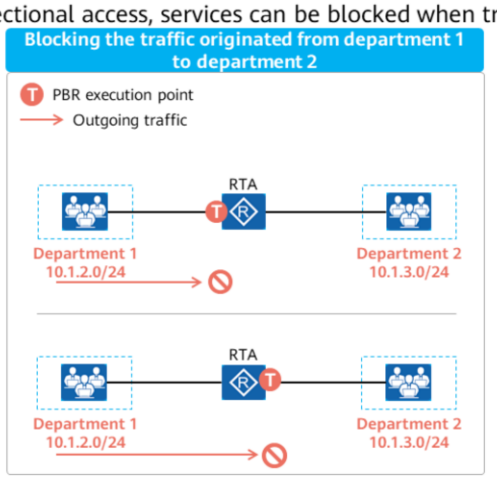


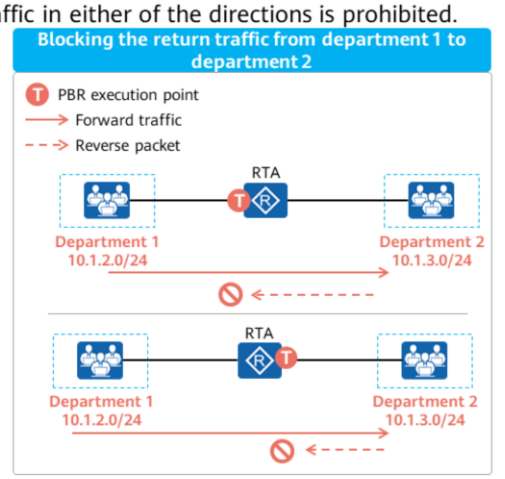
Traffic-filter can be enabled only in the interface view, whereas MQC can be enabled in multiple views.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Traffic-Filter Deployment Position

Traffic-filter can be flexibly deployed to filter the traffic entering a device or leaving a device. For bidirectional access, services can be blocked when traffic in either of the directions is prohibited.

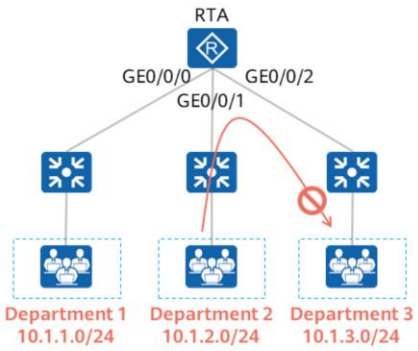




Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The content of the invoked ACL varies according to the deployment position of traffic-filter.

Using Traffic-Filter to Filter Traffic



The gateways of departments 1, 2, and 3 are all on RTA. It is required that traffic-filter be deployed on RTA to restrict the mutual access between departments 2 and 3.

RTA configuration:

1. Configure an ACL to deny the traffic originated from department 2 to department 3 and permit all the other traffic.

```
[RTA] acl number 3000
[RTA-acl-adv-3000] rule 1 deny ip source 10.1.2.0 0.0.0.255 destination 10.1.3.0 0.0.0.255
[RTA-acl-adv-3000] rule 2 permit ip
```

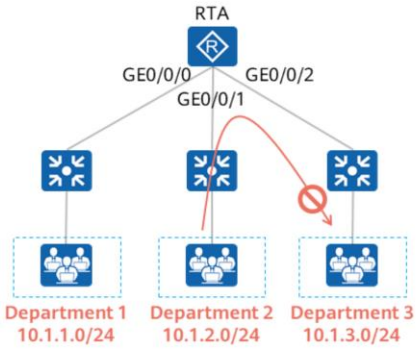
2. Apply the traffic-filter to GE0/0/2.

```
[RTA] interface GigabitEthernet 0/0/2
[RTA-GigabitEthernet0/0/2] traffic-filter outbound acl 3000
```

Question: Are there any other options to configure ACL rule and traffic-filter on interfaces?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Using MQC to Filter Traffic (1)



The gateways of departments 1, 2, and 3 are all on RTA. It is required that MQC be deployed on RTA to restrict the mutual access between departments 2 and 3.

RTA configuration:

1. Configure an ACL to match the traffic originated from department 2 to department 3.

```
[RTA] acl number 3000
[RTA-acl-adv-3000] rule 1 permit ip source 10.1.2.0 0.0.0.255 destination
10.1.3.0 0.0.0.255
```

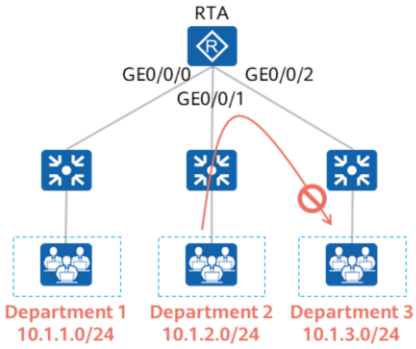
2. Create a traffic classifier named **2_3** and a traffic behavior named **2_3**.

```
[RTA] traffic classifier 2_3
[RTA-classifier-2_3] if-match acl 3000
[RTA] traffic behavior 2_3
[RTA-behavior-2_3] deny
```

When you filter packets matching an ACL rule, if the ACL rule is set to permit, the device processes the packets according to the action (deny or permit) configured in the traffic behavior. If the ACL rule is set to deny, the device discards the packets regardless of whether deny or permit is configured in the traffic behavior.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Using MQC to Filter Traffic (2)



The gateways of departments 1, 2, and 3 are all on RTA. It is required that traffic-filter be deployed on RTA to restrict the mutual access between departments 2 and 3.

3. Create a traffic policy named **2_3**. Bind traffic classifier **2_3** and traffic behavior **2_3** to the traffic policy **2_3**.

```
[RTA] traffic policy 2_3
[RTA-trafficpolicy-2_3] classifier 2_3 behavior 2_3
```

4. Apply the traffic policy **2_3** in the inbound direction of GE0/0/1.

```
[RTA] interface GigabitEthernet 0/0/1
[RTA-GigabitEthernet0/0/1] traffic-policy 2_3 inbound
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Essay) What is the difference between local PBR and interface PBR?
2. (Essay) When MQC is used to filter traffic, what is the difference between the ACL matching a traffic classifier and the ACL invoked by traffic-filter?

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. Local PBR takes effect for locally originated traffic, whereas interface PBR takes effect only for incoming traffic on an interface.
2. In an ACL invoked by MQC, permit and deny indicate whether traffic is matched, instead of the action of permitting or denying traffic. In an ACL invoked by traffic-filter, permit and deny indicate the actions of permitting or denying traffic.

Summary

- This course describes PBR application scenarios, PBR classification, as well as how to configure local PBR and interface PBR.
- MQC supports multiple flexible traffic matching modes and executable actions, and can be invoked in multiple views. MQC can be used to implement PBR and traffic filtering.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Introduction to Typical Campus Network Technologies

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- The networks that are constructed in scenarios such as factories, government agencies, shopping malls, office buildings, school campuses, and parks to implement network interconnection are all called campus networks. Campus networks vary depending on industry attributes. However, a campus network is typically divided into the egress layer, core layer, aggregation layer, and access layer.
- This course describes the overall architecture of a campus network, common technologies and protocols used at each layer of the network, and typical networking scenarios.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Other courses have illustrated various routing protocols, which are not provided here.

Objectives

- Upon completion of this course, you will be able to:
  - Describe the typical architecture of a large-scale campus network.
  - Describe common Ethernet switching technologies.
  - Describe common WLAN application scenarios.
  - Describe common technologies used to ensure campus network reliability.
  - Describe common technologies used for campus network service and management.
  - Describe common technologies used to ensure campus network security.
  - Describe common VPN technologies used on campus networks.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Campus Network Architecture and Common Technologies**
2. Ethernet Switching Basics
3. Wireless Access
4. Network Reliability
5. Network Service and Management
6. Network Security
7. VPN
8. Multicast
9. IPv6

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Many Campuses Exist in a City

City residents work and live in campuses.
90%+

5h&22h
Daily smart terminal usage time & time in the campus every day



80%+
GDP is created in campuses.

90%+
Innovations are made in campuses.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Introduction to a Campus Network

- A campus network is an internal network of an enterprise or organization.
- A campus network is built to ensure that enterprise or organization services are running efficiently.
- Campus networks can be classified into the following types based on the scale:
  - Large-sized campus network: number of terminal users > 2000, number of NEs > 100
  - Medium-sized campus network: 2000 > number of terminal users > 200, 100 > number of NEs > 25
  - Small-sized campus network: number of terminal users < 200, number of NEs < 25
- Some enterprises have branches dispersed in different areas. Each branch network can be considered as a single campus network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

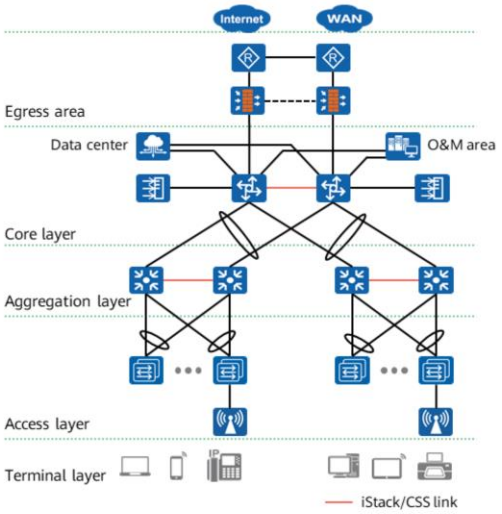
Common Industry Campus Networks

Enterprise campus network	School campus network	Government campus network	Entrepreneur campus network
			
• It focuses on network reliability and advancement, and aims to continuously improve employees' office experience and ensure the efficiency and quality of operation and production.	• School campus networks are classified into primary and secondary education and higher education campus networks. • Higher education campus networks are complex and include teaching and research networks, student networks, and operational dormitory networks. • There are high requirements on network manageability, security, and advancement.	• It usually refers to the internal network of a government agency. • High security is required. In most cases, the internal network and external network are isolated to ensure high security of confidential information.	• Such campus networks include networks of shopping malls, supermarkets, hotels, and parks. • The campus network is mainly used to serve consumers and includes the internal office network. • It provides Internet access services and deploys an intelligent service system to improve user experience, reduce O&M costs, improve business efficiency, and achieve value transfer.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

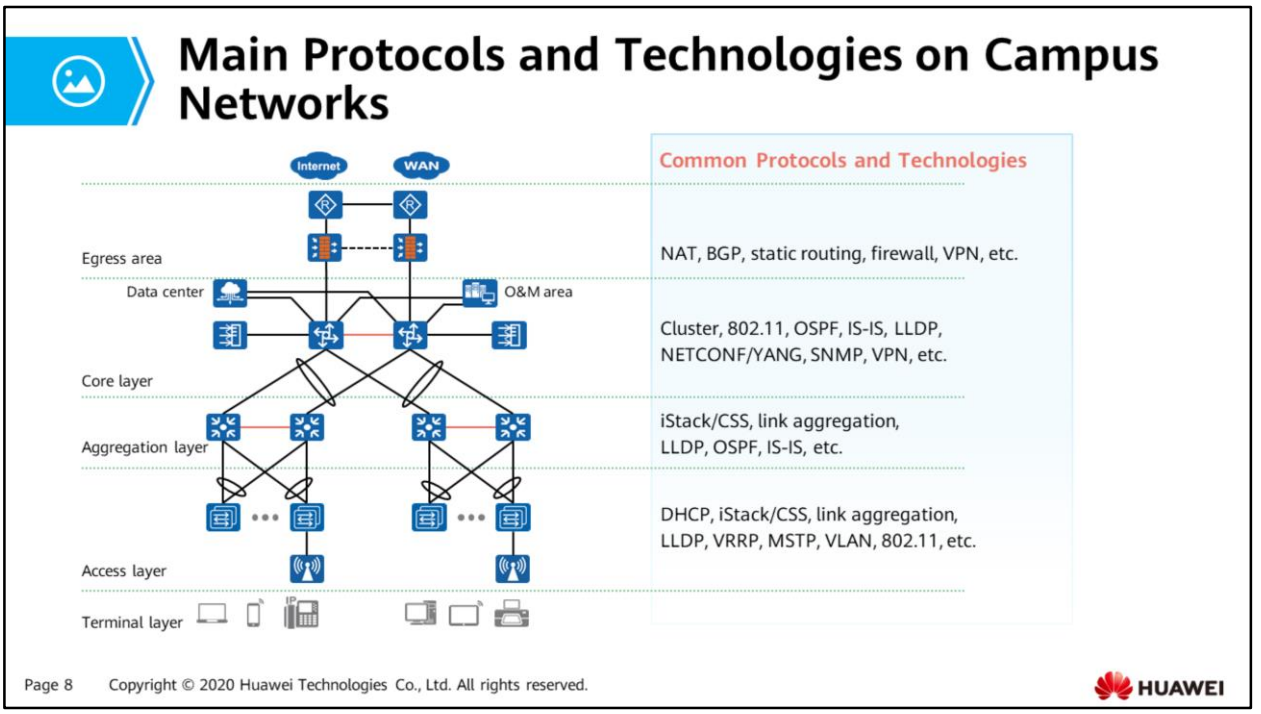
- To meet requirements of different industry campuses, the campus network architecture is designed based on the characteristics of the industry that the campus network serves. The campus network solution is based on industry attributes.

Typical Architecture of a Campus Network



- **Core layer**: is the backbone area of a campus network and is the data switching core. It connects various parts of the campus network, such as the data center, aggregation layer, and egress area.
- **Aggregation layer**: is located at the middle layer of a campus network to implement data aggregation or switching and provide key network functions, such as routing and security.
- **Access layer**: connects end users to a campus network and is the edge of a campus network.
- **Egress area**: is the edge between the campus internal network and the external network. Through this egress area, internal users can access the public network and external users (including customers, partners, branch users, and remote users) can access the internal network.
- **Data center area**: has servers and application systems deployed to provide data and application services for internal and external users of the enterprise.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



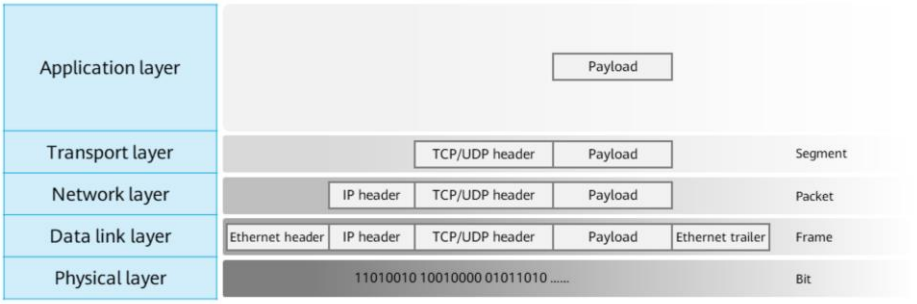
- NAT: Network Address Translation
- The Link Layer Discovery Protocol (LLDP) is a Layer 2 discovery protocol defined in the IEEE 802.1ab standard. Using LLDP, the NMS can rapidly obtain the Layer 2 network topology and topology changes when the network scale expands.
- Network Configuration Protocol (NETCONF) is a communication management protocol for NEs. It uses Extensible Markup Language (XML) for configuration data and protocol messages, allowing you to install, operate, and delete NEs.
- Yet Another Next Generation (YANG) is a data modeling language for data sent using NETCONF. It can be used to model configuration and status data of NEs.
- SNMP: Simple Network Management Protocol
- VRRP: Virtual Router Redundancy Protocol
- MSTP: Multiple Spanning Tree Protocol

Contents

1. Campus Network Architecture and Common Technologies
2. **Ethernet Switching Basics**
3. Wireless Access
4. Network Reliability
5. Network Service and Management
6. Network Security
7. VPN
8. Multicast
9. IPv6

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

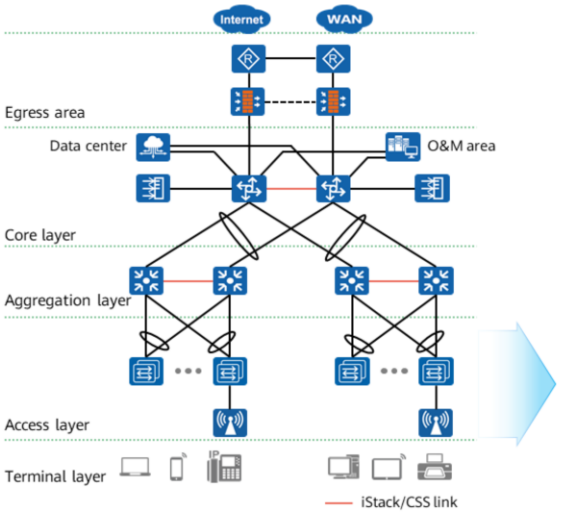
TCP/IP Model



- Dividing network communication into small components facilitates development, design, and troubleshooting of each component.
- Network components are standardized so that they can be developed by multiple vendors.
- The functions of each layer can be defined to impel industry standardization.
- Communication between network hardware and software of different types is allowed.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

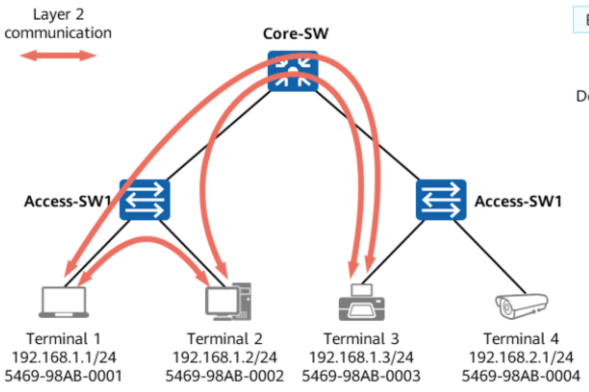
Campus Network and Ethernet Layer 2 Switching

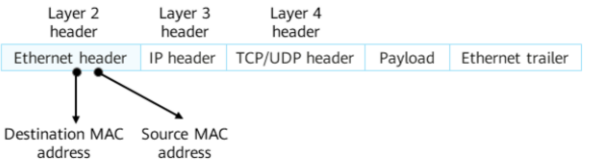


- The access layer provides various access modes for users and is the first network layer to which terminals connect.
- The access layer is usually composed of access switches. There are a large number of access switches that are sparsely distributed in different places on the network. In most cases, an access switch is a simple Layer 2 switch.
- If wireless terminals are present at the terminal layer, wireless access points (APs) need to be deployed at the access layer and access the network through access switches.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Overview of Layer 2 Switching



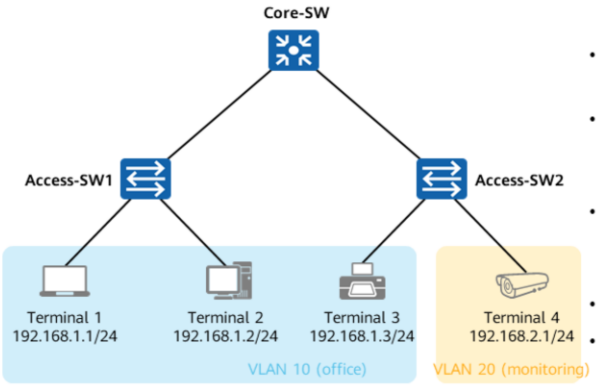


- Layer 2 switching is a basic function of Ethernet switches.
- In Layer 2 switching, a switch forwards a frame according to the destination MAC address in the Layer 2 header of the frame.
- Each switch maintains a MAC address table to guide data frame forwarding.
- When receiving a data frame, the switch queries the destination MAC address of the frame in the MAC address table and performs the corresponding operation according to the matched entry. When a switch receives a data frame, it also learns the source MAC address of the frame.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- A Layer 2 device works at the second layer of the OSI model and forwards data packets based on MAC addresses.
- A Layer 2 device parses and learns source MAC addresses of Ethernet frames and maintains a mapping table of MAC addresses and interfaces. This table is called a MAC address table. When receiving an Ethernet frame, the device searches for the destination MAC address of the frame in the MAC table to determine the interface to which the frame is forwarded.
- Interfaces on a Layer 2 device send and receive data independently and belong to different collision domains. Collision domains are isolated at the physical layer so that collisions will not occur between hosts (or networks) connected through this Layer 2 device due to uneven traffic rates on these hosts (or networks).

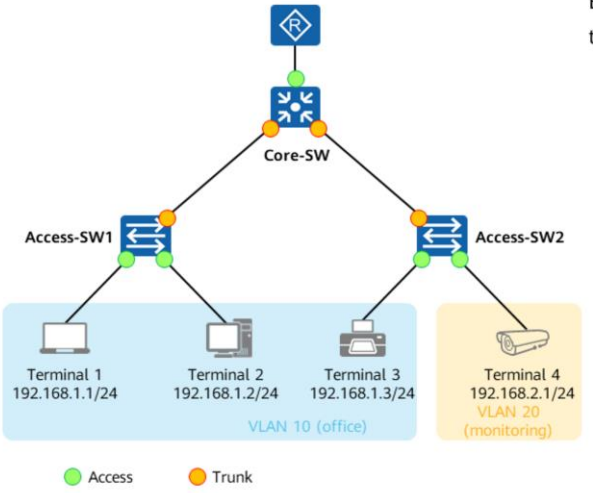
VLAN



- Virtual Local Area Network (VLAN) technology logically divides a physical LAN into multiple broadcast domains, each of which is called a VLAN.
- All devices in a VLAN belong to the same broadcast domain. Different VLANs belong to different broadcast domains.
- Devices in a VLAN can directly communicate with each other, whereas devices in different VLANs cannot communicate directly.
- Services in different VLANs are isolated from each other. Devices in different VLANs communicate with each other through Layer 3 devices.
- Generally, a VLAN is a logical subnet.
- Members of a VLAN are allocated based on interfaces of a switch. Generally, an interface of a switch is added to a specified VLAN. In this case, the device connected to the interface is added to the VLAN.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Layer 2 Ethernet Interface Types

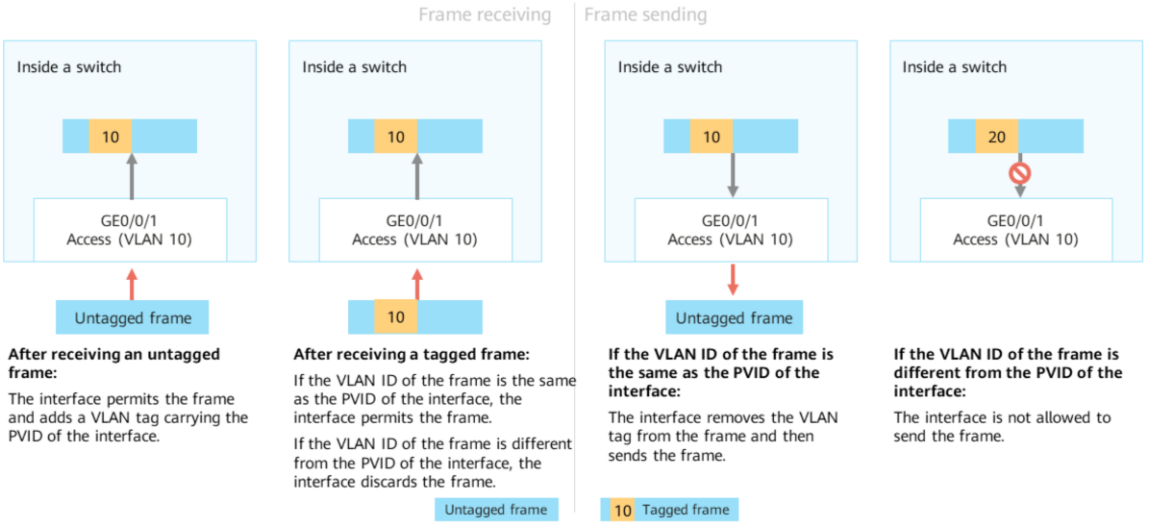


Ethernet Layer 2 interfaces on switches are classified into the following types:

- An access interface is used to connect a switch to a terminal, such as a PC or server. Typically, the NICs on such a terminal receive and send only untagged frames. An access interface can be added to only one VLAN.
- A trunk interface allows frames that belong to multiple VLANs to pass through and differentiates the frames using the 802.1Q tag. This type of interface is used to connect a switch to another switch or a sub-interface on a device, such as a router or firewall.
- Similar to a trunk interface, a hybrid interface also allows frames that belong to multiple VLANs to pass through and differentiates the frames using the 802.1Q tag. You can determine whether to allow a hybrid interface to carry VLAN tags when sending the frames of one or more VLANs.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Access Interface

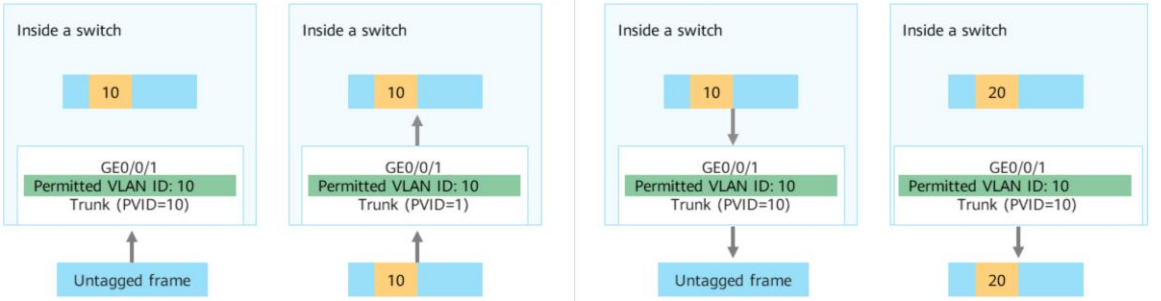


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- All Layer 2 interfaces have a default VLAN ID, which is called Port Default VLAN ID (PVID). On Huawei switches, the default PVID is 1. In addition, all data frames carry tags inside the switch to improve the processing efficiency of data frames.

Trunk Interface

Frame receiving | Frame sending



After receiving an untagged frame:

The interface adds a VLAN tag with the VID being the PVID of the interface to the frame and permits the frame only when the PVID is in the list of VLAN IDs permitted by the interface. If the PVID is not in the list, the interface discards the frame.

After receiving a tagged frame:

If the VLAN ID of the frame is in the list of VLAN IDs permitted by the interface, the interface permits the frame. Otherwise, the interface discards the frame.

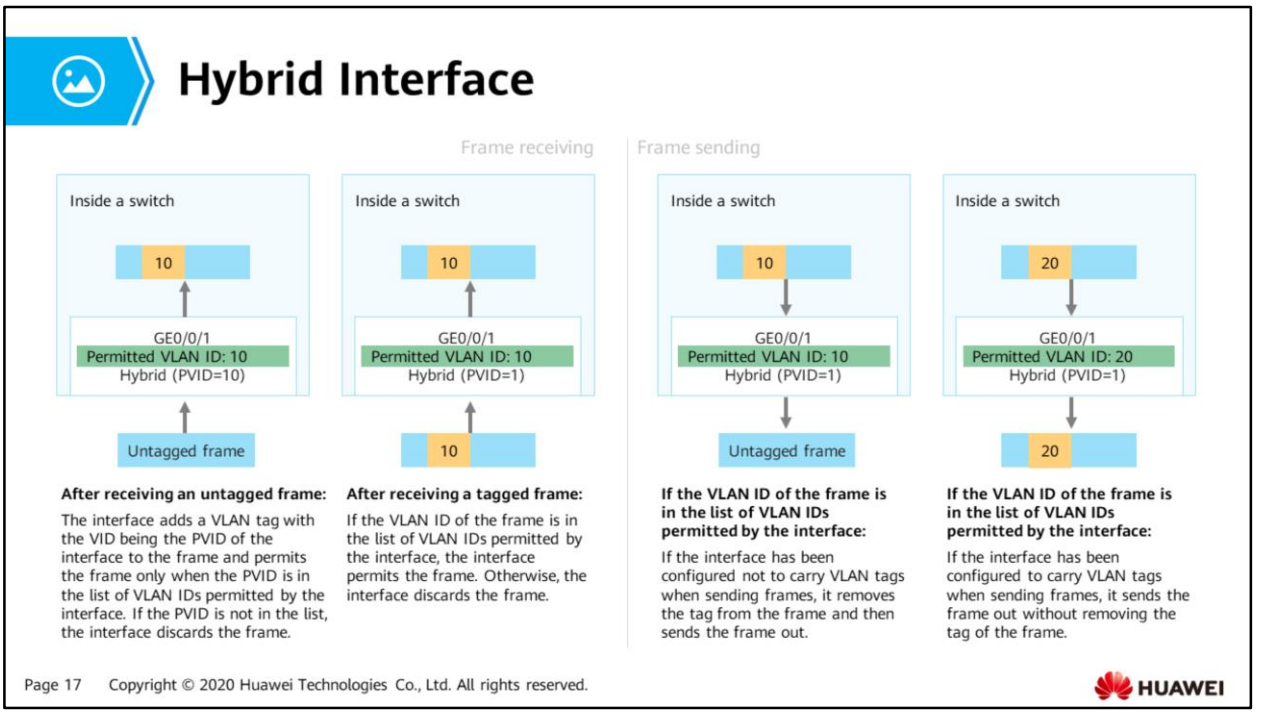
If the VLAN ID of the frame is the same as the PVID of the interface:

If the VLAN ID of the frame is in the list of VLAN IDs permitted by the interface, the interface removes the tag from the frame and sends the frame out. Otherwise, the interface is not allowed to send the frame.

If the VLAN ID of the frame is different from the PVID of the interface:

If the VLAN ID of the frame is in the list of VLAN IDs permitted by the interface, the interface sends the frame out without removing the tag of the frame. Otherwise, the interface is not allowed to send the frame.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



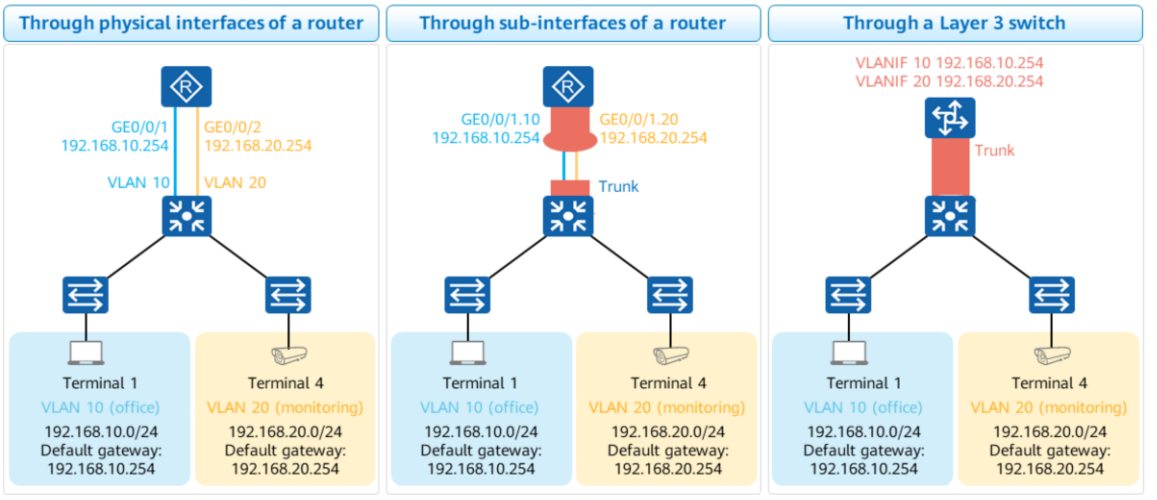
- A hybrid interface can transmit data of multiple VLANs. The behavior of a hybrid interface is similar to that of a trunk interface in receiving data frames. When a trunk interface sends a data frame, the switch removes the tag of the data frame only when the VLAN ID of the data frame is the same as the PVID of the interface. In addition, data frames of other VLANs sent by the interface carry tags. A hybrid interface sends data frames in a different way from a trunk interface. You can run commands to configure a hybrid interface to send untagged data frames of a certain VLAN or some VLANs.

VLAN Assignment Methods

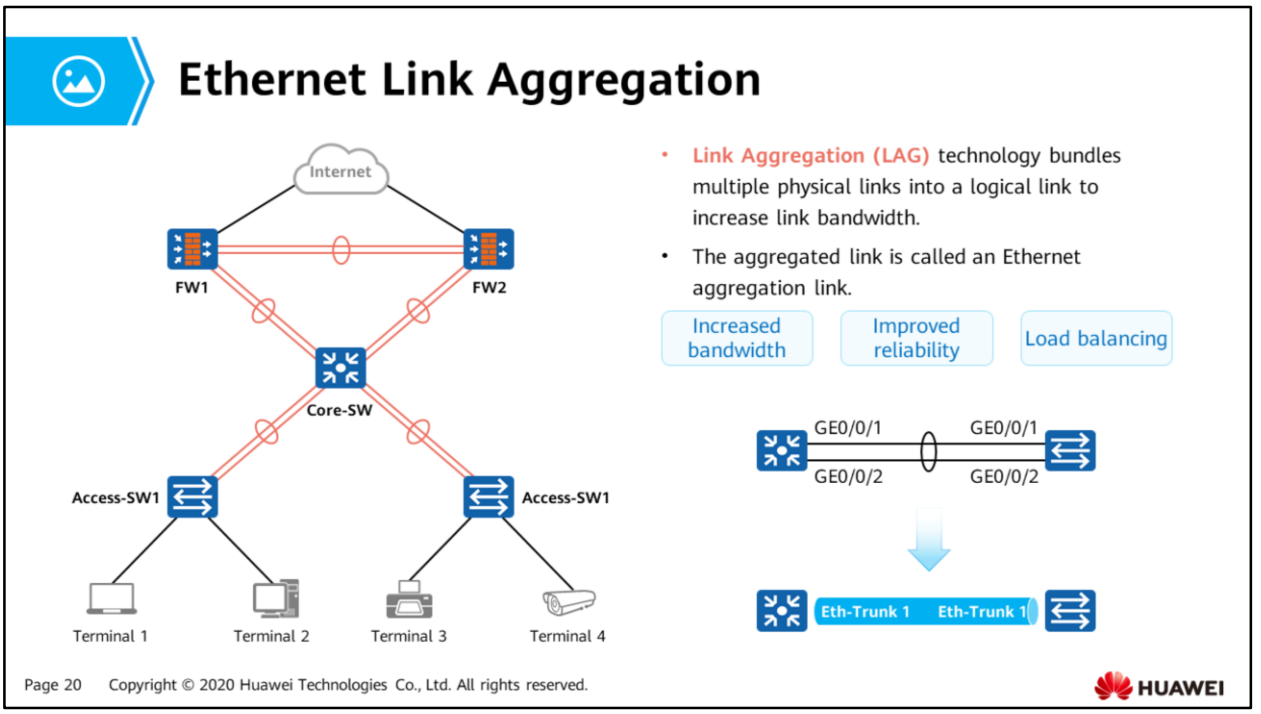
VLAN Assignment Method	Description
Interface-based assignment	VLANs are assigned based on switch interfaces.
MAC address-based assignment	VLANs are assigned based on the source MAC addresses of frames.
IP subnet-based assignment	VLANs are assigned based on the source IP addresses of frames.
Protocol-based assignment	VLANs are assigned based on protocol types and encapsulation formats of frames.
Policy-based assignment	VLANs are assigned based on policies such as combinations of interfaces, MAC addresses, and IP addresses.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Inter-VLAN Communication

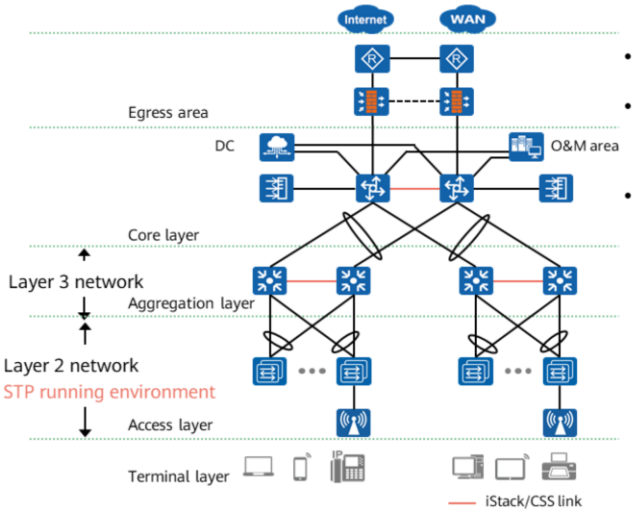


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- As networks grow in scale, users require Ethernet backbone networks to provide higher bandwidth and availability. In the past, the only way to increase bandwidth was to upgrade the network with high-speed LPUs, which is costly and inflexible.
- In contrast, link aggregation increases bandwidth by bundling a group of physical port into a single logical port, without the need to upgrade hardware. In addition, link aggregation provides link backup mechanisms, greatly improving link availability.
- An LAG is the logical link bundled by many Ethernet links, and is short for Eth-Trunk. Each LAG corresponds to a unique logical interface, which is called an aggregation interface or Eth-Trunk interface.
- Link aggregation has the following advantages:
    - Improved bandwidth: The maximum bandwidth of a link aggregation group (LAG) is the combined bandwidth of all member links.
    - Improved reliability: If an active link fails, traffic can be switched to other available member links.
    - Load balancing: The traffic load can be balanced among the active member links in a LAG.

STP: Loop Prevention + Layer 2 Network Reliability



- Eliminates loops on a switching network.
- Dynamically adapts to network topology changes.
- Ensures Layer 2 network reliability together with redundant links.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

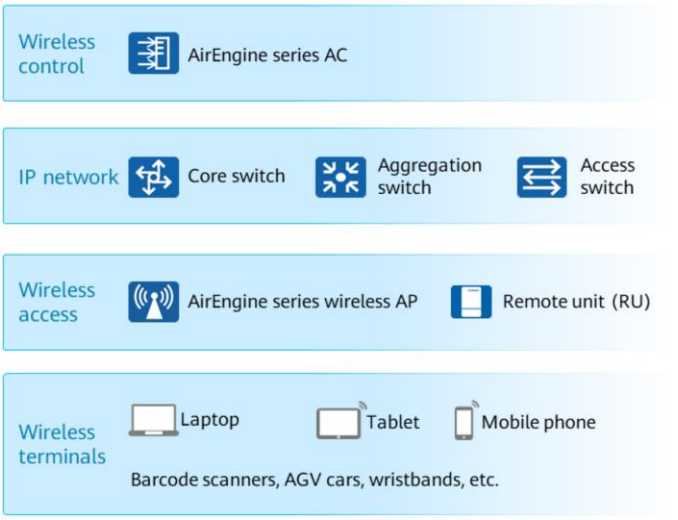
1. Campus Network Architecture and Common Technologies
2. Ethernet Switching Basics
3. **Wireless Access**
4. Network Reliability
5. Network Service and Management
6. Network Security
7. VPN
8. Multicast
9. IPv6

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

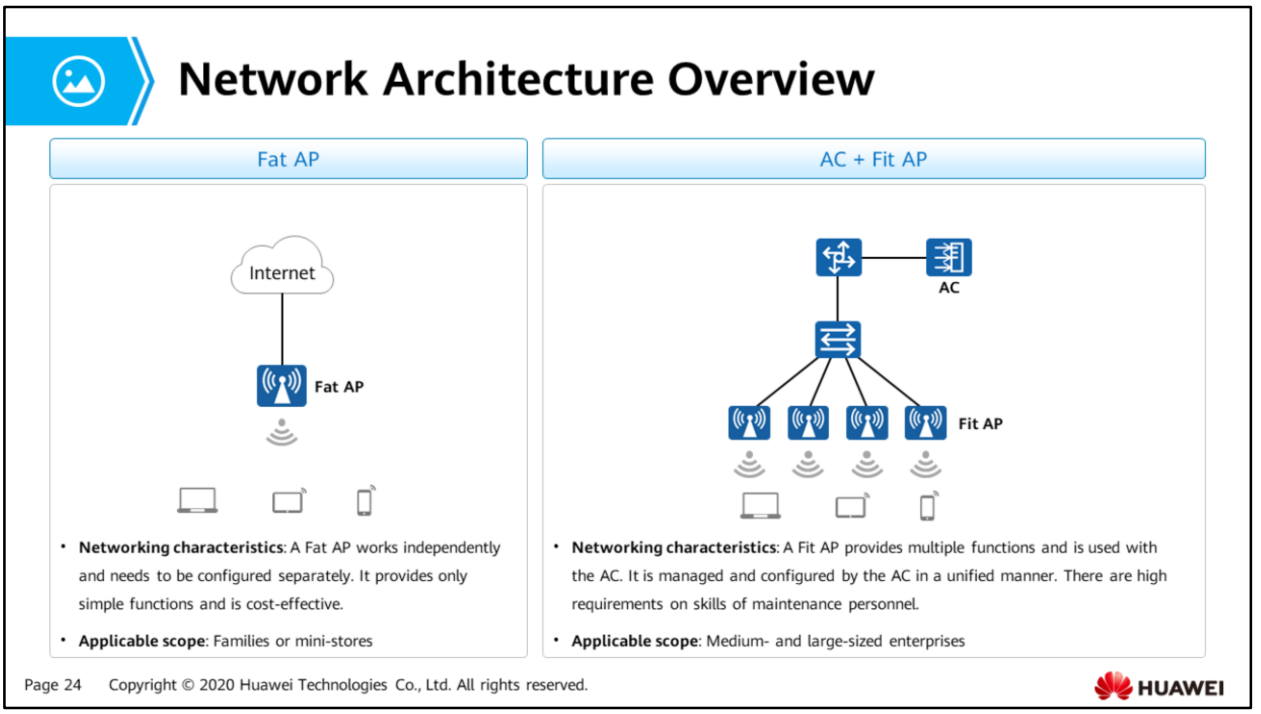
WLAN and Main NEs

Overview of WLAN

- WLAN is a network that uses high-frequency (2.4 GHz or 5 GHz) signals such as radio waves, lasers, and infrared rays to replace the traditional media used for transmission on a wired LAN.
- The WLAN technologies described in this document are based on 802.11 standards.
- 802.11 was originally a wireless LAN communications standard defined by the Institute of Electrical and Electronics Engineers (IEEE) in 1997.
- The IEEE then made amendments to the standard, forming the 802.11 family, including 802.11, 802.11a, 802.11b, 802.11e, 802.11g, 802.11i, 802.11n, and 802.11ac.

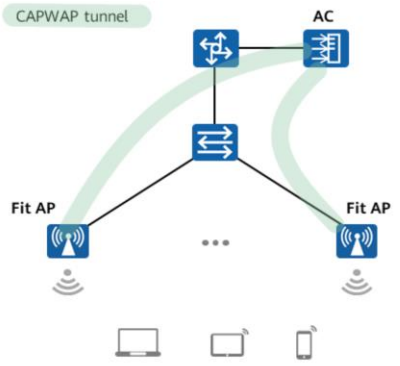


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI



- AP
  - The AP can switch flexibly among the Fat, Fit, and cloud modes based on the network plan.
  - Fat AP: applies to home WLANs. A Fat AP works independently and requires separate configurations. It provides only simple functions and is cost-effective. The Fat AP independently implements functions such as user access, authentication, data security, service forwarding, and QoS.
  - Fit AP: applies to medium- and large-sized enterprises. Fit APs are managed and configured by the AC in a unified manner, provide various functions, and have high requirements on network maintenance personnel's skills. Fit APs must work with a AC for user access, AP going-online, authentication, routing, AP management, security, and QoS.
  - Cloud AP: applies to small- and medium-sized enterprises. Cloud APs are managed and configured by a cloud management platform in a unified manner, provide various functions, support plug-and-play, and have low requirements on network maintenance personnel's skills.
- AC
  - An AC is usually deployed at the aggregation layer of a network to provide high-speed, secure, and reliable WLAN services.
  - Huawei ACs provide a large capacity and high performance. They are highly reliable, easy to install and maintain, and feature such advantages as flexible networking and energy conservation.

AC + Fit AP Architecture (1)



Basic Concepts

- **AC**: controls and manages all Fit APs on a AC + Fit AP network.

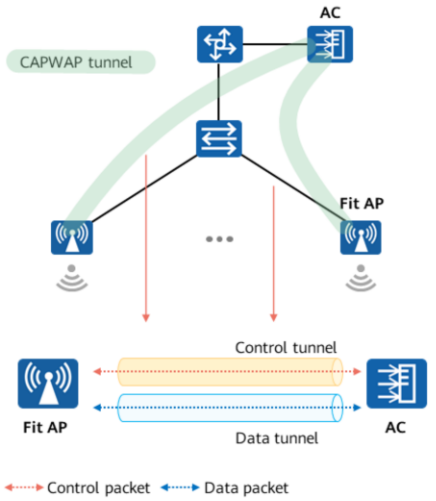
- The AC is responsible for WLAN access control, forwarding and statistics, AP configuration monitoring, roaming management, AP network management agent, and security control.
- A Fit AP encrypts and decrypts 802.11 packets, provides 802.11 physical layer functions, collects air interface statistics, and is managed by an AC.
- The AC and Fit APs communicate through CAPWAP.
- Compared with the Fat AP architecture, the AC + Fit AP architecture has the following advantages:
  - **Easier configuration and deployment**
  - **Higher security**
  - **Easier update and expansion**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The AC and Fit APs communicate through CAPWAP. With CAPWAP, APs automatically discover the AC, the AC authenticates the APs, and the APs obtain the software package and the initial and dynamic configurations from the AC. CAPWAP tunnels are established between the AC and APs. CAPWAP tunnels include control and data tunnels. The control tunnel is used to transmit control packets (also called management packets, which are used by the AC to manage and control APs). The data tunnel is used to transmit data packets. The CAPWAP tunnels allow for Datagram Transport Layer Security (DTLS) encryption, so that transmitted packets are more secure.
- Compared with the Fat AP architecture, the AC + Fit AP architecture has the following advantages:
  - Configuration and deployment: The AC centrally configures and manages the wireless network so that you do not need to configure each AP separately. In addition, the channels and power of APs on the entire network are automatically adjusted, eliminating the need for manual adjustment.

AC + Fit AP Architecture (2)



Overview of CAPWAP

- **Control And Provisioning of Wireless Access Points (CAPWAP)**: defines how to manage and configure APs. That is, a AC manages and controls APs in a centralized manner through CAPWAP tunnels.
- With CAPWAP, APs automatically discover the AC, the AC authenticates the APs, and the APs obtain the software package and the initial and dynamic configurations from the AC. CAPWAP tunnels are established between the AC and APs.

CAPWAP Tunnel Functions

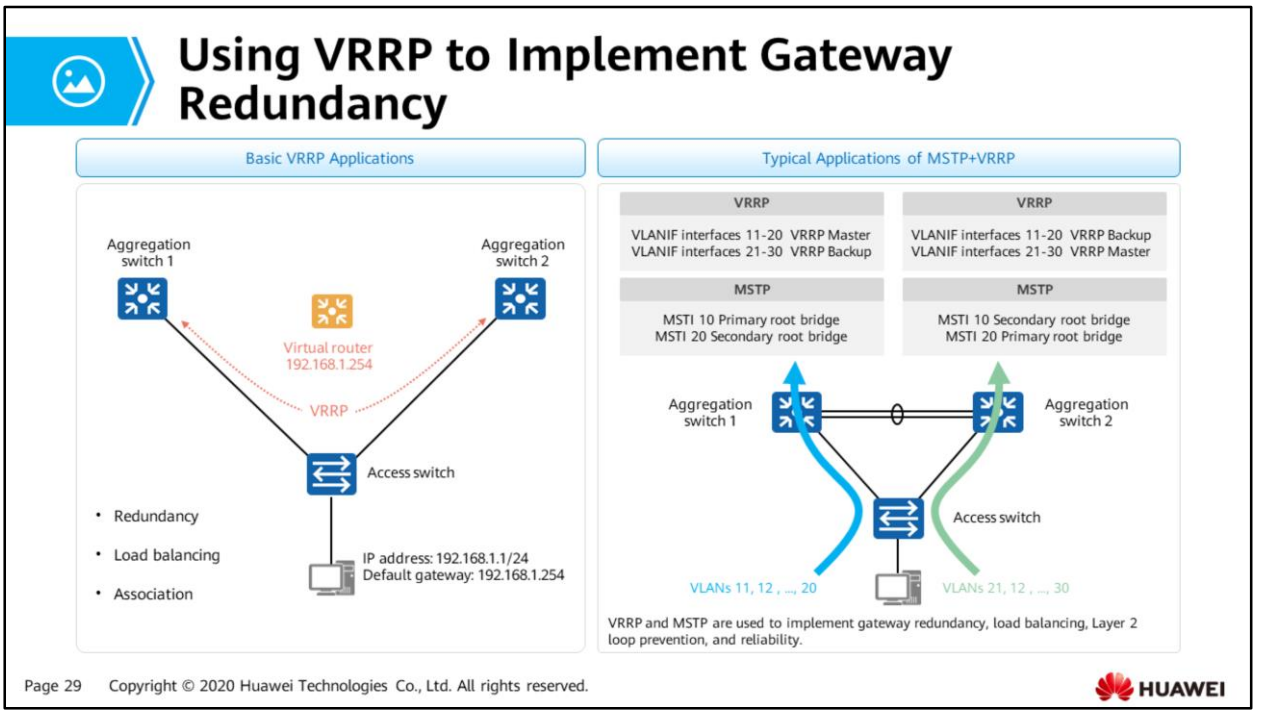
- CAPWAP tunnels include control and data tunnels.
- The control tunnel is used to transmit control packets (also called management packets, which are used by the AC to manage and control APs). The data tunnel is used to transmit data packets. The CAPWAP tunnels allow for Datagram Transport Layer Security (DTLS) encryption, so that transmitted packets are more secure. Allows APs to exchange data sent by STAs with the AC through CAPWAP tunnels when the tunnel forwarding mode is used.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

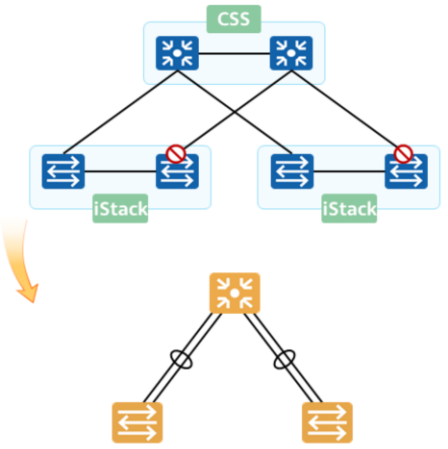
1. Campus Network Architecture and Common Technologies
2. Ethernet Switching Basics
3. Wireless Access
4. **Network Reliability**
5. Network Service and Management
6. Network Security
7. VPN
8. Multicast
9. IPv6

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- The Virtual Router Redundancy Protocol (VRRP) specifies an election protocol that dynamically assigns responsibility for a virtual router to VRRP routers on a LAN. It allows several routers on a subnet to use the same virtual IP address, with the physical routers representing a virtual logical router. If a gateway fails, VRRP selects a different gateway to forward traffic, thereby ensuring reliable communication.
- Generally, all hosts on the same network segment are configured with the same default route with the gateway address as the next hop address. The hosts use the default route to send packets to the gateway and the gateway forwards the packets to other network segments. When the gateway fails, hosts with the same default route cannot communicate with external networks. To improve network reliability, multiple egress gateways can be configured. However, route selection between the gateways becomes an issue.
- VRRP resolves this issue. The virtual router IP address is configured as the default gateway address. If a gateway fails, VRRP selects a different gateway to forward traffic, thereby ensuring reliable communication.
    - Redundancy: Multiple routing devices enabled with VRRP constitute a VRRP group and the VRRP group is used as the default gateway. When a single point of failure occurs, services are transmitted through the backup link. This reduces the possibility of network faults and ensures uninterrupted transmission of various services.

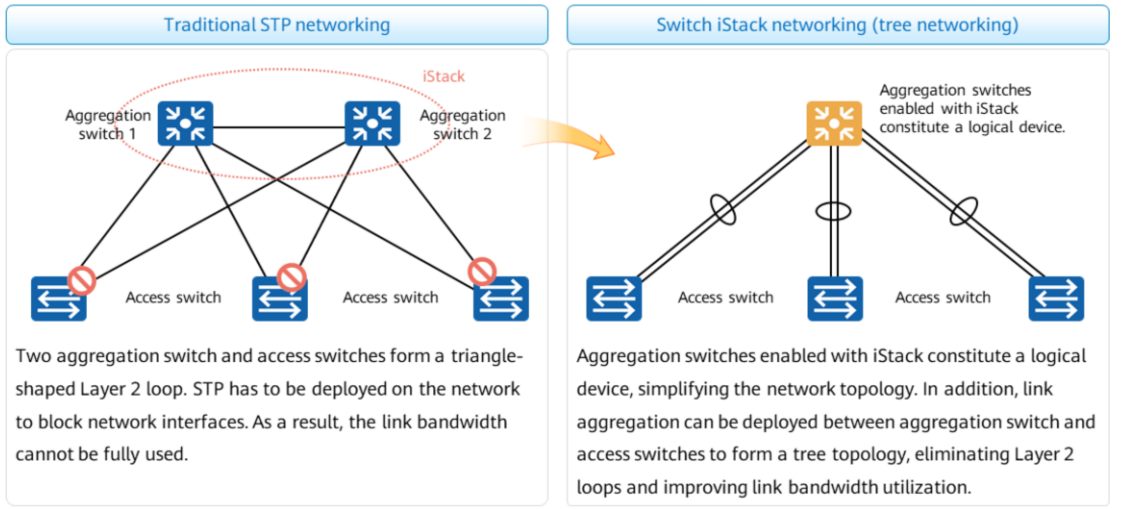
Overview of CSS/iStack



- Intelligent stack (iStack) enables multiple stacking-capable switches to function as a single logical switch. iStack is applicable to Huawei fixed switches.
- A cluster switch system (CSS) combines two clustering-capable switches into a single logical switch. A CSS is also called a cluster. CSS is applicable to Huawei modular switches.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Traditional STP Networking and iStack Networking



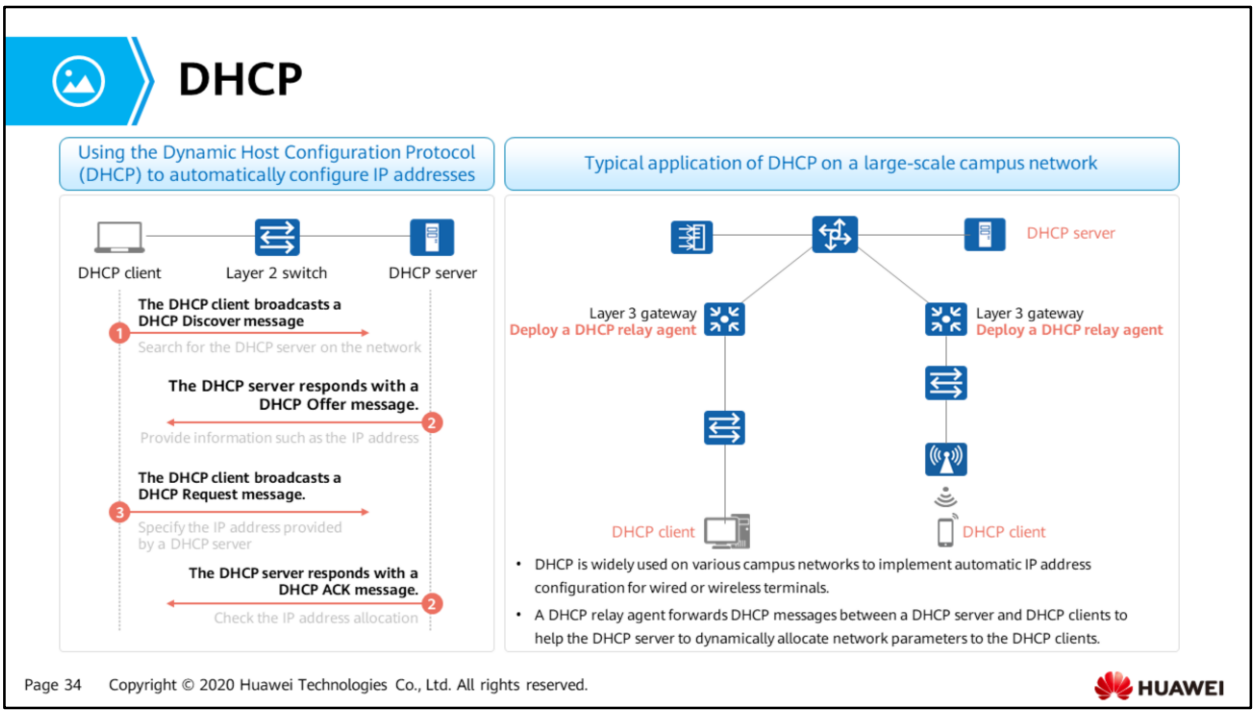
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- iStack enables multiple stacking-capable switches to function as a logical device.
- Before a stack is set up, each switch is independent and has its own IP address and MAC address. You need to manage the switches separately. After a stack is set up, switches in the stack form a logical entity and can be managed and maintained using a single IP address. iStack technology improves forwarding performance and network reliability, and simplifies network management.

Contents

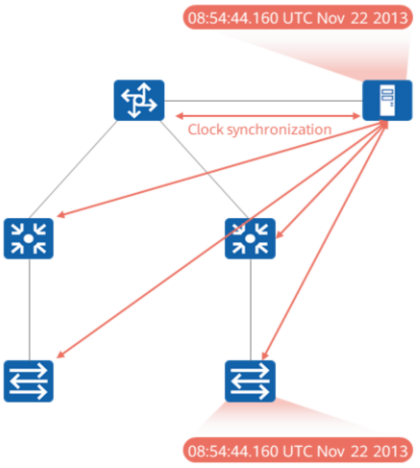
1. Campus Network Architecture and Common Technologies
2. Ethernet Switching Basics
3. Wireless Access
4. Network Reliability
5. **Network Service and Management**
6. Network Security
7. VPN
8. Multicast
9. IPv6

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- DHCP dynamically configures and uniformly manages IP addresses of hosts. It simplifies network deployment and scale-out, even for small networks.
- DHCP enables a host to obtain an IP address dynamically, but does not specify an IP address for each host.
- DHCP can allocate other configuration parameters, such as the boot file of a client, so that the client can obtain all the required configuration information by using only one message.
- DHCP is defined in RFC 2131 and uses the client/server communication mode. A DHCP client requests configuration information from a DHCP server, and the DHCP server returns the configuration information allocated to the DHCP client.
- DHCP supports dynamic and static IP address allocation.
    - Dynamic allocation: DHCP allocates an IP address with a limited validity period (known as a lease) to a client. This mechanism applies to scenarios where hosts temporarily access the network and the number of idle IP addresses is less than the total number of hosts.
    - Static allocation: DHCP allocates fixed IP addresses to clients as configured. Compared with manual IP address configuration, DHCP static allocation prevents manual configuration errors and enables unified maintenance and management.
- DHCPv4 offers the following benefits:
    - Reduced client configuration and maintenance costs
    - Centralized management

NTP

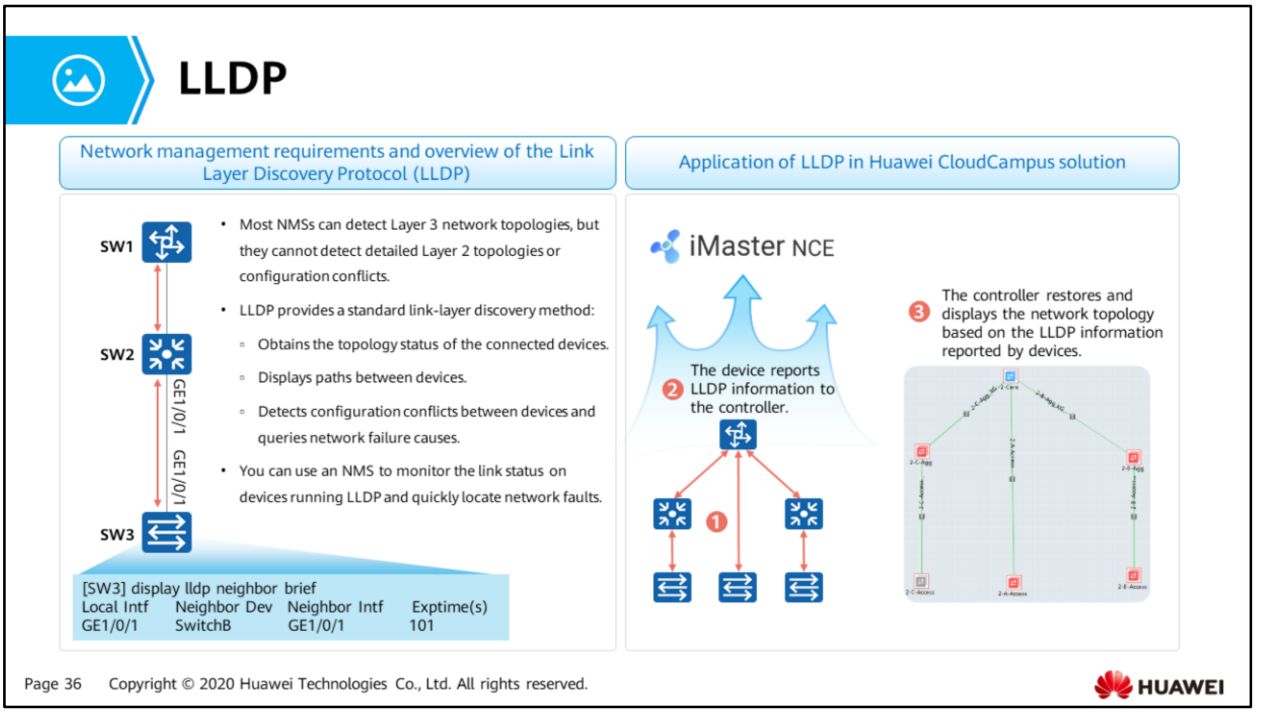


The Network Time Protocol (NTP) is mainly used in the following scenarios where the clocks of all network devices must be consistent:

- **Network management**: Analysis on logs or debugging information collected from different routers must be performed based on time.
- **Charging system**: The clocks of all devices must be consistent.
- **Several systems interworking on the same complex event**: The systems must use the same clock for reference to ensure proper sequencing of operations.
- **Incremental backup between the backup server and clients**: Clocks on the backup server and clients should be synchronized.
- **System time**: Some applications need to know the time when user logs in to the system and the file revision time.

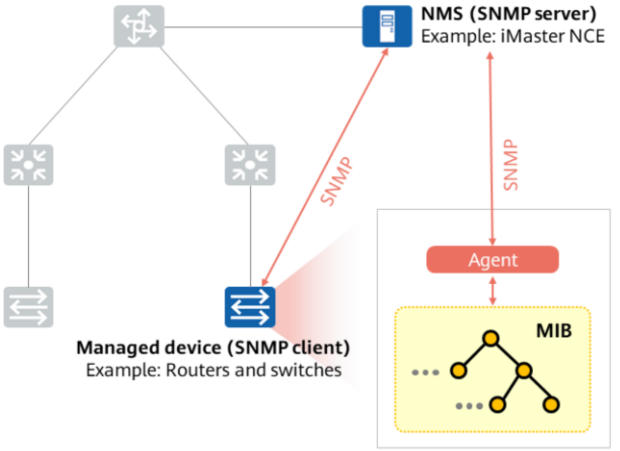
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- NTP is an application layer protocol belonging to the Transmission Control Protocol/Internet Protocol (TCP/IP) suite. NTP synchronizes time between time servers and clients. NTP implementation is based on IP and User Datagram Protocol (UDP). NTP packets are transmitted by UDP over port 123.
- As network topologies become increasingly complex, clock synchronization becomes more important for all devices within a network. Manual configuration of system clocks by network administrators is both labor-intensive and error-prone, potentially affecting clock precision. To address this problem, NTP is introduced to synchronize the clocks of devices within a network.
- NTP is used when clocks of all devices on a network need to be consistent. For example, in network management, the logs and debugging information collected from different routers need to be analyzed based on time.
    - Charging system: The clocks of all devices must be consistent.
    - Several systems interworking on the same complex event: The systems must use the same clock for reference to ensure proper sequencing of operations.
    - Incremental backup between the backup server and clients: Clocks on the backup server and clients should be synchronized.
    - System time: Some applications need to know the time when user logs in to the system and the file revision time.
- A switch can function as both an NTP server and an NTP client.



- LLDP is a standard Layer 2 topology discovery protocol defined in IEEE 802.1ab. LLDP collects local device information including the management IP address, device ID, and port ID and advertises the information to neighbors. Neighbors save the received information in their management information bases (MIBs). The NMS can use data in MIBs to query the link status.
- An NMS must be capable of managing multiple network devices with diverse functions and complex configurations. Most NMSs can detect Layer 3 network topologies, but they cannot detect detailed Layer 2 topologies or configuration conflicts. A standard protocol is required to exchange Layer 2 information between network devices.
- LLDP provides a standard link-layer discovery method. Layer 2 information obtained from LLDP allows the NMS to detect the topology of neighboring devices, and display paths between clients, switches, routers, application servers, and network servers. The NMS can also detect configuration conflicts between network devices and identify causes of network failures. Enterprise users can use an NMS to monitor the link status on devices running LLDP and quickly locate network faults.

SNMP

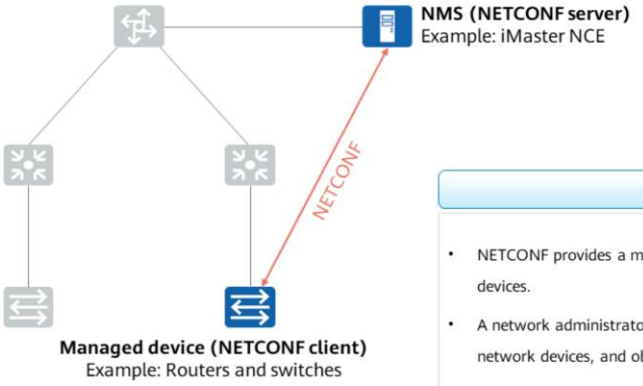


Overview of SNMP

- The Simple Network Management Protocol (SNMP) is a network management standard widely used on TCP/IP networks.
- SNMP uses a central computer (a network management station) that runs network management software to manage devices.
- By employing the "network management over networks" mode, SNMP is used for efficient and batch network device management. In addition, SNMP enables unified management of network devices of different types and from different vendors.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

NETCONF/YANG



SNMP Weaknesses

SNMP is not a configuration-oriented protocol. On a large-sized network with a complex topology, SNMP cannot meet network management requirements, especially configuration management requirements.

NETCONF

- NETCONF provides a mechanism for communication between the NMS and network devices.
- A network administrator can use NETCONF to add, modify, or delete configurations of network devices, and obtain configurations and status of network devices.
- NETCONF is based on Extensible Markup Language (XML).

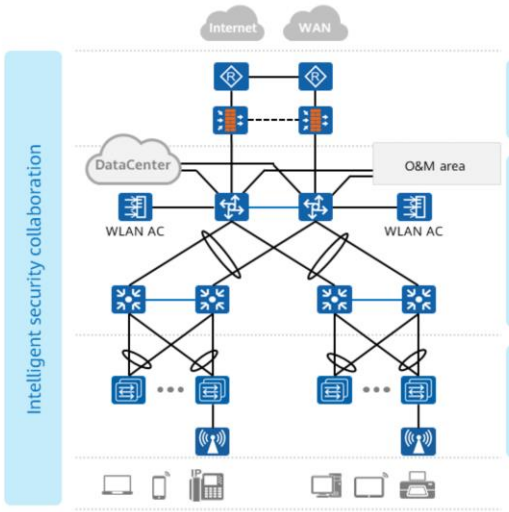
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Campus Network Architecture and Common Technologies
2. Ethernet Switching Basics
3. Wireless Access
4. Network Reliability
5. Network Service and Management
6. **Network Security**
7. VPN
8. Multicast
9. IPv6

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Overview of Network Security



Campus border

- Deploy firewalls to divide security zones, isolate services, and implement security management and control.
- Deploy anti-DDoS devices to defend against DDoS attacks.
- Deploy IPS devices to provide intrusion detection and security posture awareness.

Core and aggregation layer

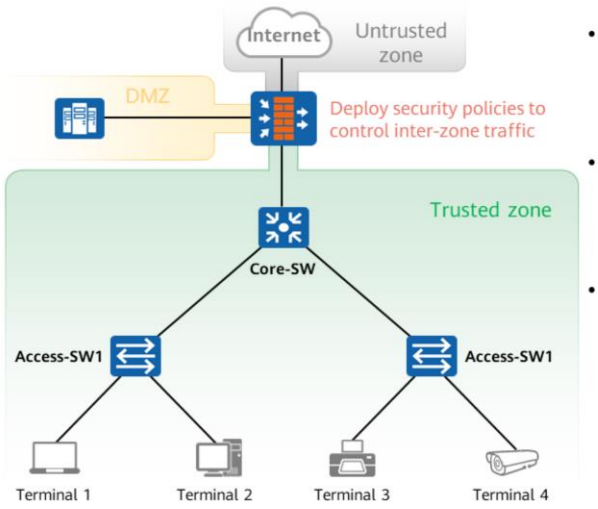
- Deploy service isolation and mutual access traffic control.
- Deploy network admission control on demand.
- Deploy local attack defense to improve anti-attack capabilities of devices.

Access layer

- It is recommended that security functions such as broadcast storm control, DHCP snooping, IPSG, and DAI be enabled and port isolation be deployed to enhance user communication security.
- Deploy network access control to authenticate identities or terminals of users who access the campus network and grant network access permissions based on the authentication results.
- Deploy wireless air interface security and wireless service security.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Firewall-based Security Zone Division and Security Policies

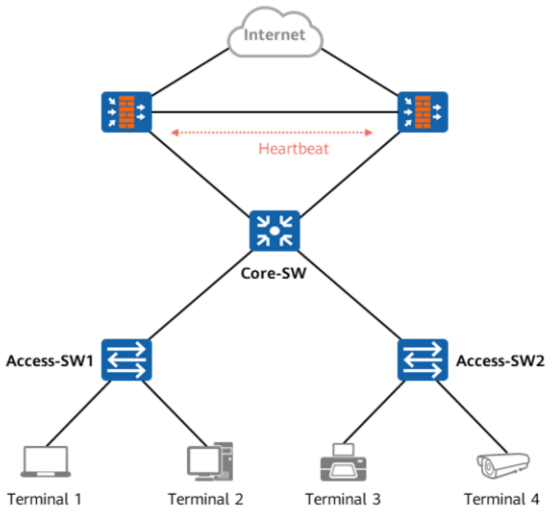


- A security zone is introduced on a device. Most security policies are implemented based on security zones.
- A security zone is a set of the networks connected by interfaces. Users on these networks have the same security attributes.
- Employees, servers, and external networks can be assigned to different security zones for inter-zone traffic inspection and protection.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Two security zones with the same security level cannot be created on a firewall.
- Interfaces on a firewall must be added to a security zone. Otherwise, traffic cannot be forwarded properly.
- An interface on a firewall can belong to a security zone.
- A security zone of a firewall can have multiple interfaces.
- The default security zones of the system cannot be deleted. You can create a user-defined security zone as required.

Overview of Firewall HSB



Requirement

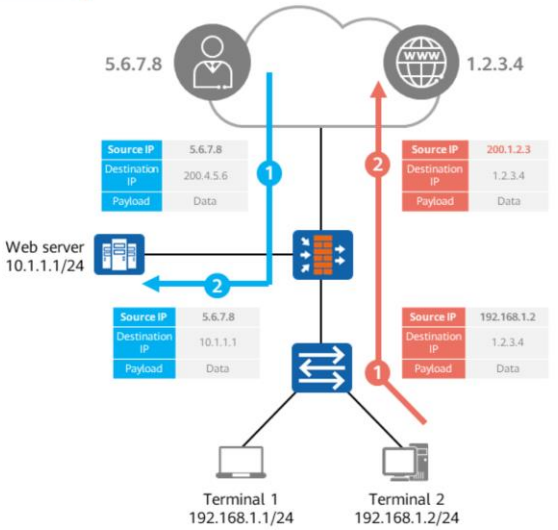
- When one firewall is deployed at the egress of the network and a fault occurs on it, services on the entire network will be adversely affected. Network reliability needs to be improved.
- Multiple firewalls can be deployed to improve reliability. Service continuity must be ensured during device switchover.

Solution

- Two firewalls are deployed and work in hot standby (HSB) mode.
- HSB requires two firewalls with the same hardware and software configurations.
- The two firewalls are connected through an independent link, which is usually called a heartbeat link. The two firewalls learn the health status of each other and back up configurations and entries (such as session table entries and IPsec SAs) to each other through the heartbeat link.
- If one firewall fails, services can be smoothly switched over to the other firewall to ensure service continuity.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

NAT



Requirement

- When an intranet user (using a private IP address) needs to access the Internet, the source IP address needs to be translated into a public IP address.
- Extranet users need to access the server (such as the web server) that uses a private IP address.

NAT

- NAT translates between private and public IP addresses carried in the headers of IP data packets.
- Common NAT:
  - Source IP address-based NAT:
    - No-PAT
    - Network address and port translation (NAPT)
  - Destination IP address-based NAT:
    - NAT server

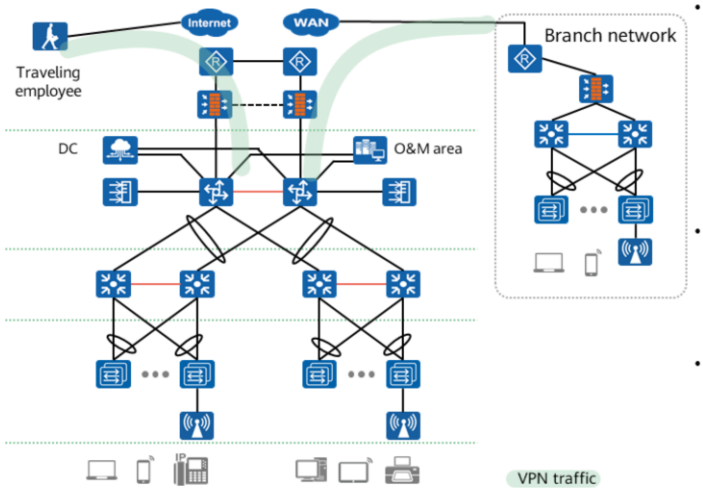
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Campus Network Architecture and Common Technologies
2. Ethernet Switching Basics
3. Wireless Access
4. Network Reliability
5. Network Service and Management
6. Network Security
7. **VPN**
8. Multicast
9. IPv6

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VPN Technology Application Scenario



- For a large-scale enterprise, network access requirements are not limited to the headquarters network. Branches, traveling employees, and partners also need to access network resources of the headquarters. Generally, Virtual Private Network (VPN) technology is used to meet this requirement.
- VPN technology builds a virtual private network without any change in the network topology. Compared with leased lines, VPN is a more cost-effective solution to build a private network.
- Different VPN technologies have different features and implementation modes. Common VPN technologies include IPsec VPN, GRE VPN, L2TP VPN, and MPLS VPN.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

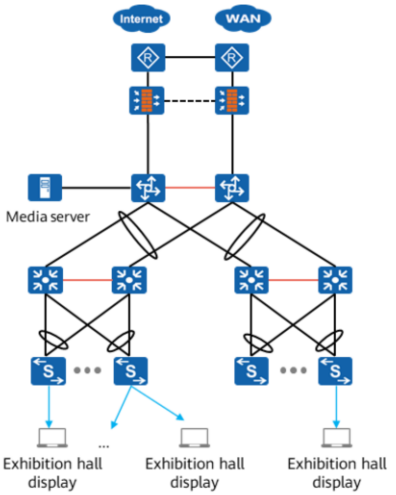
- Internet Protocol Security (IPsec)
- Generic Routing Encapsulation (GRE)
- Layer 2 Tunneling Protocol (L2TP)
- Multiprotocol Label Switching (MPLS)

Contents

1. Campus Network Architecture and Common Technologies
2. Ethernet Switching Basics
3. Wireless Access
4. Network Reliability
5. Network Service and Management
6. Network Security
7. VPN
8. **Multicast**
9. IPv6

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Multicast Application Scenarios



- An enterprise has some bulletins, such as the weather, duty schedule, precautions in the equipment room, and promotion videos. To facilitate employees and visitors to obtain the information in a timely manner, display screens are usually deployed in densely populated areas.
- The content displayed on each screen is the same. This is a typical point-to-multipoint communication scenario. If information is transmitted in unicast mode, device performance and link bandwidth on the network are wasted to a certain extent.
- Multicast technology efficiently implements point-to-multipoint data transmission over an IP network, while conserving network bandwidth and reducing network loads.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

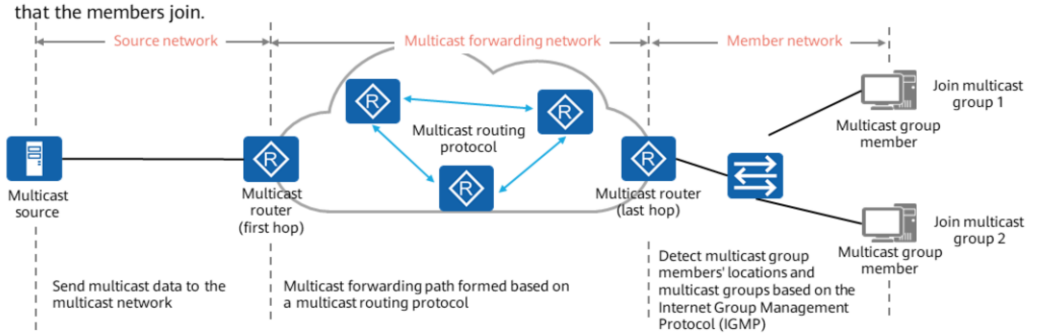
HUAWEI

- Unicast transmission is implemented between a source IP host and a destination IP host. Most of data is transmitted in unicast mode on a network. For example, email and online banking applications are implemented in unicast mode.
    - In unicast mode, each data packet has a specific destination IP address. For the same data, if there are multiple receivers, the server needs to send unicast data packets with the same number as the number of receivers. When there are hundreds or thousands of receivers, the server consumes a lot of resources to create the same data and send multiple copies of the same data. As a result, the device performance and link bandwidth on the network are wasted to a certain extent. The unicast mode is applicable to networks with a small number of users. When there are a large number of users, the unicast mode cannot ensure the network transmission quality.
- Broadcast transmission is implemented between a source IP host and all the other IP hosts on the local network. All hosts can receive data from the source host, regardless of whether they require the data.
    - Broadcast data packets are restricted in a broadcast domain. Once a device sends broadcast data, all devices in the broadcast domain receive the data packet and have to consume resources to process the data packet. A large number of broadcast data packets consume network bandwidth and device resources. The broadcast mode applies only to shared network segments, and cannot ensure information security and paid services.
- Multicast transmission is implemented between one source IP host and a group of IP hosts. Intermediate routers and switches selectively replicate and forward data based on demands of receivers.

Basic Multicast Network Architecture

A multicast network can be divided into three parts:

- Source network: sends multicast data generated by the multicast source to the multicast network.
- Multicast forwarding network: A loop-free multicast forwarding path is formed. This forwarding path is also called a multicast distribution tree (MDT).
- Member network: enables the multicast network to detect the locations of multicast group members and the multicast groups that the members join.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Multicast source: indicates a multicast traffic sender, for example, a multimedia server. A multicast source does not need to run any multicast protocol. It only needs to send multicast data.
- Multicast receiver: is also called a multicast group member and a device that expects to receive traffic of a specific multicast group, for example, a PC running the multimedia live broadcast client software.
- Multicast group: indicates a group of receivers identified by a multicast IP address. User hosts (or other receiver devices) that have joined a multicast group become members of the group and can identify and receive the IP packets destined for the multicast group address.
- Multicast router: indicates a network device that supports multicast and runs multicast protocols. In addition to routers, switches and firewalls support multicast (depending on device models). A router is only a representative.
- First-hop router: indicates a router that directly connects to the multicast source on the multicast forwarding path and is responsible for forwarding multicast data from the multicast source.
- Last-hop router: indicates a router that directly connects to multicast group members (receivers) on the multicast forwarding path and is responsible for forwarding multicast data to these members.
- In the TCP/IP protocol suite, IGMP manages IP multicast members, and sets up and maintains multicast member relationships between receivers and their directly connected multicast routers.

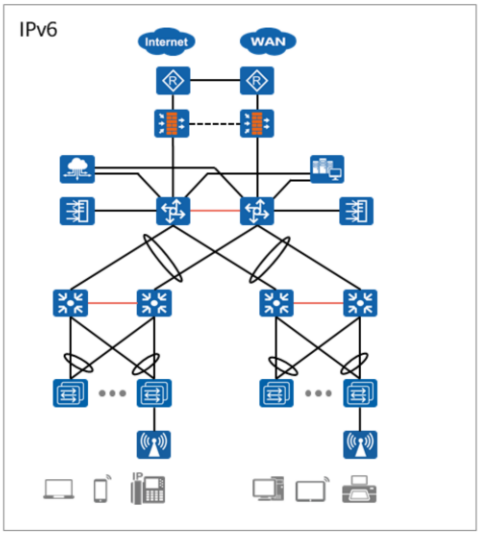
Contents

1. Campus Network Architecture and Common Technologies
2. Ethernet Switching Basics
3. Wireless Access
4. Network Reliability
5. Network Service and Management
6. Network Security
7. VPN
8. Multicast
9. **IPv6**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Overview of IPv6

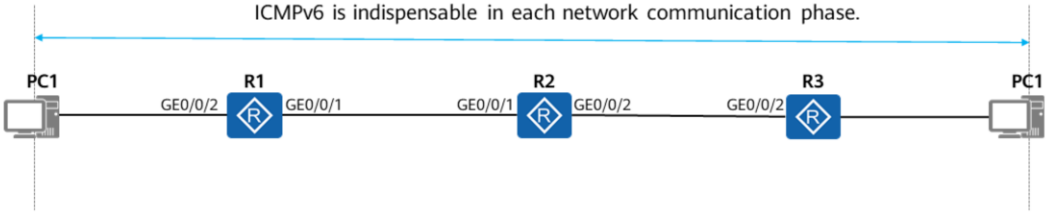


- IPv4 is a widely used Internet protocol. In the early stage of the Internet, IPv4 developed rapidly due to its simple design, easy implementation, and good interoperability. However, with the rapid development of the Internet, the shortage of IPv4 address spaces becomes increasingly serious, and it is imperative that IPv6 replace IPv4.
- Internet Protocol version 6 (IPv6), also called IP Next Generation (IPng), is a second-generation network layer protocol designed by the Internet Engineering Task Force (IETF), IPv6 is an upgraded version of IPv4.
- The 128-bit address length provides numerous addresses, meeting requirements of emerging services such as the IoT and facilitating service evolution and expansion.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

ICMPv6

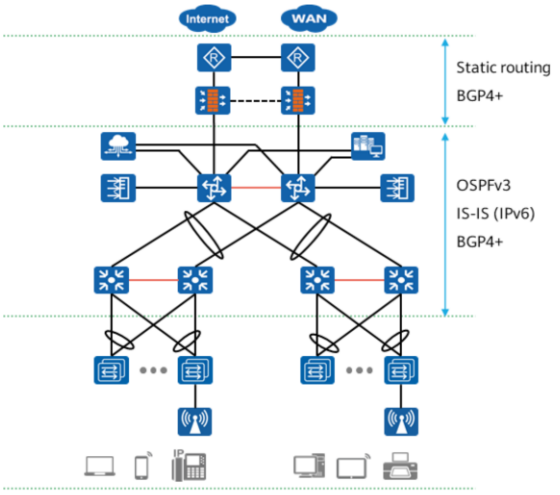
- The Internet Control Message Protocol version 6 (ICMPv6) is one of the basic IPv6 protocols.
- ICMPv6 messages are widely used in other protocols, such as NDP and path MTU discovery.
- ICMPv6 controls key IPv6 processes, such as address autoconfiguration, address resolution, address conflict detection, route selection, and error control.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- NDP: Neighbor Discovery Protocol

IPv6 Routing



IPv6 networks support static and dynamic routing protocols.

- Static route:
  - The method of configuring an IPv6 static route is the same as that of configuring an IPv4 static route.
- Dynamic routing protocols:
  - OSPFv3
  - IS-IS
  - BGP4+

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Multiple) Which of the following statements about Layer 2 interfaces of switches are true?( )
   A. Layer 2 interfaces on switches include access, trunk, and hybrid interfaces.
   B. An access interface can be added to only one VLAN.
   C. By default, a trunk interface allows traffic from VLAN 1 to pass through and sends untagged packets.
   D. On a hybrid interface, you can run a command to specify whether the packets sent from a specified VLAN carry VLAN tags.
2. (Multiple) What are advantages of Ethernet link aggregation?( )
   A. Improved bandwidth: The maximum bandwidth of a link aggregation group (LAG) is the combined bandwidth of all member links.
   B. Improved reliability: If an active link is faulty, traffic can be switched to other available member links.
   C. Load balancing: The traffic load can be balanced among the active member links in a LAG.
   D. Intelligent optimization: The number of active interfaces can be automatically adjusted based on the traffic volume.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. ABCD
2. ABC

Summary

- This course describes typical application scenarios of campus network technologies, including Ethernet switching basics, wireless access, network reliability, network service and management, routing and network reachability, network security, and VPN.
- This document cannot list all technologies used on campus networks. After completing this course, you will be able to understand the typical architecture of campus networks and the application of typical technologies on campus networks. This will help you continue to learn the following courses.
- In subsequent courses, we will discuss key technologies of campus networks to lay a solid foundation for learning campus network solutions in the future.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



RSTP Implementation and Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- On an Ethernet switching network, redundant links are used to implement link backup and enhance network reliability. The downside of this is that it may produce loops, leading to broadcast storms and an unstable MAC address table. As a result, communication on the network may deteriorate or even be interrupted. To prevent loops, IEEE introduced the Spanning Tree Protocol (STP), which is standardized as IEEE 802.1d.
- The convergence speed of an STP topology slows as the number of LANs increases. Therefore, IEEE introduced the Rapid Spanning Tree Protocol (RSTP), standardized as 802.1w, in 2001 to improve the network convergence speed.
- This document describes the improvements of RSTP compared with STP, working mechanism of RSTP, and RSTP configurations.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe defects of STP technology.
  - Describe RSTP improvements compared with STP.
  - Describe the working mechanism of RSTP.
  - Perform basic RSTP configurations.

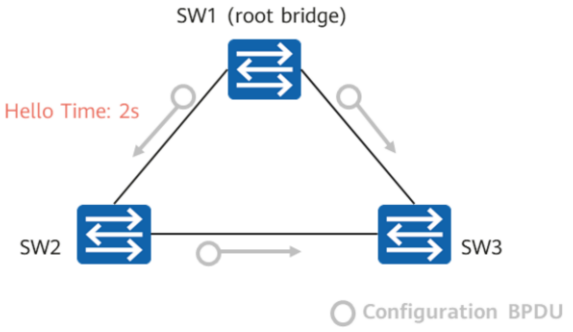
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Introduction to RSTP**
   - **STP Review and Defects**
   - RSTP Overview
2. Improvements Made in RSTP
3. Working Mechanism of RSTP
4. RSTP Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Review: STP Implementation



STP Configuration BPDUs

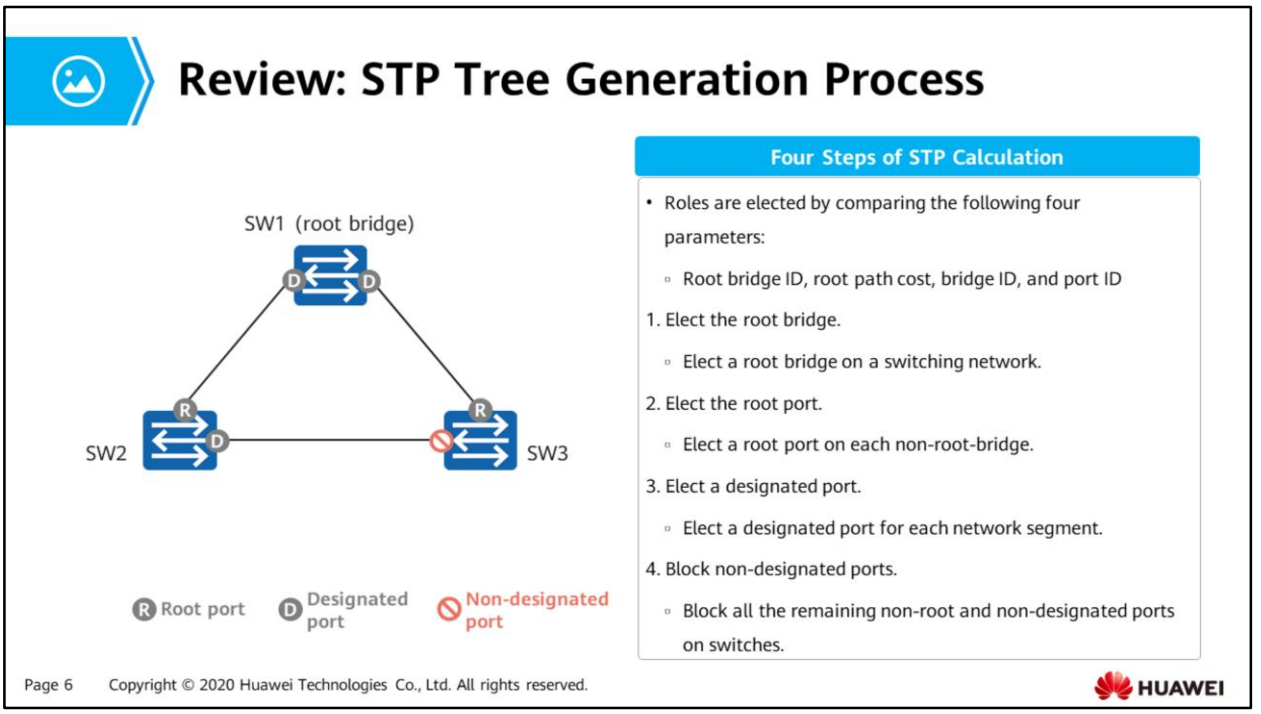
- STP transmits configuration BPDUs between switches to elect the root switch (or root bridge) and determine the role and status of each switch port.
  - Each switch actively sends configuration BPDUs during initialization.
  - After the network topology becomes stable, only the root bridge proactively sends configuration BPDUs. Other bridges send configuration BPDUs only after receiving configuration BPDUs from uplink devices.
- A configuration BPDU contains parameters such as the bridge ID, path cost, and port ID.

Format of configuration BPDUs

PID	PVI	BPDU Type	Flag	Root ID	RPC	Bridge ID	Port ID	Message Age	Max Age	Hello Time	Forward Delay

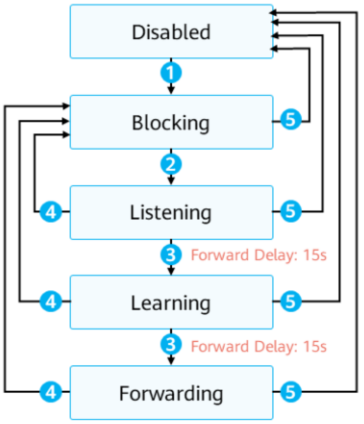
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Flooding of configuration BPDUs:
  - During generation of the STP tree, all STP switches generate and send configuration BPDUs periodically (Hello Time, 2s by default). All STP switches consider themselves as the root bridge.
  - When BPDUs are flooded and collected, switches compare information in BPDUs and elect the root bridge.
  - After the STP tree is formed, only the root bridge generates and sends configuration BPDUs periodically (2s by default). A non-root bridge periodically receives configuration BPDUs from its root port, immediately generates configuration BPDUs, and sends the configuration BPDUs through its designated port. During the process, configuration BPDUs from the root bridge pass through the other switches hop by hop. As shown in the figure, the link between SW1 and SW2 is the uplink of SW2, and the link between SW2 and SW3 is the downlink of SW2.
- Packet format:
  - Parameters in BPDUs are classified into the following types:
  - Type 1: BPDU identifiers, including Protocol ID, Protocol version ID, BPDU Type, and Flag.
    - Protocol ID (PID): The value has 2 bytes and is always 0x000.
    - Protocol version ID (PVI): The value has 1 byte and is always 0x00.
    - BPDU Type: The value has 1 byte and is always 0x00.
    - Flag: It refers to the network topology change flag. The value has 1 byte. Only the least significant bit and most significant bit are used.



- STP working mechanism:
  - On a switching network with loops, switches run STP to automatically generate a loop-free working topology, which is also called an STP tree. A tree node is a specific switch, and a tree branch is a specific link.
- STP uses the following four steps to prevent Layer 2 loops (a spanning tree is generated):
  - Elect a root bridge on a switching network; elect a root port on each non-root bridge; elect a designated port for each network segment; block all the remaining non-root and non-designated ports (alternate ports) on switches.
- How is an STP tree generated?
  - Compare the root bridge ID, root path cost, bridge ID, and port ID. A smaller value indicates a higher priority. These parameters are all fields in BPDUs.
    - Root bridge election: The device with the smallest root bridge ID is the root bridge.
    - Root port election: The system compares the RPC, peer BID, peer PID, and local PID in sequence and selects the port with the smallest value.
    - Designated port election: The system compares the RPC, local BID, and local PID in sequence and selects the port with the smallest value.
    - After the root port and designated port are determined, all the remaining non-root ports and non-designated ports on the switch are blocked.
- On Huawei switches, a blocked non-designated port is an alternate port.

Review: STP Port State Transition



STP Port State Transition

1. The port is initialized or enabled and enters the Blocking state.
2. If the port is selected as the root port or designated port, it enters the Listening state.
3. The port enters the Learning state after the Forward Delay timer expires. After another Forward Delay timer, the port enters the Forwarding state.
4. If a port is no longer the root port or designated port, it enters the Blocking state.
5. The port is disabled or the link is terminated.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- STP defines five port states: Disabled, Blocking, Listening, Learning, and Forwarding, depending on whether the port can receive and send STP BPDUs and whether the port can forward user data frames.
  - Disabled: The port cannot receive or send any frame. That is, the port does not process BPDUs or forward user data frames. The port is in Down state.
  - Blocking: The port can only receive and process BPDUs but cannot send BPDUs or forward user data frames.
  - Listening: The port can receive and send BPDUs but cannot learn MAC addresses or forward user data frames. This is a transitional state. It is used to determine the port role, elect the root bridge, root port, and designated port, and prevent temporary loops.
  - Learning: The port can receive and send BPDUs, learn MAC addresses, and create a MAC address table based on received user data frames. However, the port cannot forward user data frames. This is a transitional state, which is used to prevent the flooding of a large number of user data frames on the network when the MAC address table is not created.
  - Forwarding: Ports can receive and send BPDUs, or they can learn MAC addresses, while forwarding user data frames. Only the root port and designated port can enter the Forwarding state.

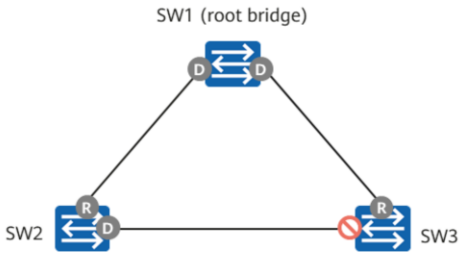
Disadvantages of STP

- STP ensures a loop-free network but is slow to converge, leading to service quality deterioration. If the network topology changes frequently, connections on the STP network are frequently torn down, causing frequent service interruption.
- STP has the following disadvantages:
  - STP does not differentiate between port roles according to their states, making it difficult for less experienced administrators to learn about and deploy this protocol.
    - Ports in Listening, Learning, and Blocking states are the same for users because they are all prevented from forwarding service traffic.
    - In terms of port use and configuration, the essential differences between ports lie in the port roles but not port states.
  - The STP algorithm does not determine topology changes until the timer expires, delaying network convergence.
  - The STP algorithm requires the root bridge to send configuration BPDUs after the network topology becomes stable, and other devices process and spread the configuration BPDUs through the entire network. This also delays convergence.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

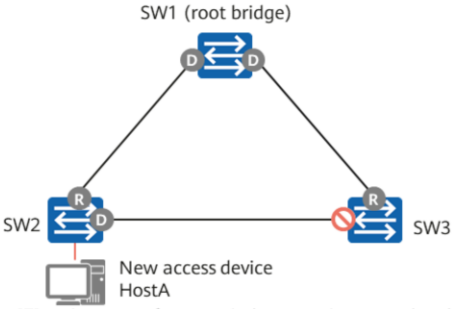
STP Dependency on Timers

Initialization



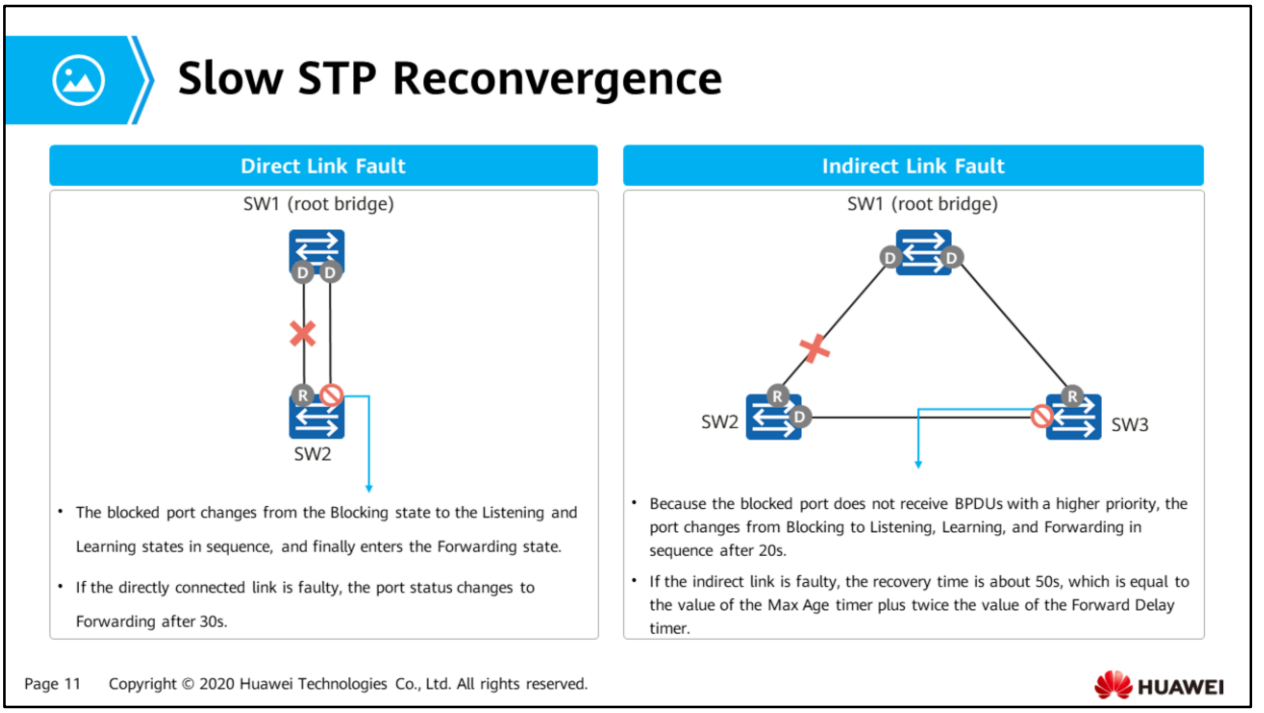
STP uses a timer to prevent temporary loops. After STP elects a port role, even if the port is a designated port or a root port, it still needs to wait for two intervals of the Forward Delay timer (30s) before forwarding packets.

Terminal Access



In the STP environment, after a terminal or server is connected to the network, the port needs to switch from the Disabled state to the Blocking, Listening, Learning, and Forwarding states in sequence. In this case, HostA needs to wait for two intervals of the Forward Delay timer before accessing the network service.

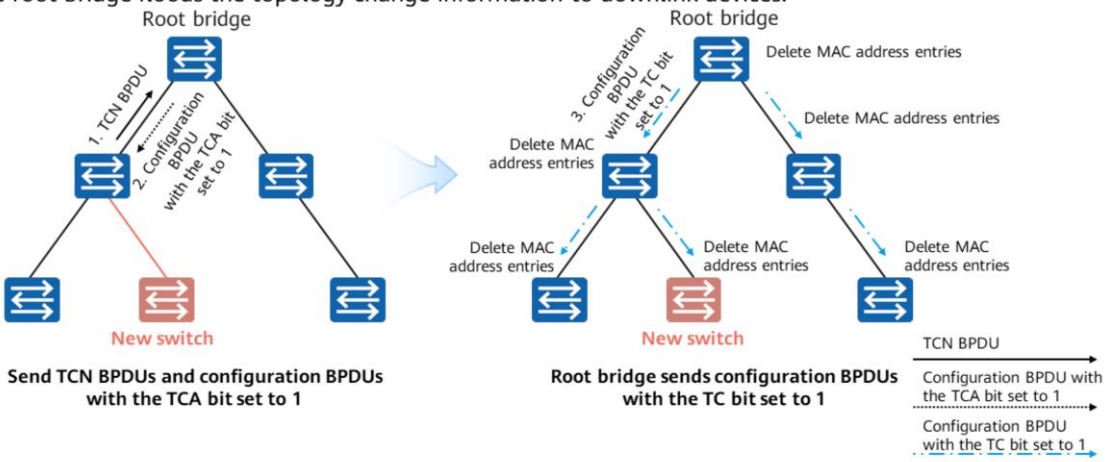
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Direct link fault:
  - There are two links between two switches. One is the active link and the other is the standby link.
  - When the network is stable, SW2 detects that the link of the root port is faulty. The blocked port starts the port state transition and finally enters the Forwarding state to forward user traffic.
- Indirect link fault:
  - When the network is normal, the blocked port of SW3 periodically receives BPDUs from the root bridge.
  - When the link between SW1 and SW2 is faulty, SW2 can detect the fault immediately. In this case, SW2 considers itself as the new root bridge and sends its own configuration BPDU to SW3. The root bridge ID is its own bridge ID.
  - The blocked port on SW3 receives the configuration BPDU, but the configuration BPDU is inferior to the configuration BPDU buffered on the port. Therefore, SW3 ignores the configuration BPDU.
  - When the Max Age timer expires, the configuration BPDUs buffered on SW3 age and SW3 starts to send configuration BPDUs to SW2. The configuration BPDUs are triggered by the configuration BPDUs sent by the root bridge SW1. The value of the root bridge ID field in the configuration BPDUs is the bridge ID of SW1.
  - After SW2 receives the configuration BPDU, it parses the BPDU and determines that SW1 is the root bridge. Therefore, SW2 changes the port connected to SW3 to the root port.

STP Topology Change Mechanism

The STP topology change mechanism transmits topology change information to the root bridge, and then the root bridge floods the topology change information to downlink devices.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- STP processing when the topology changes:
  - When a switch detects a topology change, it notifies the root bridge of the spanning tree. The root bridge then floods the topology change information to the entire network.
  - Topology change process:
    - If a switch is added to the network and the working topology changes, the switch at the change point can directly detect the change through the port status, but other switches cannot directly detect the change.
    - The switch at the change point continuously sends TCN BPDUs to the uplink device through the root port at an interval of Hello time (2s by default) until it receives the configuration BPDUs with the TCA bit set to 1 from the uplink switch. The TCA bit is set to 1 to instruct the downlink device to stop sending TCN BPDUs.
    - After receiving a TCN BPDU, the uplink switch replies with a configuration BPDU with the TCA bit set to 1 through the designated port and sends TCN BPDUs to the uplink switch through the root port at an interval of Hello time.
    - This process repeats until the root bridge receives a TCN BPDU.
    - After receiving the TCN BPDU, the root bridge sends a configuration BPDU in which the TC bit is set to 1 to notify all switches of the network topology change and instruct downlink devices to delete the bridge MAC address entry.

Contents

1. **Introduction to RSTP**
   - STP Review and Defects
   - **RSTP Overview**
2. Improvements Made in RSTP
3. Working Mechanism of RSTP
4. RSTP Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

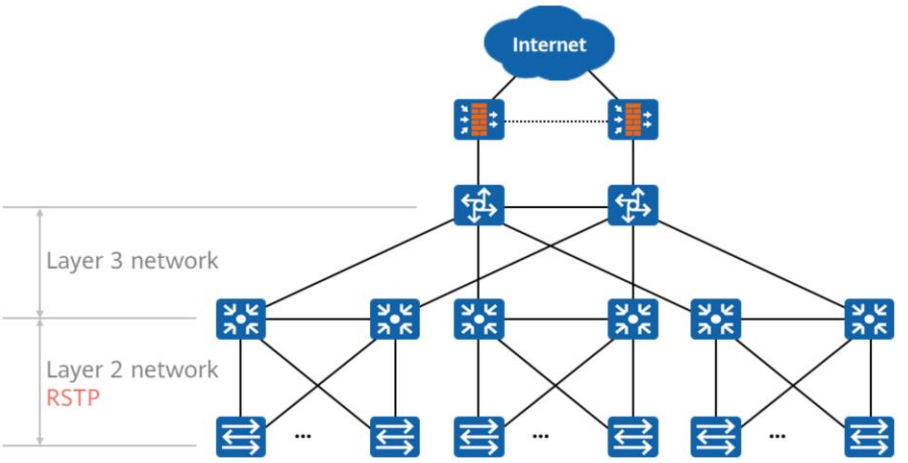
RSTP Overview

- RSTP defined in IEEE 802.1w was developed based on STP. RSTP optimizes STP in many aspects, provides a faster convergence speed, and is compatible with STP.
- RSTP has the following improvements:
  - Defines additional port roles to simplify the learning and deployment of STP.
  - Redefines port states.
  - Changes the configuration BPDU format and uses the Flags field to describe port roles.
  - Processes configuration BPDUs differently from STP.
  - Provides fast convergence.
  - Adds protection functions.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- RSTP can interoperate with STP, but doing so causes RSTP to lose its advantages, such as fast convergence.
  - On a network with both STP-capable and RSTP-capable devices, STP-capable devices discard RST BPDUs. If a port on an RSTP-capable device receives a configuration BPDU from an STP-capable device, the port switches to the STP mode and starts to send configuration BPDUs after two Hello timer intervals.
  - After STP-capable devices are removed, Huawei RSTP-capable devices can be switched back to the RSTP mode.

RSTP Application on a Campus Network

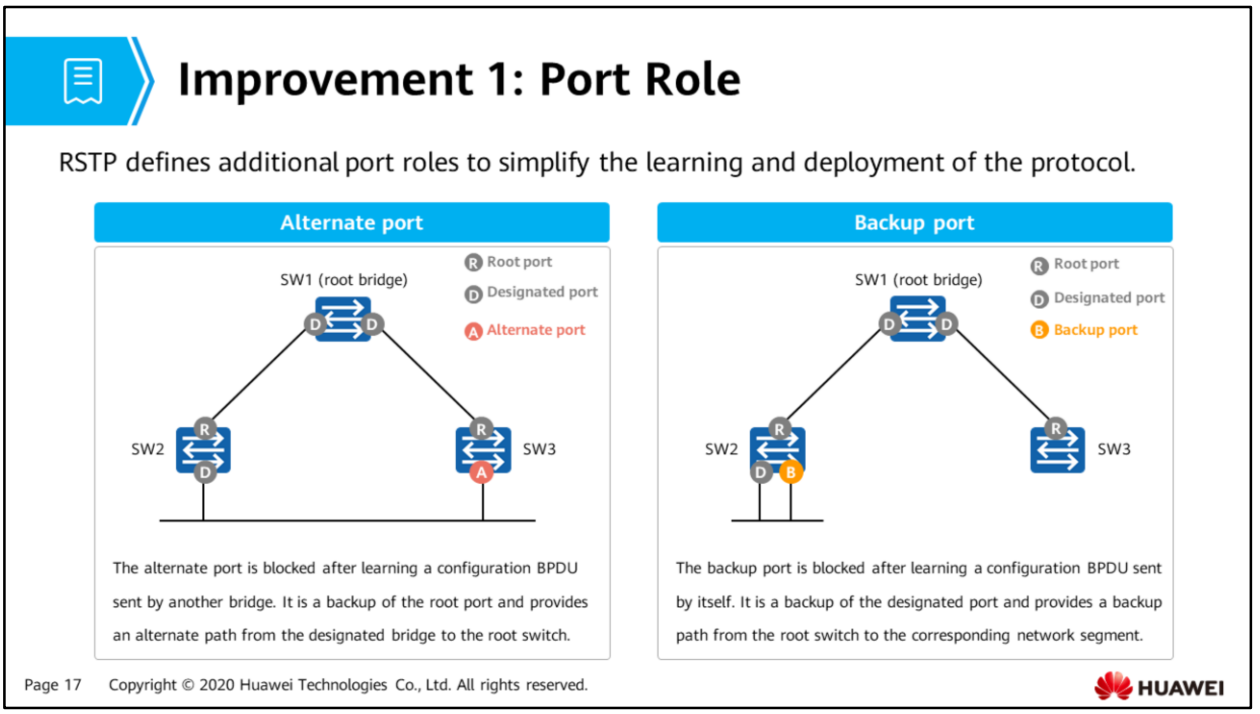


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Contents

1. Introduction to RSTP
2. **Improvements Made in RSTP**
3. Working Mechanism of RSTP
4. RSTP Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- RSTP defines four port roles: root port, designated port, alternate port, and backup port.
- The functions of the root port and designated port are the same as those defined in STP. The alternate port and backup port are defined as follows:
  - From the perspective of configuration BPDU transmission:
    - An alternate port is blocked after learning a configuration BPDU sent from another bridge.
    - A backup port is blocked after learning a configuration BPDU sent from itself.
  - From the perspective of user traffic:
    - An alternate port acts as a backup of the root port and provides an alternate path from the designated bridge to the root bridge.
    - A backup port backs up a designated port and provides a backup path from the root bridge to the related network segment.
- After roles of all RSTP ports are determined, the topology convergence is completed.

Improvement 2: Port States

RSTP defines three states, depending on whether a port forwards user traffic and learns MAC addresses.

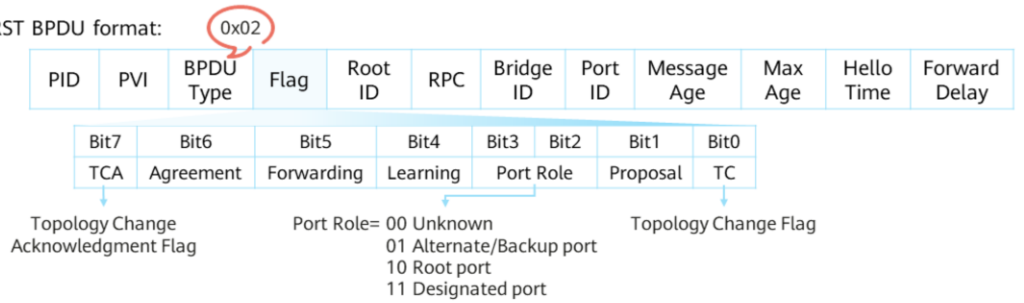
- Discarding: The port does not forward user traffic or learn MAC addresses.
- Learning: The port does not forward user traffic but learns MAC addresses.
- Forwarding: The port forwards user traffic and learns MAC addresses.

STP Port State	RSTP Port State	Port Role
Forwarding	Forwarding	Root port or designated port
Learning	Learning	Root port or designated port
Listening	Discarding	Root port or designated port
Blocking	Discarding	Alternate port or backup port
Disabled	Discarding	Disabled port

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Improvement 3: Configuration BPDU — RST BPDU

- RSTP configuration BPDUs use the Flag field in STP BPDUs to determine the port role.
- RSTP has the following changes except that the format of RSTP is the same as that of STP:
  - The value of the Type field is changed from 0 to 2. Devices running STP will discard configuration BPDUs sent from devices running RSTP.
  - The Flags field uses the six bits reserved in STP. This configuration BPDU is called a Rapid Spanning Tree Bridge Protocol Data Unit (RST BPDU).
- RST BPDU format:



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The format of an RST BPDU is different from that of an STP configuration BPDU, including the BPDU type and Flag field.
  - BPDU Type: 1 byte. The value of an RST BPDU is 0x02.
  - Flag: 1 byte
    - Bit 7: TCA, indicating that the topology change is acknowledged
    - Bit 6: Agreement, which is used in the P/A mechanism
    - Bit 5: Forwarding
    - Bit 4: Learning
    - Bits 3 and 2: port role

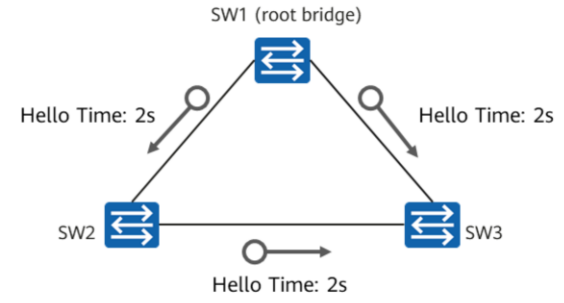
      00 — unknown port

      01 — alternate or backup port

      10 — root port

      11 — designated port
    - Bit 1: Proposal, which is used in the P/A mechanism
    - Bit 0: TC, indicating a topology change

Improvement 4: Configuration BPDU Processing (1)



RST BPDU

Configuration BPDU Transmission After the Topology Becomes Stable

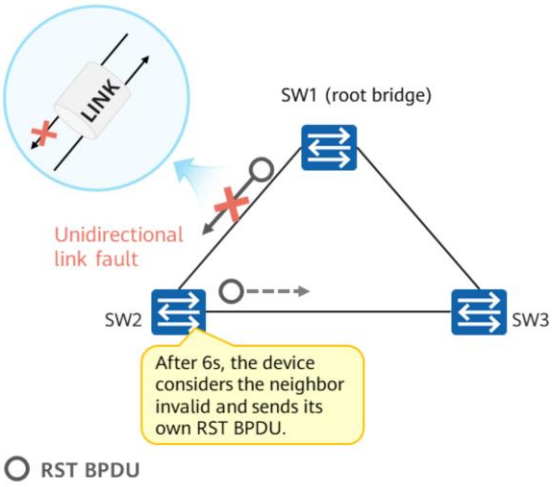
RSTP improves the transmission mode of configuration BPDUs.

- RSTP allows non-root bridges to send configuration BPDUs at an interval of Hello Time after the topology becomes stable, regardless of whether they have received configuration BPDUs from the root bridge.

In STP, the root bridge sends configuration BPDUs at an interval of Hello Time after the topology becomes stable. Non-root bridges send configuration BPDUs only after they receive configuration BPDUs from uplink devices. This complicates the STP calculation and slows down network convergence.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Improvement 4: Configuration BPDU Processing (2)



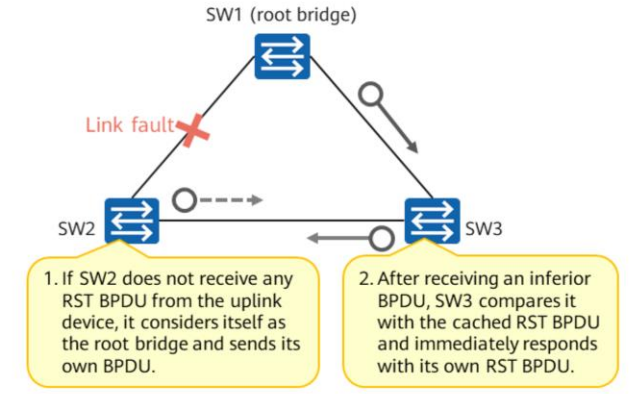
Shorter BPDU Timeout Interval

If a port does not receive any configuration BPDU from the uplink device within the timeout interval (three intervals of the Hello timer), the device considers that the negotiation with the neighbor fails.

STP needs to wait for the time specified by the Max Age timer.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Improvement 4: Configuration BPDU Processing (3)



Processing an Inferior BPDU

- When a port receives an RST BPDU from the uplink designated bridge, the port compares the cached RST BPDU with its own RST BPDU.
- If the number of RST BPDUs cached on the port is superior to the received RST BPDU, the port discards the received RST BPDU and immediately responds with the cached RST BPDU. This speeds up network convergence.

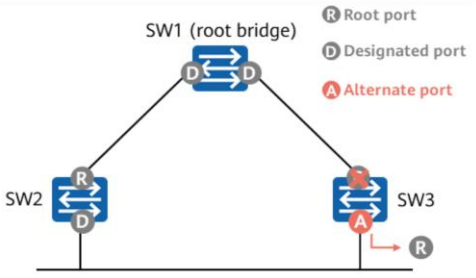
In STP, only the designated port can process the inferior BPDU immediately.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- STP
  - In STP, only the designated port immediately processes inferior BPDUs. Other ports ignore the inferior BPDUs. After the Max Age timer expires, the buffered inferior BPDUs age out and these ports send their superior BPDUs to implement a new round of topology convergence.
- RSTP
  - RSTP processes inferior BPDUs without using any timer (no longer depending on BPDU aging) to implement topology convergence. In addition, any port in RSTP can process inferior BPDUs to speed up topology convergence.

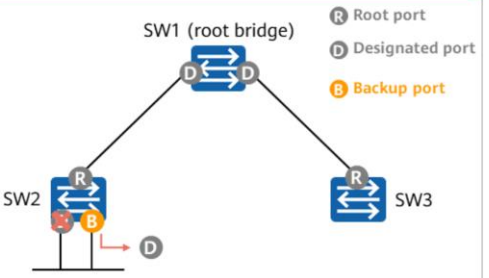
Improvement 5: Fast Convergence Mechanism (1)

Fast Switchover of the Root Port



If a root port fails, the best alternate port becomes the root port and enters the Forwarding state. This is due to the fact that the network segment connected to this alternate port has a designated port that can access the root bridge.

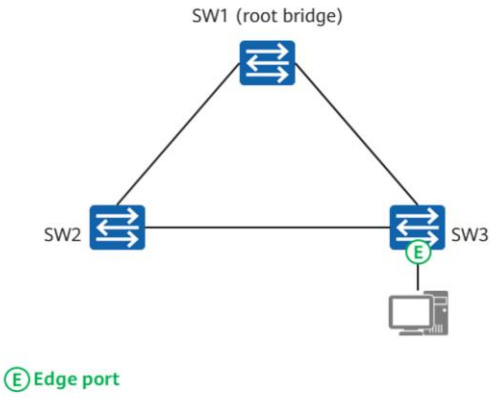
Fast Switchover of the Designated Port



If a designated port fails, the best backup port becomes the designated port and enters the Forwarding state. A backup port backs up a designated port and provides a backup path from the root bridge to the related network segment.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Improvement 5: Fast Convergence Mechanism (2)



E Edge port

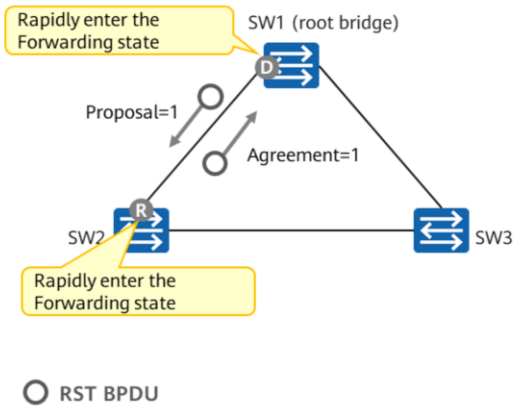
Edge Port

- RSTP introduces the edge port. If a port is located at the edge of a network and directly connects to a terminal, it can be configured as an edge port.
- An edge port does not participate in RSTP calculation and can directly enter the Forwarding state from the Discarding state.
- An edge port becomes a common STP port once it is connected to a switching device and receives a configuration BPDU. The spanning tree needs to be recalculated, which leads to network flapping.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The Up and Down states of an edge port do not change the network topology.

Improvement 5: Fast Convergence Mechanism (3)



P/A Mechanism

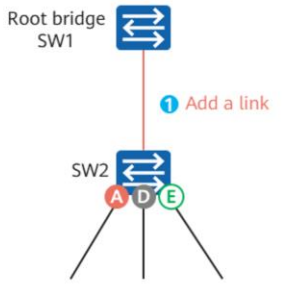
- The Proposal/Agreement (P/A for short) mechanism enables the uplink port to quickly transition to Forwarding state.
- In RSTP, after a port is elected as the designated port, the port enters the Discarding state and then rapidly enters the Forwarding state through the P/A mechanism.

In STP, the port enters the Forwarding state after at least one interval of Forward Delay (Learning state).

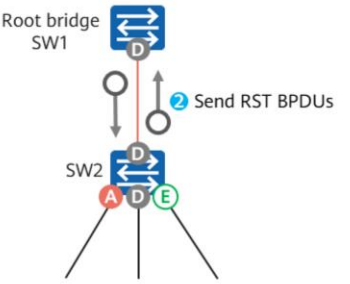
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Although STP can select designated ports quickly, to prevent loops, all ports must wait at least one interval of the Forward Delay timer before forwarding traffic.
- RSTP solves this problem by blocking non-root ports to prevent loops. The P/A mechanism shortens the time that an uplink port waits before transitioning to Forwarding state.

P/A Mechanism (1)



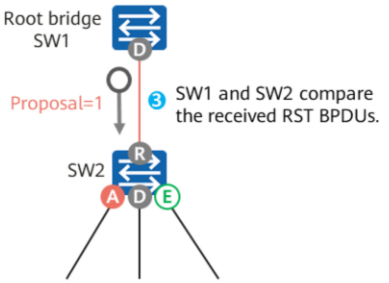
- A link is added between root bridge SW1 and SW2.
- The three downlink ports of SW2 are the alternate port, designated port in Forwarding state, and edge port, respectively.



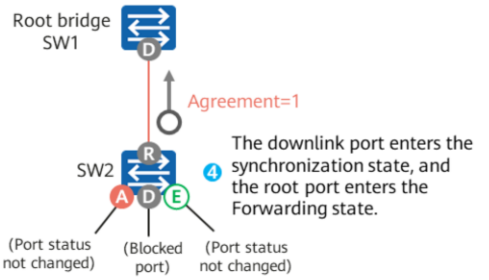
- The two ports between SW1 and SW2 become designated ports and send RST BPDUs.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

P/A Mechanism (2)



- SW2's port connected to SW1 receives a superior RST BPDU, so the port becomes a root port and stops sending RST BPDUs.
- The designated port of SW1 enters the Discarding state and sends an RST BPDU with the Proposal bit set to 1.

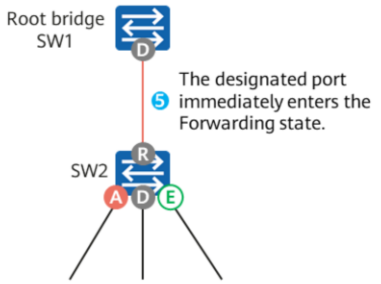


- After receiving the RST BPDU with the Proposal bit set to 1 from the root bridge, SW2 starts to synchronize all its ports.
- After all ports are synchronized, all downlink ports (except edge ports) enter the Discarding state, and the uplink root port enters the Forwarding state and returns an RST BPDU with the Agreement bit set to 1 to SW1.

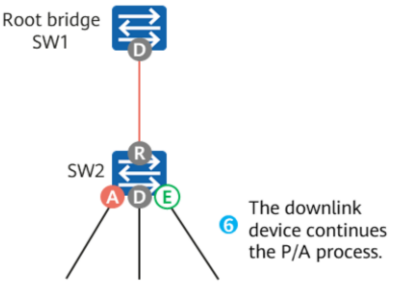
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The downlink ports of SW2 are synchronized as follows: The alternate port status remains unchanged; the edge port does not participate in calculation; the non-edge designated port is blocked.

P/A Mechanism (3)



The RST BPDU with the Agreement bit set to 1 received by SW1 is a response to the sent RST BPDU with the Proposal bit set to 1 on SW1. Therefore, the designated port immediately enters the Forwarding state.

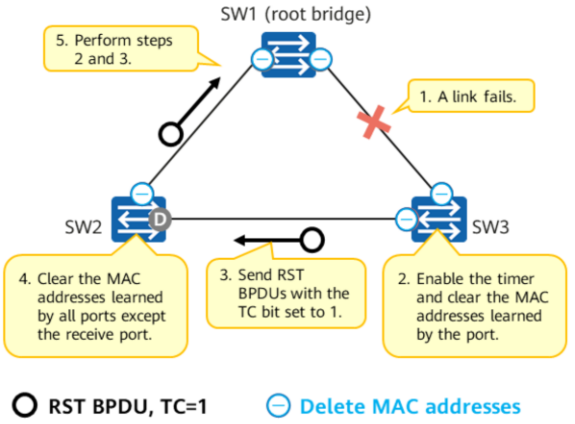


The downlink device continues P/A negotiation.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Improvement 6: Topology Change Mechanism

RSTP considers that the network topology has changed when a non-edge port transitions to the Forwarding state.



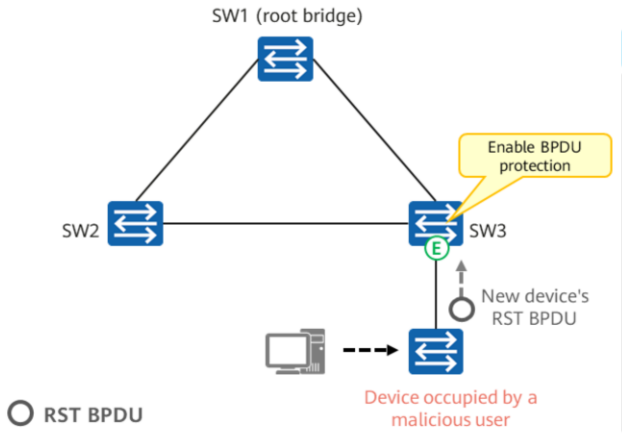
Topology Change Mechanism

- When detecting a topology change, RSTP devices react as follows:
  - The local device starts a TC While timer on each non-edge designated port and root port. The TC While timer value is twice the Hello Time value. Within the TC While time, the local device deletes MAC address entries learned on ports whose states have changed.
  - These ports send out RST BPDUs with the TC bit set to 1. When the TC While timer expires, the ports stop sending RST BPDUs.
  - When other switches receive RST BPDUs, they clear MAC address entries learned on all their ports except the ports that receive the RST BPDUs. These switches also start a TC While timer on each non-edge designated port and repeat the preceding process.
- In this manner, RST BPDUs are flooded on the network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- If the topology of an STP network changes, TCN BPDUs are first sent to the root bridge. Then, the root bridge notifies the topology change and floods the configuration BPDUs with the TC bit set to 1.
- RSTP uses a new topology change mechanism to rapidly flood RST BPDUs with the TC bit set to 1.
- In the figure:
  - If the root port of SW3 cannot receive RST BPDUs from the root bridge, the alternate port quickly becomes the new root port, starts the TC While timer, and clears MAC addresses learned by all ports. Then, the new root port sends RST BPDUs with TC bits set to 1.
  - After receiving the RST BPDU, SW2 clears the MAC addresses learned by all ports except the receive port, starts the timer, and sends the RST BPDU with the TC bit set to 1.
  - RST BPDUs are flooded on the entire network.

Improvement 7: Protection Functions (1)



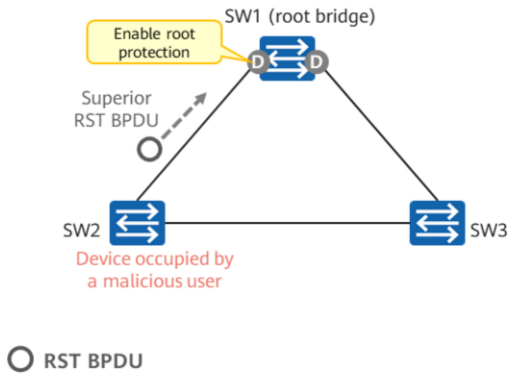
BPDU protection

- On an RSTP network, an edge port does not receive RST BPDUs in normal situations. If a switching device receives malicious RST BPDUs on an edge port, the switching device automatically sets the edge port to a non-edge port and performs STP calculation. This causes network flapping.
- BPDU protection enables a switch to set the state of an edge port to Error-Down if the edge port receives an RST BPDU. In this case, the port remains as the edge port, and the switch sends a notification to the NMS.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- On a switching device, ports directly connected to a user terminal such as a PC or file server are edge ports.
- As shown in the figure:
  - SW3 is connected to a host and is configured as an edge port.
  - Then the host is used by a malicious user to forge RST BPDUs to attack SW3. Therefore, the edge port receives the RST BPDUs, loses the edge port role, and calculates the spanning tree.

Improvement 7: Protection Functions (2)



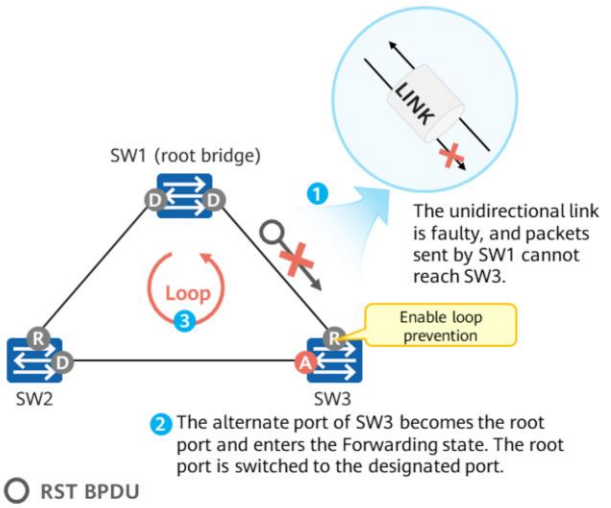
Root protection

- If root protection is enabled on a designated port, the port role cannot be changed.
- Once a designated port that is enabled with root protection receives superior RST BPDUs, the port enters the Discarding state and does not forward packets. If the port does not receive any superior RST BPDUs within a specified period (two intervals of the Forward Delay timer by default), the port automatically enters the Forwarding state.
- Root protection ensures that the role of the root bridge does not change due to network problems.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The root bridge on a network may receive superior RST BPDUs due to incorrect configurations or malicious attacks. When this occurs, the root bridge can no longer serve as the root bridge and the network topology will incorrectly change. As a result, traffic may be switched from high-speed links to low-speed links, leading to network congestion.
- As shown in the figure:
  - When the network is stable, SW1 functions as the root bridge and sends the optimal RST BPDU to downlink devices.
  - If SW2 is occupied by a malicious user, for example, the bridge priority of SW2 is modified to make SW2 have a higher bridge priority than SW1, SW2 sends its own RST BPDU.
  - After receiving the RST BPDU, the designated port of SW1 recalculates the spanning tree. SW1 then loses its role as the root bridge, causing the topology change.

Improvement 7: Protection Functions (3)



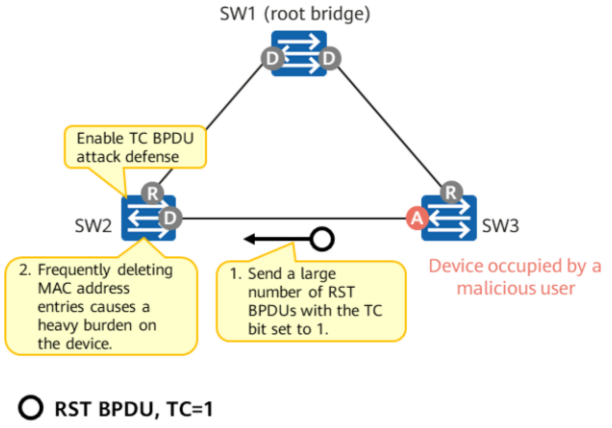
Loop Prevention

- If the root port or alternate port does not receive BPDUs from the uplink device for a long time, the device enabled with loop prevention sends a notification to the NMS. If the root port is used, the root port enters the Discarding state and becomes the designated port. If the alternate port is used, the alternate port keeps blocked and becomes the designated port. In this case, loops will not occur.
- After link congestion is cleared or unidirectional link failures are rectified, the port receives BPDUs for negotiation and restores its original role and status.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- On an RSTP network, a switching device maintains the states of the root port and blocked ports based on RST BPDUs received from the uplink switching device. If the ports cannot receive RST BPDUs from the uplink switching device because of link congestion or unidirectional link failures, the switching device re-selects a root port.
- As shown in the figure, when the unidirectional link between SW1 and SW3 fails, because the root port on SW3 does not receive BPDUs from the uplink device within the timeout interval, the alternate port becomes the root port and the root port becomes the designated port. As a result, a loop occurs.

Improvement 7: Protection Functions (4)



TC BPDU attack defense

- After enabling TC BPDU attack defense on a switching device, you can set the number of TC BPDUs that the device can process within a given period of time.
- If the number of TC BPDUs that the switching device receives within a given time period exceeds the specified threshold, the switching device processes only the specified number of TC BPDUs.
- After the time period expires, the device processes all the excess TC BPDUs in a batch. In this way, the switching device does not need to frequently delete MAC entries.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

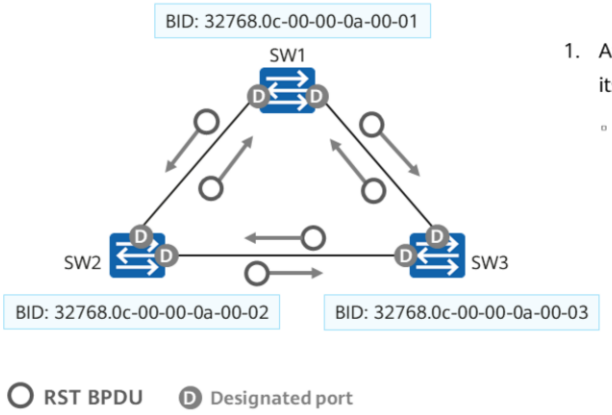
- A switching device deletes its MAC address entries after receiving TC BPDUs. If an attacker sends a large number of malicious RST BPDU with the TC bit set to 1 to the switching device within a short period, the device will constantly delete MAC address entries. This increases the load on the switching device and threatens network stability.
- As shown in the figure:
  - If SW3 is occupied by a malicious user, the attacker forges a large number of RST BPDUs with TC bit set to 1 and sends them. After receiving the RST BPDUs, SW2 frequently deletes MAC address entries, which causes a heavy burden.

Contents

1. Introduction to RSTP
2. Improvements Made in RSTP
3. **Working Mechanism of RSTP**
4. RSTP Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

RSTP Topology Convergence Process (1)

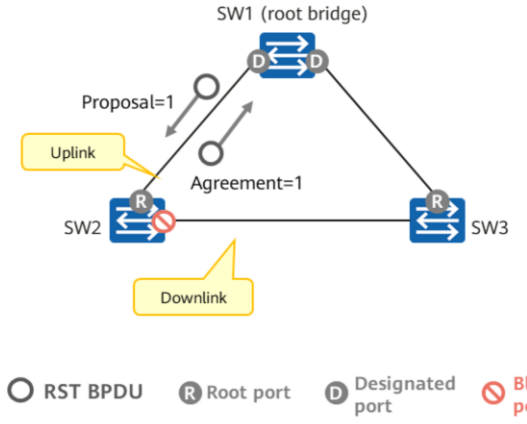


1. After RSTP is enabled on a switch, the switch considers itself as the root bridge and sends RST BPDUs.
   - All ports are designated ports and are in Discarding state.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The RSTP convergence process is similar to the STP convergence process.
- During network initialization, all RSTP switches on the network consider themselves as the root bridge, configure each port as a designated port, and send RST BPDUs. SW1 has the optimal bridge ID and is elected as the root bridge.

RSTP Topology Convergence Process (2)

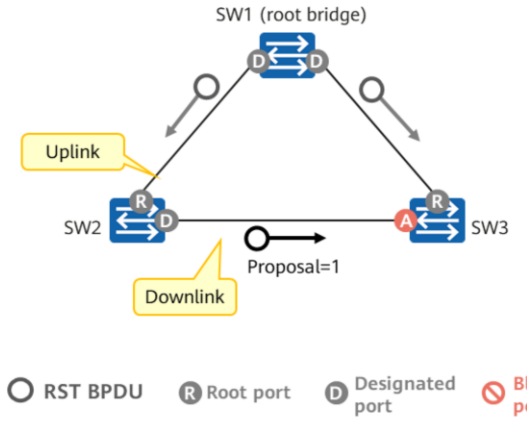


2. The uplink quickly enters the Forwarding state through the P/A mechanism.
   - After receiving a superior RST BPDU, SW2 considers that SW1 is the root bridge and the port on SW2 becomes the root port instead of the designated port. Then SW2 stops sending RST BPDUs.
   - The port on SW1 enters the Discarding state and sends RST BPDUs with the Proposal bit set to 1. After receiving the BPDU, SW2 blocks all ports except the edge port. This process is called synchronization.
   - After ports on SW2 synchronize information, the root port enters the Forwarding state and sends an RST BPDU with the Agreement bit set to 1 to SW1. After SW1 receives the BPDU, the designated port immediately enters the Forwarding state.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Each switch that considers itself as the root bridge generates an RST BPDU to negotiate the port status on the specified network segment. The Proposal bit in the Flag field of the RST BPDU needs to be set.
- When a port receives an RST BPDU, it compares the received RST BPDU with the local RST BPDU. If the local RST BPDU is superior to the received RST BPDU, the port discards the received RST BPDU and sends a local RST BPDU with the Proposal bit set to 1 to reply to the peer device.
- As shown in the preceding figure, the link between SW1 and SW2 is used as an example to describe the uplink convergence process.

RSTP Topology Convergence Process (3)



3. The interconnection port of the downlink starts a new round of P/A negotiation.
   - The downlink port of SW2 is configured as the designated port and continuously sends RST BPDUs with the Proposal bit set to 1.
   - After receiving the BPDU, the downlink port of SW3 finds that the received BPDU is not the optimal one. Therefore, SW3 ignores the received BPDU and does not send an RST BPDU with the Agreement bit set to 1.
   - The downlink interface of SW2 does not receive any response packet with the Agreement bit set to 1. SW2 enters the Forwarding state after two intervals of the Forward Delay timer.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The interconnection port of the downlink enters the slow convergence process. SW2 and SW3 are used as an example.

Contents

1. Introduction to RSTP
2. Improvements Made in RSTP
3. Working Mechanism of RSTP
4. **RSTP Configurations**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic RSTP Configuration Commands (1)

1. Configure a working mode.

   [Huawei] **stp mode** { **stp** | **rstp** | **mstp** }

   The switch supports three working modes: STP, RSTP, and Multiple Spanning Tree Protocol (MSTP). By default, a switch works in MSTP mode.

2. (Optional) Configure the switch as the root bridge.

   [Huawei] **stp root primary**

   By default, a switch does not function as the root bridge of any spanning tree. After you run this command, the priority value of the switch is set to 0 and cannot be changed.

3. (Optional) Configure the switch as the secondary root bridge.

   [Huawei] **stp root Secondary**

   By default, a switch does not function as the secondary root bridge of any spanning tree. After you run this command, the priority value of the switch is set to 4096 and cannot be changed.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic RSTP Configuration Commands (2)

1. (Optional) Configure the STP priority of a switch.

   [Huawei] **stp priority** *priority*

   The value ranges from 0 to 61440, with an increment of 4096. By default, the priority value of a switch is 32768.

2. (Optional) Configure a path cost for a port.

   [Huawei] **stp pathcost-standard** { **dot1d-1998** | **dot1t** | **legacy** }

   Configure a path cost calculation method. By default, the IEEE 802.1t standard (dot1t) is used to calculate the path costs.
   All switches on a network must use the same path cost calculation method.

   [Huawei-GigabitEthernet0/0/1] **stp cost** *cost*

   Set the path cost of the port.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The following describes the supported cost range for different calculation methods:
  - dot1d-1998: Uses the IEEE 802.1d-1998 standard to calculate the path cost. The value ranges from 1 to 65535.
  - dot1t: Uses the IEEE 802.1t standard to calculate the path cost. The value ranges from 1 to 200,000,000.
  - Legacy: Uses Huawei calculation method to calculate the path cost. The value ranges from 1 to 200,000.

Basic RSTP Configuration Commands (3)

1. (Optional) Configure the interface priority.

   [Huawei-GigabitEthernet0/0/1] **stp priority** *priority*

   The value is an integer that ranges from 0 to 240, with an increment of 16. By default, the priority of a switch port is 128.

2. Enable STP or RSTP.

   [Huawei] **stp enable**

   By default, STP or RSTP is enabled on a switch.

3. Configure the port as an STP edge port.

   [Huawei-GigabitEthernet0/0/1] **stp edged-port enable**

   By default, all the ports on a switch are non-edge ports.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

RSTP Protection Configuration Commands (1)

1. Enable BPDU protection on an edge port of a switch.

   [Huawei] **stp bpdu-protection**

   By default, BPDU protection is disabled on a switch.

2. Configure root protection.

   [Huawei-GigabitEthernet0/0/1] **stp root-protection**

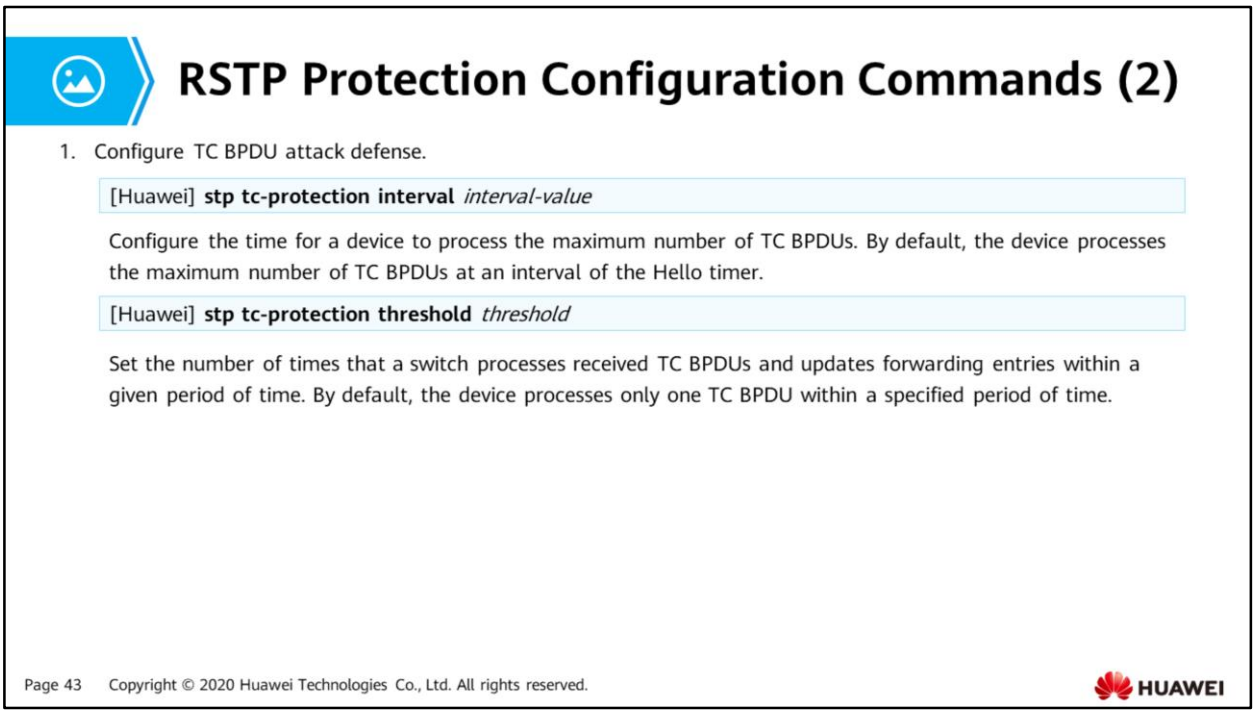
   By default, root protection is disabled on a port. Root protection takes effect only on designated ports. Root protection and loop prevention cannot be configured on the same port.

3. Configure loop prevention on the root port or alternate port.

   [Huawei-GigabitEthernet0/0/1] **stp loop-protection**

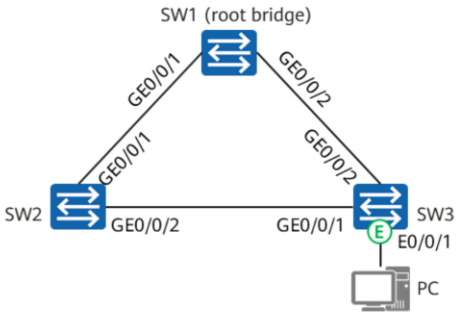
   By default, loop prevention is disabled on a port.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Within the time specified by **stp tc-protection interval**, the switch processes the number of TC BPDUs specified by **stp tc-protection threshold**. Packets that exceed this threshold are delayed, so spanning tree convergence may be affected. For example, the period is set to 10s and the threshold is set to 5. After the switch receives TC BPDUs, the switch processes the first five TC BPDUs within 10s. After 10s, the switch processes subsequent TC BPDUs.

Case: Basic RSTP Configuration (1)



- RSTP is configured on the three switches to eliminate Layer 2 loops.
- The configuration roadmap is as follows:
  - Configure SW1 as the root bridge and SW2 as the secondary root bridge.
  - Configure the port connected to the PC as the edge port because this port does not participate in RSTP calculation.
  - Configure root protection and BPDU protection to protect devices or links.

Enable RSTP on SW1.

```
[SW1] stp mode rstp
[SW1] stp enable
[SW1] stp root primary
```

Enable RSTP on SW2.

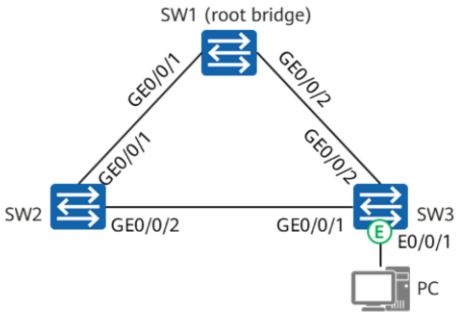
```
[SW2] stp mode rstp
[SW2] stp enable
[SW2] stp root secondary
```

Enable RSTP on SW3.

```
[SW3] stp mode rstp
[SW3] stp enable
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Basic RSTP Configuration (2)



- RSTP is configured on the three switches to eliminate Layer 2 loops.
- The configuration roadmap is as follows:
  - Configure SW1 as the root bridge and SW2 as the secondary root bridge.
  - Configure the port connected to the PC as the edge port because this port does not participate in RSTP calculation.
  - Configure root protection and BPDU protection to protect devices or links.

Enable the edge port on SW3.

```
[SW3-Ethernet0/0/1] stp edged-port enable
```

Enable root protection on SW1.

```
[SW1-GigabitEthernet0/0/1] stp root-protection
[SW1-GigabitEthernet0/0/2] stp root-protection
```

Enable BPDU protection on SW3.

```
[SW3] stp bpdu-protection
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Multiple) Which of the following are RSTP port states? ( )
   A. Idle
   B. Discarding
   C. Forwarding
   D. Learning
2. (TorF) RSTP root protection must be configured on the root port of the device. ( )

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. BCD
2. False

Summary

- STP prevents loops on a LAN. Devices running STP exchange information with one another to discover loops on the network, and block certain ports to eliminate loops. With the growth in scale of LANs, STP has become an important protocol for a LAN.
- Based on STP, RSTP has many improvements and greatly speeds up network convergence.
- This document describes seven improvements of RSTP compared with STP, including the port role, port status, BPDU format, BPDU processing mode, fast convergence mechanism, topology change mechanism, and four protection features.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



MSTP Implementation and Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- The Rapid Spanning Tree Protocol (RSTP), an enhancement to the Spanning Tree Protocol (STP), allows for fast network topology convergence. When RSTP/STP runs on a VLAN-based network, all VLANs on a local area network (LAN) use the same spanning tree. The blocked link does not carry any traffic, and traffic cannot be load balanced among VLANs. As a result, the link bandwidth usage and device resource usage are low.
- To offset disadvantages of RSTP/STP, IEEE introduced the Multiple Spanning Tree Protocol (MSTP) in 2002, which is standardized as IEEE 802.1s. MSTP is compatible with STP and RSTP. Multiple loop-free trees are set up to prevent broadcast storms and implement redundancy.
- This document describes the improvements of MSTP compared with RSTP/STP, basic concepts and working mechanism of MSTP, and MSTP configurations.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe weaknesses of RSTP/STP.
  - Describe MSTP improvements compared with RSTP/STP.
  - Describe concepts of MSTP.
  - Describe the working mechanism of MSTP.
  - Complete basic MSTP configurations.

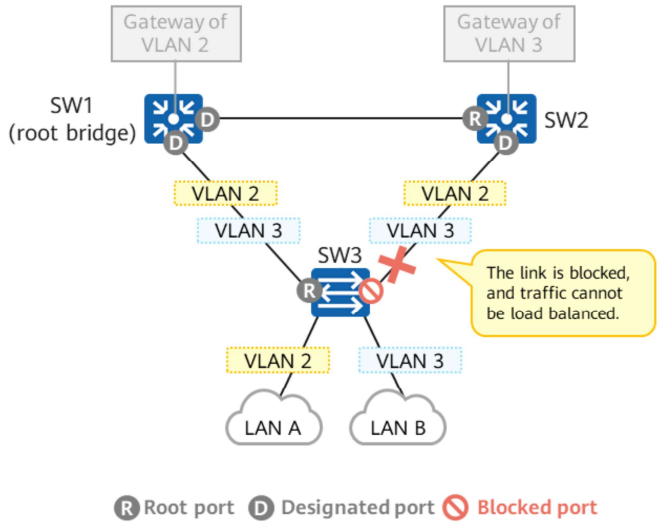
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Introduction to MSTP**
2. Basic Concepts of MSTP
3. Working Mechanism of MSTP
4. MSTP Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Disadvantages of RSTP/STP (1)

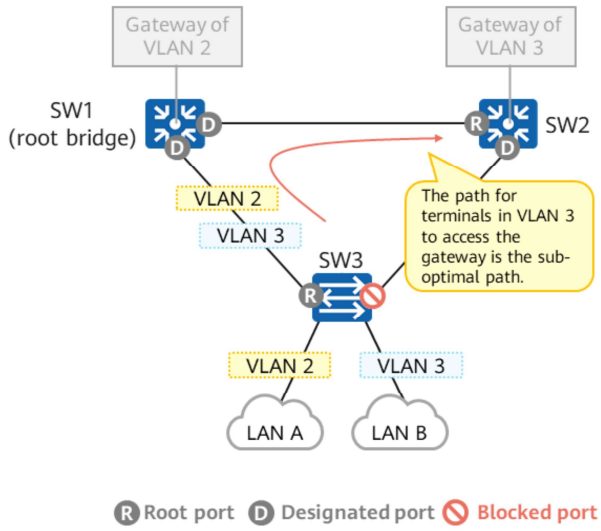


Disadvantage 1: Traffic Cannot Be Load Balanced

- Background:
  - SW3 is an access switch connected to a terminal network segment. SW3 is connected to SW1 and SW2 through two links, and all the links allow packets from VLAN 2 and VLAN 3 to pass through.
  - SW1 is configured as the gateway of terminals in VLAN 2 and SW2 as the gateway of terminals in VLAN 3. Terminals in VLAN 2 and VLAN 3 are required to use different links to connect to the corresponding gateways.
- Issue:
  - If there is only one spanning tree on the network and we assume that the port connecting SW3 to SW2 is a blocked port, data of VLAN 2 and VLAN 3 can be transmitted to aggregation switch through only one link. This means that traffic cannot be load balanced.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Disadvantages of RSTP/STP (2)



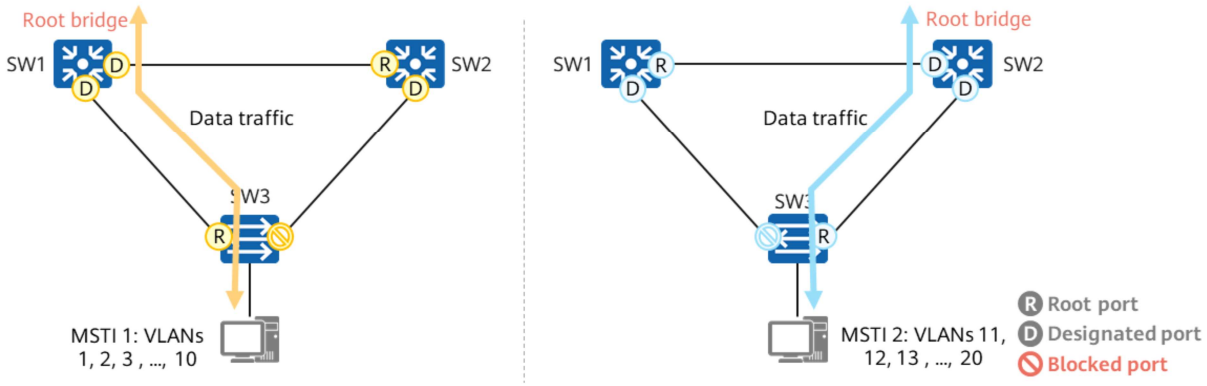
Disadvantage 2: Layer 2 Sub-optimal Path

- Background:
  - SW3 is an access switch connected to a terminal network segment. SW1 and SW2 are aggregation switches. SW1 is configured as the gateway of terminals in VLAN 2 and SW2 as the gateway of terminals in VLAN 3. All links are configured to allow packets from VLAN 2 and VLAN 3 to pass through.
  - After a single spanning tree is run, the loop is broken, and data of VLAN 2 and VLAN 3 is directly sent to SW1.
- Issue:
  - The link between SW3 and SW2 is blocked, so the path from SW3 to the gateway is the sub-optimal path. The optimal path should be the path from SW3 to SW2.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Overview of MSTP

- MSTP, which is standardized as IEEE 802.1s, is compatible with STP and RSTP. It can implement fast convergence and provide multiple redundant paths for forwarding data, effectively load balancing traffic for VLANs.
- MSTP maps one or more VLANs to a Multiple Spanning Tree Instance (MSTI), and then calculates the spanning tree based on the MSTI. The VLANs mapped to the same MSTI share the same spanning tree.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- As shown in the figure, two spanning trees are generated after calculation.
  - The spanning tree corresponding to MSTI 1 uses SW1 as the root bridge to forward packets of VLAN 1 to VLAN 10.
  - The spanning tree corresponding to MSTI 2 uses SW2 as the root bridge to forward packets of VLAN 11 to VLAN 20.
  - Packets of different VLANs are forwarded along different paths, implementing load balancing.

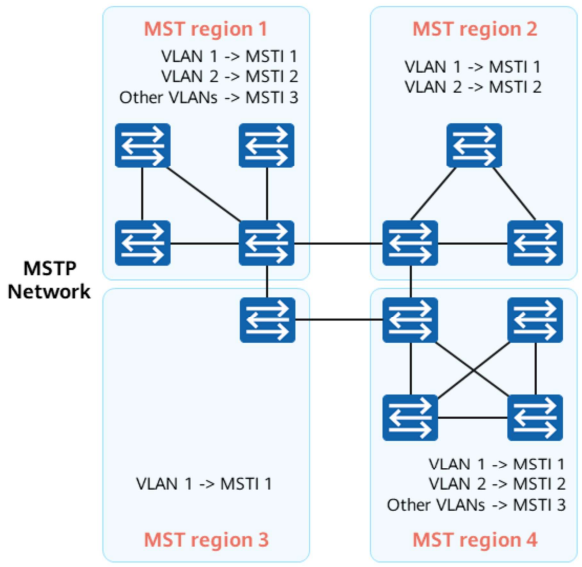
- Note: The spanning tree runs based on MSTIs instead of VLANs.

Contents

1. Introduction to MSTP
2. **Basic Concepts of MSTP**
3. Working Mechanism of MSTP
4. MSTP Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

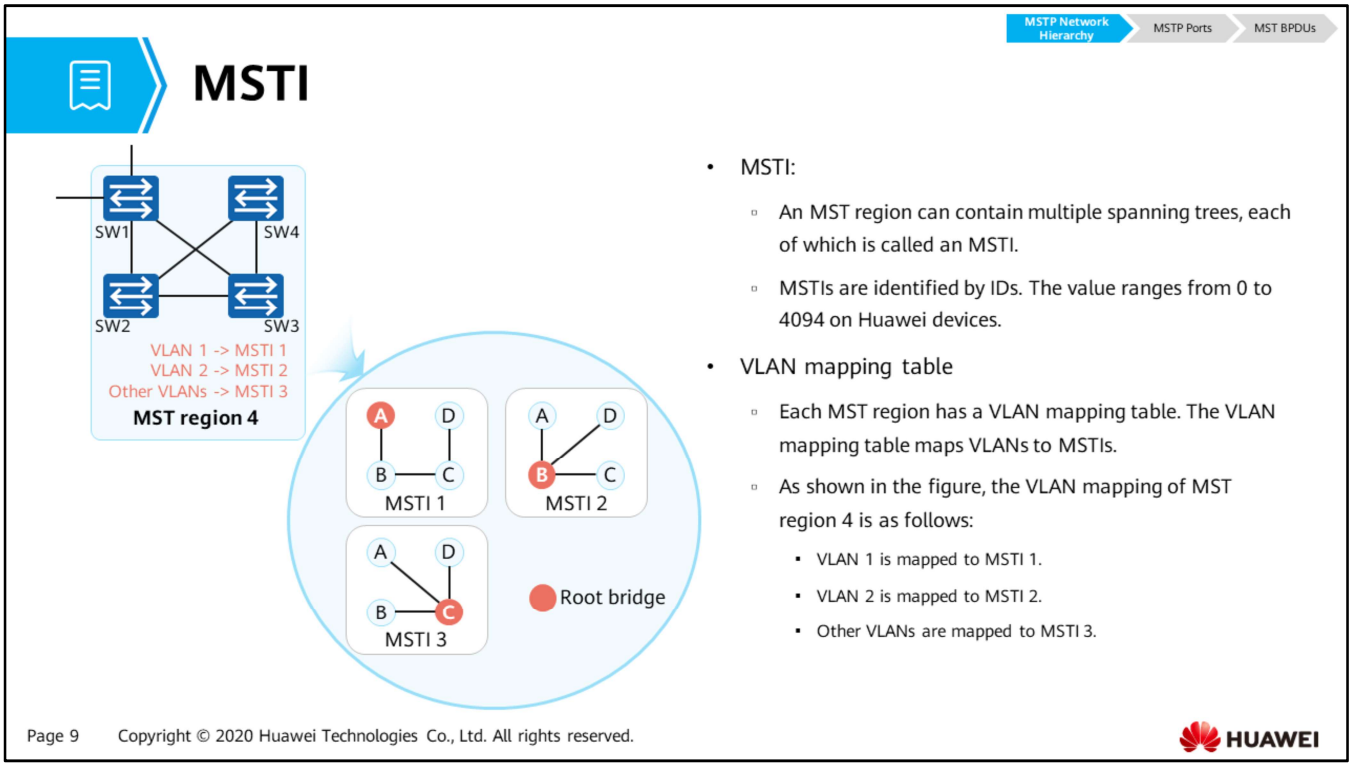
MST Region



- MSTP network hierarchy:
  - MSTP divides a switching network into multiple Multiple Spanning Tree (MST) regions, each of which has multiple spanning trees that are independent of each other.
- MST region:
  - An MST region contains multiple switches and their network segments.
  - A LAN can comprise several MST regions that are directly or indirectly connected. You can add multiple switching devices to an MST region using MSTP configuration commands.
  - An MSTP network contains one or more MST regions, and each MST region contains one or more MSTIs.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

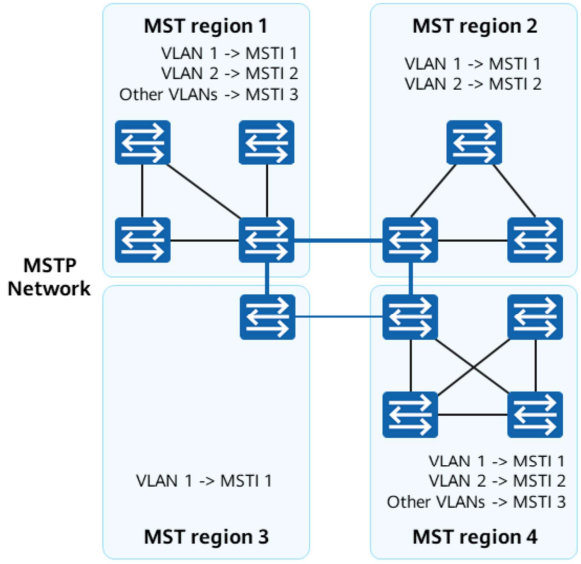
- The switches in one MST region all share the following characteristics:
  - MSTP-enabled
  - Same region name
  - Same VLAN-MSTI mappings
  - Same MSTP revision level



- MSTI 0 exists by default. By default, all VLANs on Huawei switches are mapped to MSTI 0.

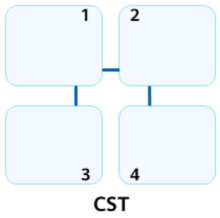
- MSTP maps VLANs to MSTIs in the VLAN mapping table.
  - Each VLAN can be mapped to only one MSTI. This means that traffic of a VLAN can be transmitted in only one MSTI. An MSTI, however, can correspond to multiple VLANs.

CST



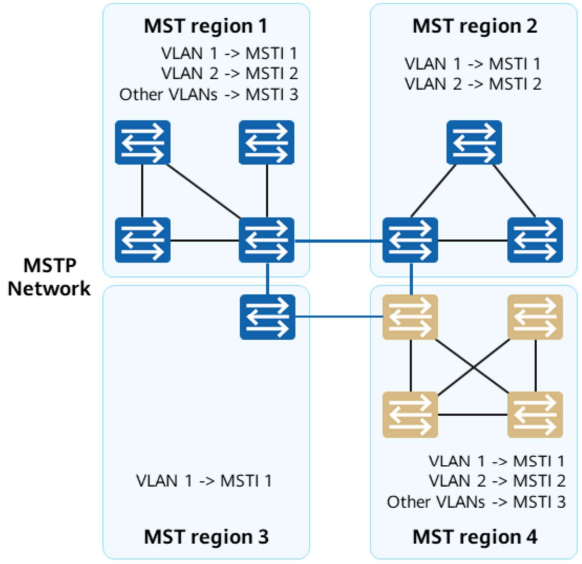
Common Spanning Tree (CST)

- A CST connects all MST regions on a switching network.
- The CST is calculated using a spanning tree protocol, with each MST region being considered as a single node.
- In the figure, the regions that are connected through dark blue thick lines form a CST.



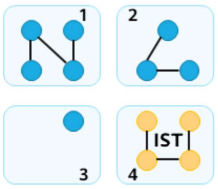
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IST



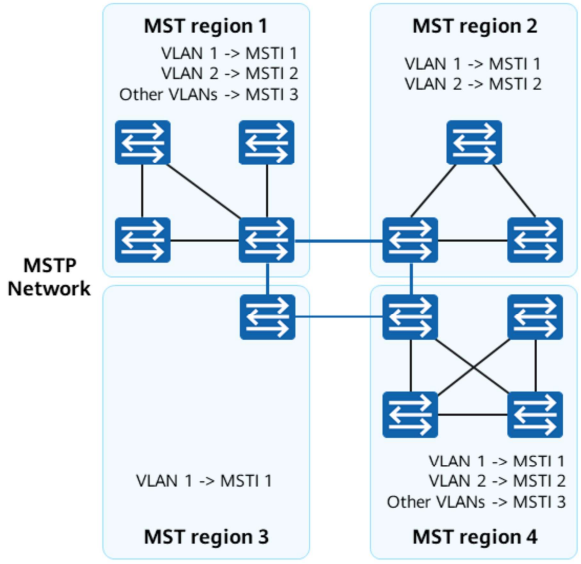
Internal Spanning Tree (IST)

- An IST resides within an MST region.
- An IST is a special MSTI with an MSTI ID of 0.
- In the figure, the switches that are connected through black thin lines in MST region 4 form an IST.



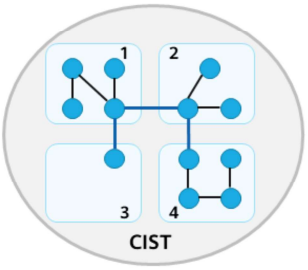
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

CIST



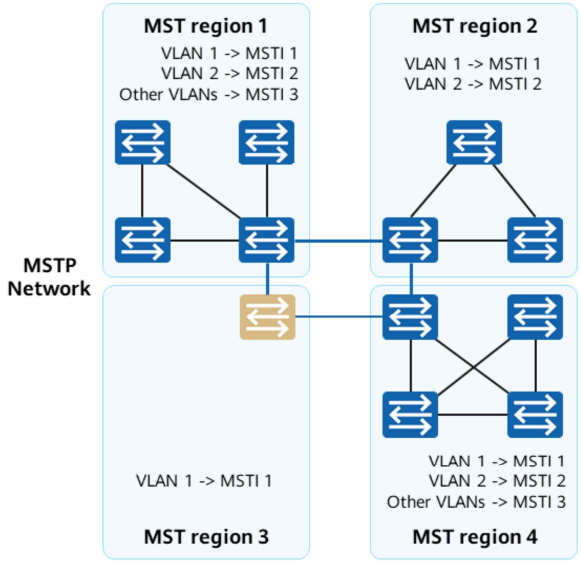
Common and Internal Spanning Tree (CIST)

- A CIST connects all the switches on a switching network and is calculated using a spanning tree protocol.
- As shown in the figure, all ISTs and the CST form a CIST.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SST

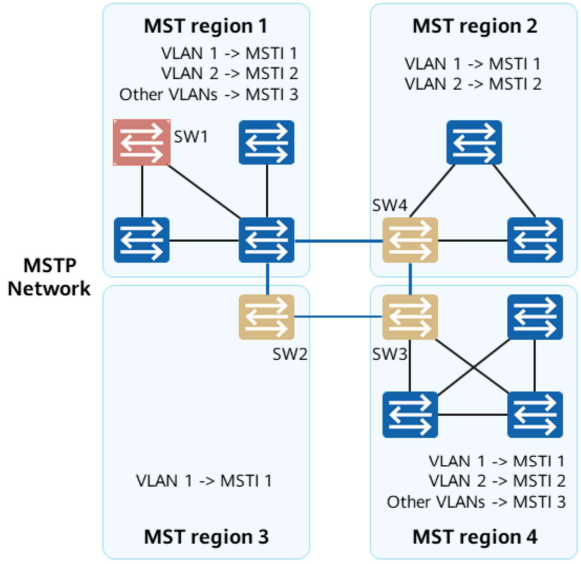


Single Spanning Tree (SST)

- A switch running a spanning tree protocol belongs to only one spanning tree.
- An MST region has only one switch.
- As shown in the figure, a switch in MST region 3 forms an SST.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

CIST Root, Regional Root, and Master Bridge



- CIST root
  - The CIST root is the root bridge of the CIST, for example, SW1 in the figure.
- Regional root
  - Regional roots are classified into IST and MSTI regional roots.
  - The switches that are closest to the CIST root are IST regional roots, for example, SW2, SW3, and SW4 in the figure.
  - An MSTI regional root is the root of the MSTI.
- Master bridge
  - The master bridge is the switch closest to the CIST root in a region, for example, SW1, SW2, SW3, and SW4 in the figure.
  - If the CIST root is in an MST region, the CIST root is the master bridge of the region.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The master bridge consists of the CIST root and the IST regional root.



Summary

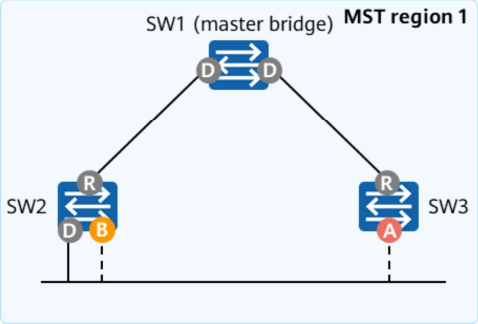
Role	Description
MST region	A switching network is divided into multiple regions. An MST region can contain one or more switches. The switches in the same MST region must be configured with the same region name, revision level, and VLAN mapping table.
MSTI	Instance-based spanning tree
VLAN mapping table	VLAN-MSTI mappings
CST	A spanning tree that connects all MST regions
IST	Internal spanning tree with the MSTI ID of 0 in an MST region
CIST	It connects all switching devices on a switching network.
SST	There is only one switching device in the MST region, and the switching device belongs to only one spanning tree.
CIST root	Root bridge of the CIST
IST regional root	Switch that is closest to the CIST root in an MST region
MSTI regional root	Root bridge in the MSTI
Master bridge	Switching device nearest to the CIST root, including the CIST root and IST regional root

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MSTP Port Roles (1)

MSTP defines the following port roles:

- Root port, designated port, alternate port, backup port, master port, regional edge port, and edge port.



Port Role	Description
Root port	A root port sends data to a root bridge and is the port closest to the root bridge.
Designated port	The designated port on a switch forwards BPDUs to a downstream switch.
Alternate port	Alternate ports provide an alternate path to the root bridge. This path is different from the path through the root port. An alternate port is blocked from sending BPDUs after a BPDU sent by another bridge is received.
Backup port	Backup ports provide a backup path to a segment already connected by a designated port. Backup ports are blocked from sending BPDUs after a BPDU sent by itself is received.

R Root port D Designated port A Alternate port B Backup port

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

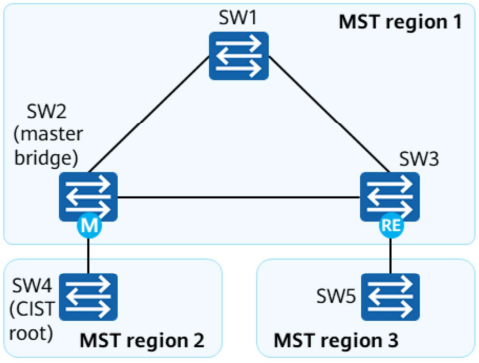
HUAWEI

- Except edge ports, all ports participate in MSTP calculation.
- A port can play different roles in different MSTIs.

MSTP Port Roles (2)

MSTP defines the following port roles:

- Root port, designated port, alternate port, backup port, master port, regional edge port, and edge port.



Port Role	Description
Master port	A master port is on the shortest path connecting MST regions to the CIST root. ▫ BPDUs of an MST region are sent to the CIST root through the master port. ▫ Master ports are special regional edge ports, functioning as root ports on ISTs or CISTs and master ports in instances.
Regional edge port	A regional edge port is located at the edge of an MST region and connects to another MST region or an SST.

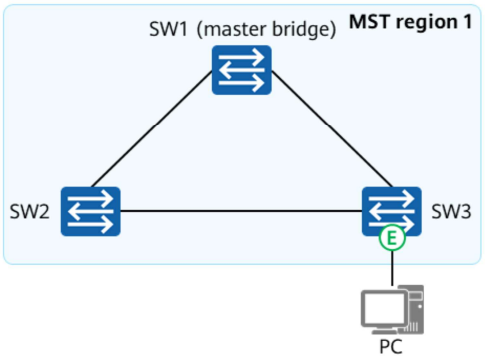
M Master port RE Regional Edge port

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MSTP Port Roles (3)

MSTP defines the following port roles:

- Root port, designated port, alternate port, backup port, master port, regional edge port, and edge port.



Port Role	Description
Edge port	An edge port is located at the edge of an MST region and does not connect to any switching device. Generally, edge ports are directly connected to terminals.

E Edge port

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MSTP Port States

MSTP port states are the same as those used in RSTP.

- Forwarding: A port in this state can send and receive BPDUs. It can also forward user traffic and learns MAC addresses.
- Learning: A port in this state can send and receive BPDUs. It learns MAC addresses but cannot forward user traffic.
- Discarding: A port in this state only receives BPDUs. It does not forward user traffic or learn MAC addresses.

MSTP Port State	Port Role
Forwarding	Root port, designated port, master port, and regional edge port
Learning	Root port, designated port, master port, and regional edge port
Discarding	Root port, designated port, master port, regional edge port, alternate port, and backup port

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

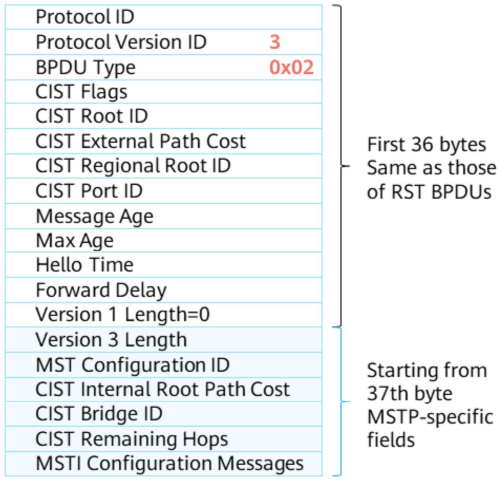
- A port in Learning state learns MAC addresses from user traffic to construct a MAC address table.

MST BPDUs

- MSTP calculates spanning trees based on Multiple Spanning Tree Bridge Protocol Data Units (MST BPDUs).
- Switches on an MSTP network transmit MST BPDUs to calculate spanning tree topologies, maintain network topologies, and communicate topology changes.

Version	Type	Name
0	0x00	Configuration BPDU
0	0x80	TCN BPDU
2	0x02	RST BPDU
3	0x02	MST BPDU

- Format of an MST BPDU



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

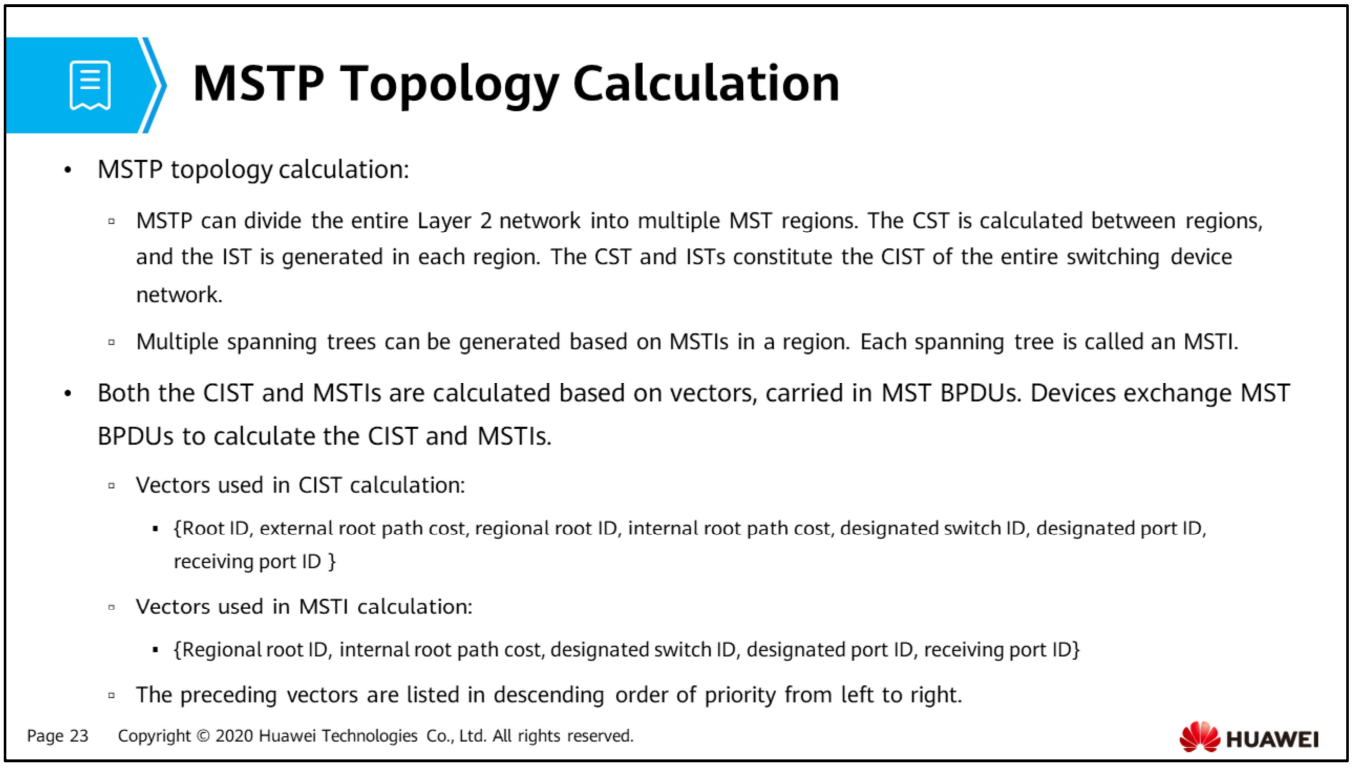
- The first 36 bytes of an MST BPDU are the same as those of an RST BPDU. Fields starting from the 37th byte of an MST BPDU are MSTP-specific. The MSTI Configuration Messages field consists of configuration messages of multiple MSTIs.
- Main fields in an MST BPDU:
  - Protocol Identifier: has 2 bytes and identifies a protocol.
  - Protocol Version Identifier: has 1 byte and indicates the protocol version identifier.
    - 0: STP
    - 2: RSTP
    - 3: MSTP
  - BPDU type: has 1 byte and indicates the BPDU type.
    - 0x00: Configuration BPDU for STP
    - 0x80: Topology Change Notification (TCN) BPDU for STP
    - 0x02: Rapid Spanning Tree (RST) BPDU or Multiple Spanning Tree (MST) BPDU
  - CIST Flags: has 1 byte and indicates the CIST flag field.
  - CIST Root Identifier: has 8 bytes and indicates the ID of the CIST root switch.
  - CIST External Path Cost: has 4 bytes and indicates the total path cost from the MST region where the switch resides to the MST region where the CIST root switch resides. This value is calculated based on link bandwidth.

- CIST Regional Root Identifier: has 8 bytes and indicates the ID of the regional root switch on the CIST. If the root is in this region, the CIST Regional Root Identifier is the same as the CIST Root Identifier.
- CIST Port Identifier: has 2 bytes and indicates the ID of the designated port in the IST.
- Message Age: has 2 bytes and indicates the lifecycle of a BPDU.
- Max Age: has 2 bytes and indicates the maximum lifecycle of a BPDU. If the Max Age timer expires, the link to the root is considered faulty.
- Hello Time: has 2 bytes and indicates the value of the Hello timer. The default value is 2s.
- Forward Delay: has 2 bytes and indicates the value of the Forward Delay timer. The default value is 15s.
- Version 1 Length: has 1 byte and indicates the BPDUv1 length, which has a fixed value of 0.
- Version 3 Length: has 2 bytes and indicates the BPDUv3 length.
- MST Configuration Identifier: has 51 bytes and indicates the MST configuration identifier, which has four fields.
- CIST Internal Root Path Cost: has 4 bytes and indicates the total path costs from the local port to the IST master. This value is calculated based on link bandwidth.
- CIST Bridge Identifier: has 8 bytes and indicates the ID of the designated switch in the CIST.
- CIST Remaining Hops: has 1 byte and indicates the number of remaining hops of a BPDU in the CIST.
- MSTI Configuration Messages: has 16 bytes and indicates an MSTI configuration message. Each MSTI configuration message occupies 16 bytes. If there are n MSTIs, MSTI configuration messages are n*16 bytes long.

Contents

1. Introduction to MSTP
2. Basic Concepts of MSTP
3. **Working Mechanism of MSTP**
4. MSTP Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

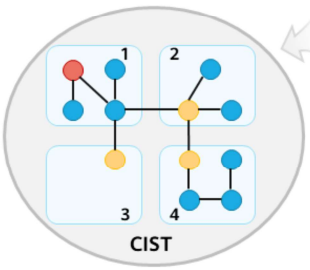


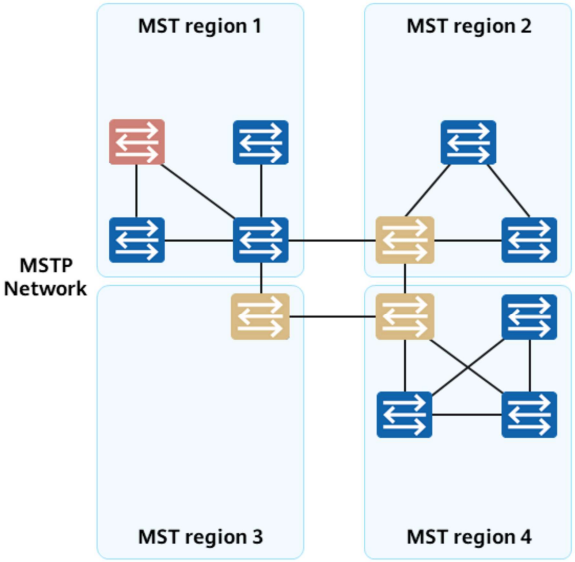
- Vectors are described as follows:
  - Root ID: identifies the root switch for the CIST.
    - The root identifier consists of the priority value (16 bits) and MAC address (48 bits).
    - The priority value is the priority of MSTI 0.
  - External root path cost (ERPC): indicates the path cost from a CIST regional root to the root.
    - ERPCs are the same on all switches in an MST region.
    - If the CIST root is in an MST region, all ERPCs in that MST region are set to 0.
  - Regional root ID: identifies the MSTI regional root.
    - It consists of the priority value (16 bits) and MAC address (48 bits).
    - The priority value is the priority of MSTI 0.
  - Internal root path cost (IRPC): indicates the path cost from the local bridge to the regional root.
    - The IRPC saved on a regional edge port must be greater than the IRPC saved on a non-regional edge port.

    - Designated switch ID: identifies the nearest upstream bridge on the path from the local bridge to the regional root (designated switch of the CIST or MSTI).
        - If the local bridge is the root or the regional root, this ID is the same as the local bridge ID.
    - Designated port ID: identifies the port on the designated switch connected to the root port on the local bridge.
        - The designated port ID consists of the priority value (4 bits) and port number (12 bits).
        - The priority value must be a multiple of 16.
    - Receiving port ID: identifies the port receiving the BPDU.
        - The receiving port ID consists of the priority value (4 bits) and port number (12 bits).
        - The priority value must be a multiple of 16.
- A smaller vector has a higher priority.
- Vectors are compared based on the following rules:
    - Root IDs
    - ERPCs
    - Regional root IDs
    - IRPCs
    - Designated switch IDs
    - Designated port IDs
    - Receiving port IDs
    - If the vectors being compared are the same, the next vector in the list is compared. If the vectors being compared are different, the remaining vectors are not compared
- If the priority of a vector carried in the configuration message of a BPDU received by a port is higher than the priority of the vector in the configuration message saved on the port, the port replaces the saved configuration message with the received one. In addition, the port updates the global configuration message saved on the device. If the priority of a vector carried in the configuration message of a BPDU received on a port is equal to or lower than the priority of the vector in the configuration message saved on the port, the port discards the BPDU.

CIST Calculation

- After comparing the vectors, the switch with the highest priority on the entire network is selected as the CIST root.
- MSTP calculates an IST for each MST region, treats each MST region as a single device, and calculates a CST to interconnect MST regions. The CST and ISTs form a CIST for the entire network.

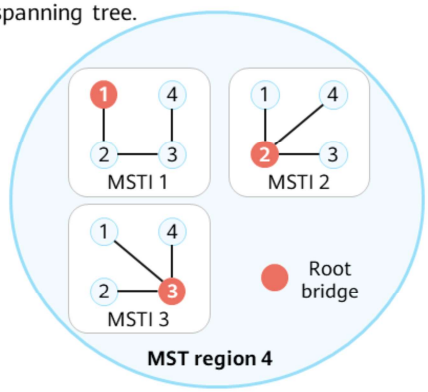


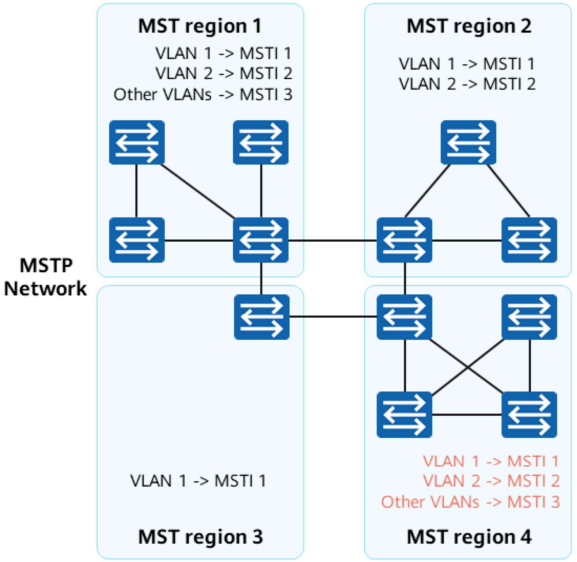


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

MSTI Calculation

- In an MST region, MSTP independently calculates an MSTI for each VLAN based on mappings between VLANs and MSTIs.
- The calculation process is similar to that used by STP to calculate a spanning tree.





Page 26 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

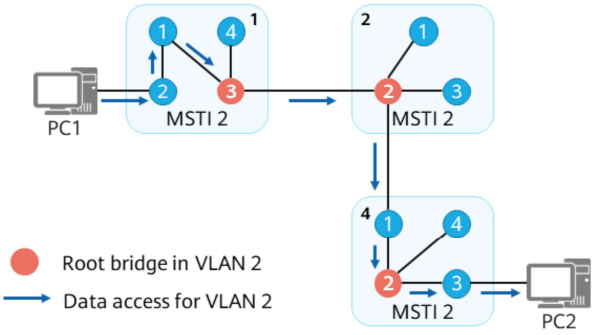
HUAWEI

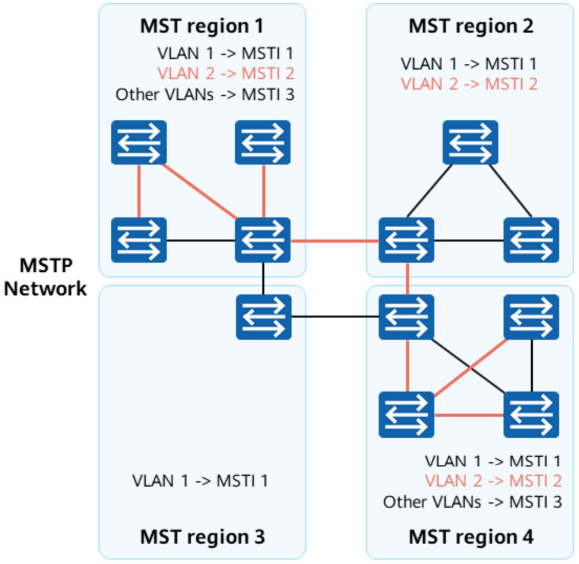
- MSTI characteristics:
  - Spanning trees of MSTIs are independent of each other.
  - The spanning tree calculation method of each MSTI is similar to that of STP.
  - Spanning trees of MSTIs can have different roots and topologies.
  - Each MSTI sends BPDUs in its spanning tree.
  - The topology of each MSTI is determined using commands.
  - A port can be configured with different parameters for different MSTIs.
  - A port can play different roles or have different states in different MSTIs.
- MSTP can calculate the root bridge or you can manually configure the root bridge or secondary root bridge for a specified spanning tree.
  - A switch can function as a root bridge or a secondary root bridge in a spanning tree. It can also function as the root bridge or secondary root bridge of another spanning tree. In a spanning tree, a device can function as either the root bridge or secondary root bridge.
  - In a spanning tree: Only one root bridge takes effect. If two or more root bridges are specified in a spanning tree, the device with the smallest MAC address is used.
  - Multiple secondary root bridges can be specified. When the root bridge fails or is powered off, a secondary root bridge becomes the new root bridge unless a new root bridge is specified. If there are multiple secondary root bridges, the one with smallest MAC address becomes the root bridge of the spanning tree.

MSTP Network Data Forwarding

On an MSTP network, a VLAN packet is forwarded as follows:

- Along MSTI in an MST region
- Along CST among MST regions





Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

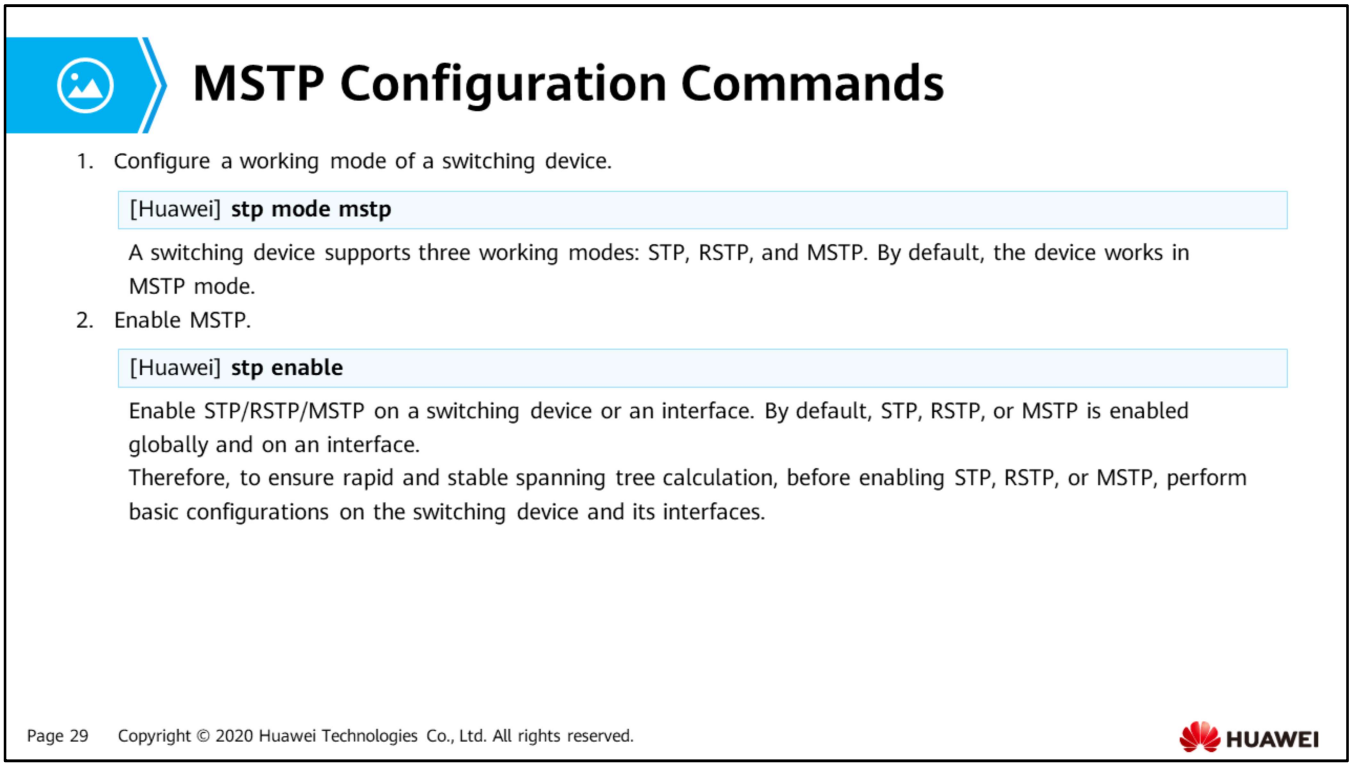
HUAWEI

- Data transmission in VLAN 2 is used as an example.

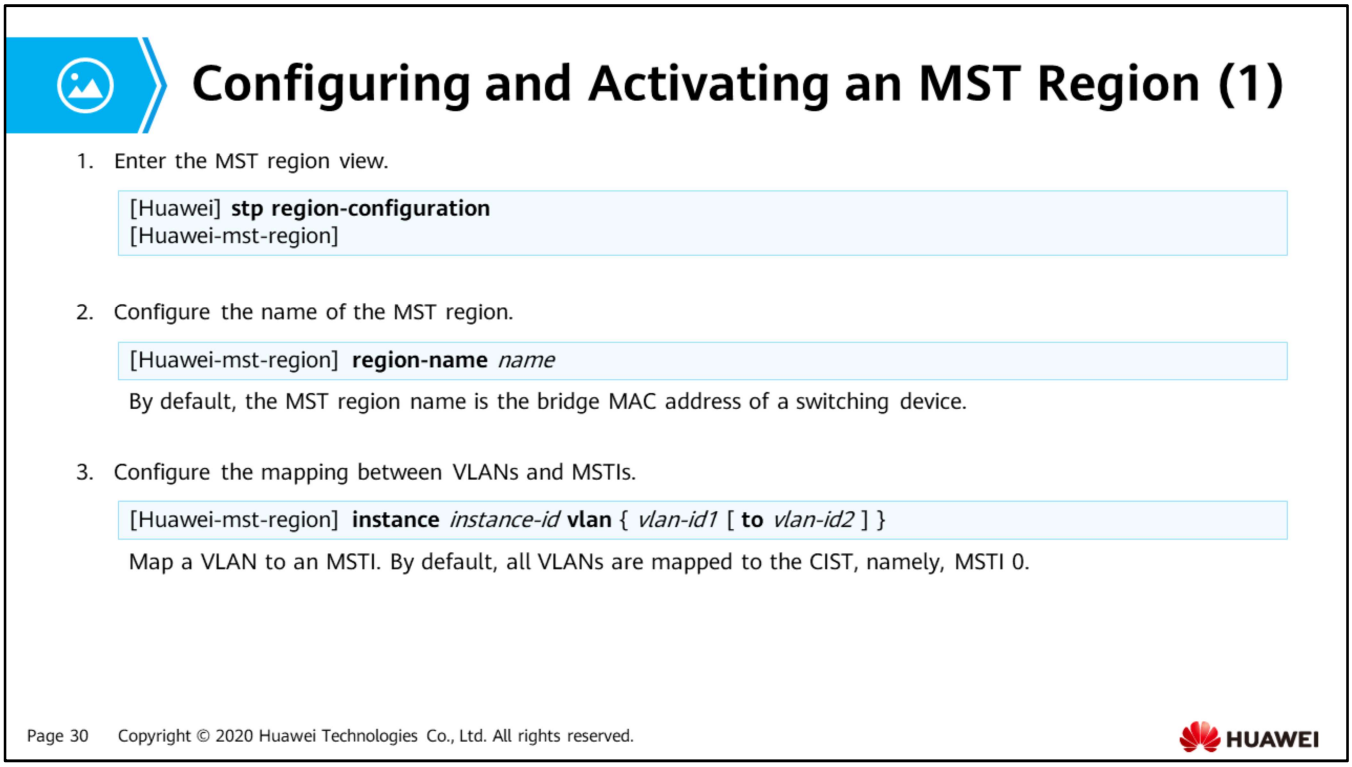
Contents

1. Introduction to MSTP
2. Basic Concepts of MSTP
3. Working Mechanism of MSTP
4. **MSTP Configurations**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- **stp mode mstp**
  - MSTP can recognize RSTP BPDUs and, conversely, RSTP can recognize MSTP BPDUs. However, MSTP and STP cannot recognize each other's BPDUs. To enable devices running different spanning tree protocols to interwork with each other, interfaces of an MSTP-enabled switch connected to devices running STP automatically transition to STP mode; other interfaces continue to work in MSTP mode.



- **stp region-configuration**
  - By default, three parameters of MST regions use default settings.
- **region-name** *name*
  - Specifies the region name of a switching device. The value is a case-sensitive string of 1 to 32 characters without spaces.
- **instance** *instance-id* **vlan** *{ vlan-id1* [**to** *vlan-id2*]
  - instance-id specifies the ID of an MSTI. The value is an integer that ranges from 0 to 4094. The value 0 indicates the CIST.

Configuring and Activating an MST Region (2)

4. (Optional) Configure the revision level of the MST region.

   [Huawei-mst-region] **revision-level** *level*

   Configure the revision level of the MST region for a switching device. By default, the revision level of an MST region is 0.

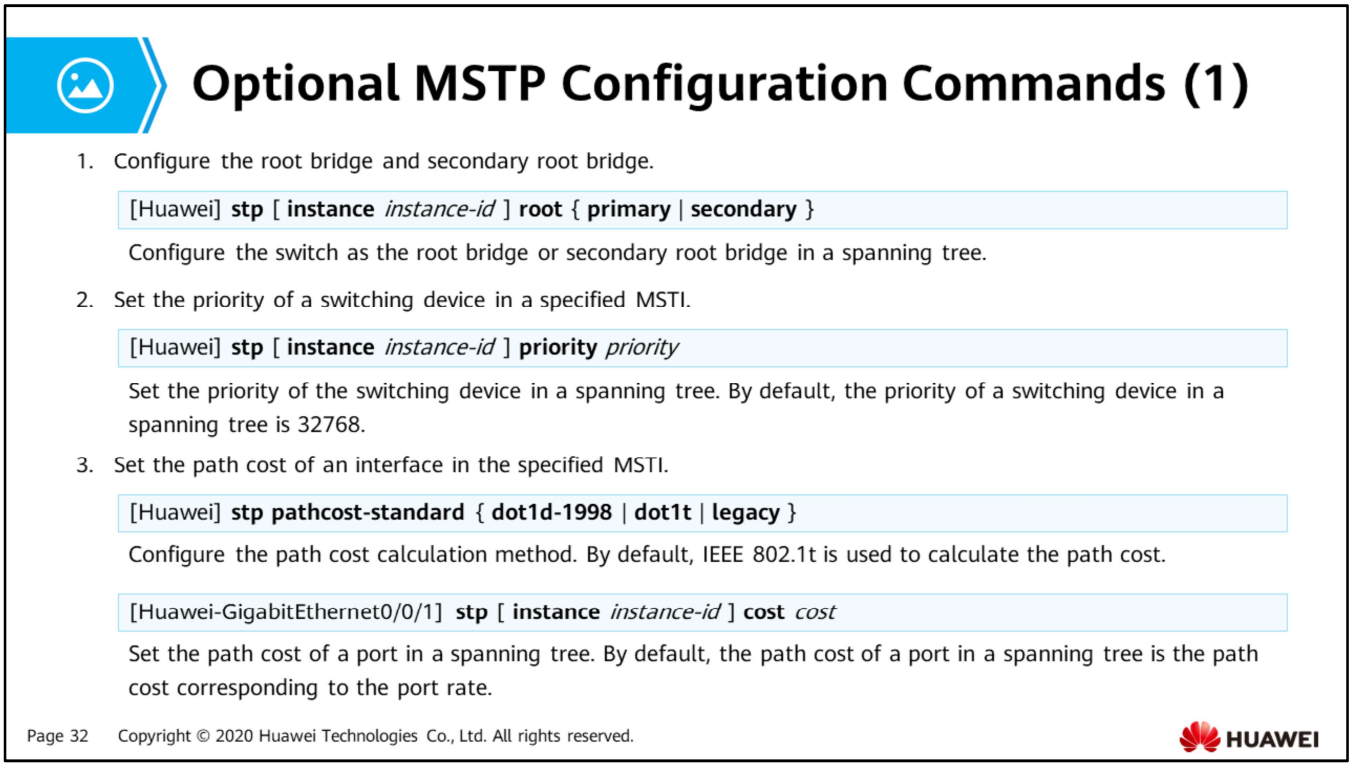
5. Activate the configuration of the MST region.

   [Huawei-mst-region] **active region-configuration**

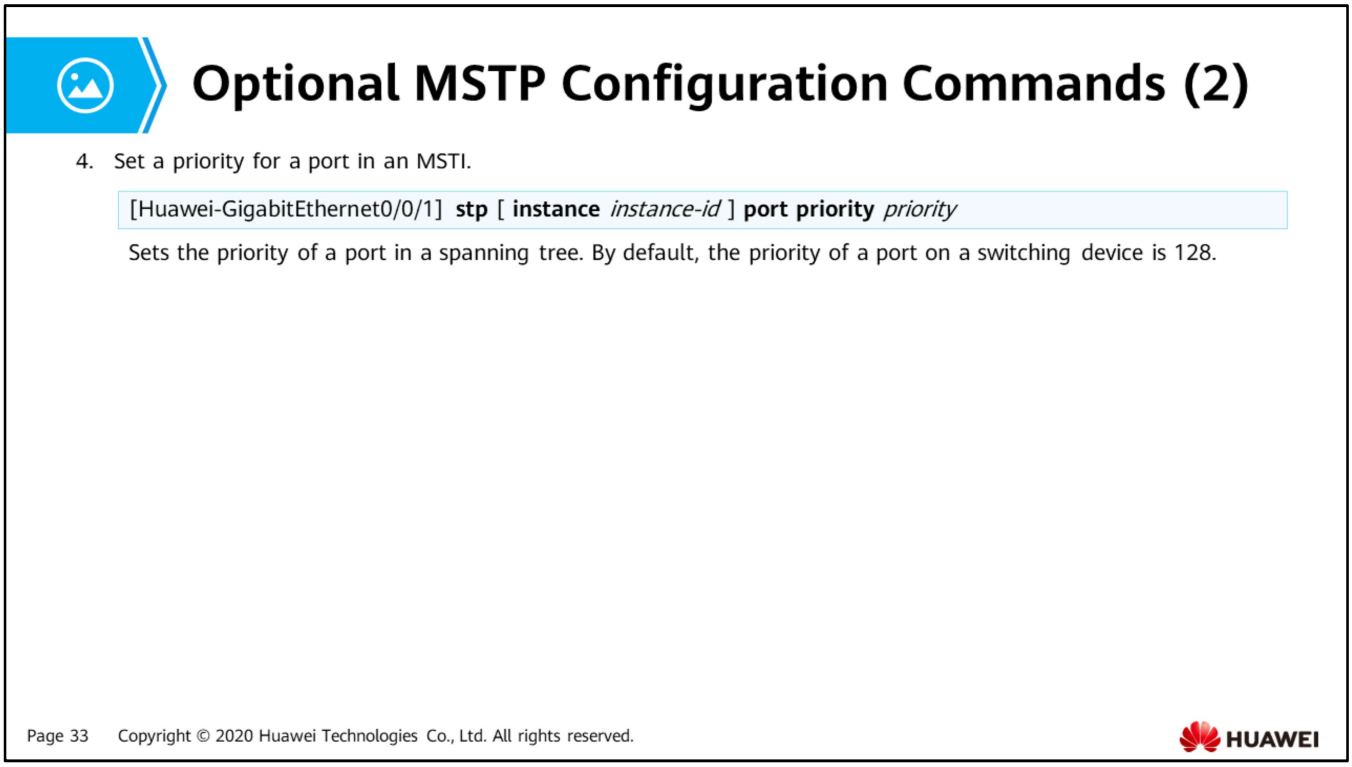
   Make the region name, VLAN mapping table, and MSTP revision level take effect.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- **revision-level** *level*
  - *level* specifies the revision level of an MST region. The value is an integer that ranges from 0 to 65535.
  - MSTP is a standard protocol; therefore, the MSTP revision level of a device is 0 by default. If the revision level of some devices from a specified manufacturer is not 0, you must change the value to 0 to facilitate tree calculation in an MST region.

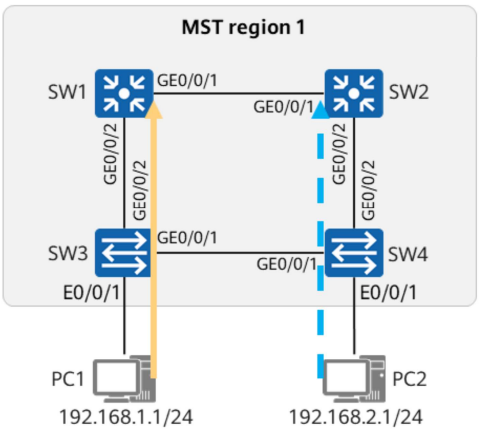


- **stp [instance** *instance-id* **] root { primary | secondary** }
  - **instance** *instance-id* specifies the ID of an MSTI. If instance instance-id is not specified, the device in MSTI 0 is a root bridge or a secondary root bridge.
  - **primary** indicates that the device functions as the root bridge of a spanning tree. After the configuration is complete, the priority of the device is 0 and cannot be changed.
  - **secondary**: indicates that the device functions as the secondary root bridge of a spanning tree. After the configuration is complete, the priority of the device is 4096 (this value cannot be modified).
- **stp [instance** *instance-id* **] priority** *priority*
  - *priority* specifies the priority of a switching device. A smaller priority value indicates a higher priority of the switching device. The value is an integer that ranges from 0 to 61440 and is a multiple of 4096, such as 0, 4096, and 8192. The default value is 32768.
- **stp pathcost-standard** { **dot1d-1998 | dot1t | legacy** }
  - **dot1d-1998**: uses the IEEE 802.1d-1998 standard to calculate the path cost.
  - **dot1t:** uses the IEEE 802.1t standard to calculate the path cost. The value ranges from 1 to 200,000,000.
  - **legacy**: uses the Huawei standard to calculate the path cost. The value ranges from 1 to 200,000.



- **stp [instance** *instance-id* **] port priority** *priority*
  - *priority:* specifies the priority of a port when it participates in spanning tree calculation. The value is an integer that ranges from 0 to 240 and is a multiple of 16, such as 0, 16, and 32.

Case: Single-Region Multi-Instance Configuration (1)

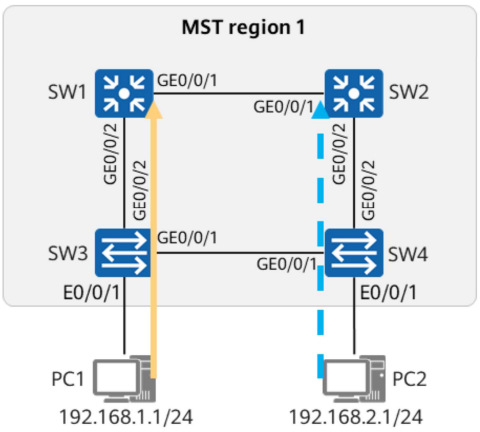


- Scenario:
  - To implement redundancy on a complex network, network designers tend to deploy multiple physical links between two devices, one of which is the primary link and the others are the backup. Loops may occur in this situation. To this end, MSTP can be deployed on the network to prevent loops.
  - MSTP blocks redundant links on a Layer 2 network and trims the network into a loop-free tree. In addition, MSTP can be deployed to implement load balancing among VLANs.
- Requirements:
  - Configure MSTP on SW1, SW2, SW3, and SW4.
  - To load balance traffic from VLANs 2 and 3, configure MSTP multi-instance.
  - Configure a VLAN mapping table to associate VLANs with MSTIs.
  - Configure the port connected to the PC as the edge port because the port does not need to participate in MSTP calculation.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Configure SW1 as the gateway of VLAN 2 (VLANIF 2: 192.168.1.254/24) and SW2 as the gateway of VLAN 3 (VLANIF 3: 192.168.2.254/24). In this way, PC1 can ping VLANIF 2 of SW1 and PC2 can ping VLANIF 3 of SW2.

Case: Single-Region Multi-Instance Configuration (2)



1. **Configure interface-based VLAN assignment to implement Layer 2 communication.**

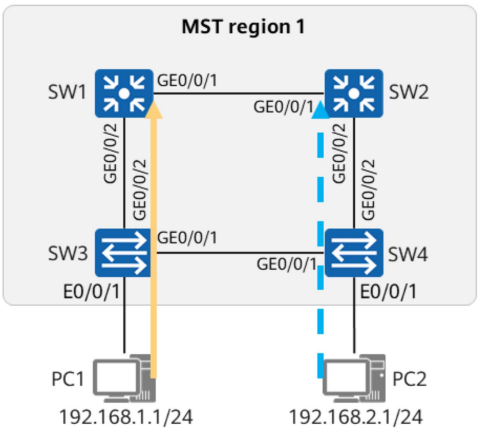
   SW1 configuration:

```
[SW1] vlan batch 2 to 3
[SW1] interface GigabitEthernet 0/0/1
[SW1-GigabitEthernet0/0/1] port link-type trunk
[SW1-GigabitEthernet0/0/1] port trunk allow-pass vlan 2 to 3
[SW1-GigabitEthernet0/0/1] quit
[SW1] interface GigabitEthernet 0/0/2
[SW1-GigabitEthernet0/0/2] port link-type trunk
[SW1-GigabitEthernet0/0/2] port trunk allow-pass vlan 2 to 3
[SW1-GigabitEthernet0/0/2] quit
```

Note: The configuration of SW2 is similar to that of SW1, and is not provided here.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Single-Region Multi-Instance Configuration (3)



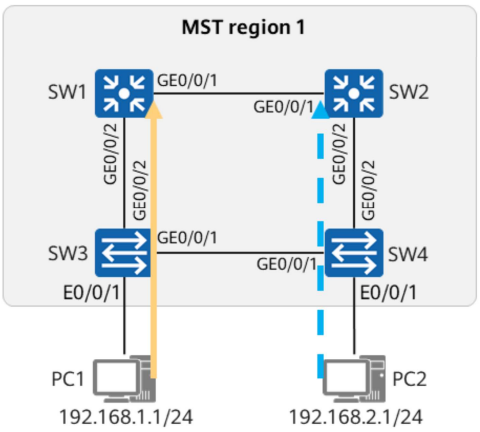
SW3 configuration:

```
[SW3] vlan batch 2 to 3
[SW3] interface GigabitEthernet 0/0/1
[SW3-GigabitEthernet0/0/1] port link-type trunk
[SW3-GigabitEthernet0/0/1] port trunk allow-pass vlan 2 to 3
[SW3-GigabitEthernet0/0/1] quit
[SW3] interface GigabitEthernet 0/0/2
[SW3-GigabitEthernet0/0/2] port link-type trunk
[SW3-GigabitEthernet0/0/2] port trunk allow-pass vlan 2 to 3
[SW3-GigabitEthernet0/0/2] quit
[SW3] interface Ethernet 0/0/1
[SW3-Ethernet0/0/1] port link-type access
[SW3-Ethernet0/0/1] port default vlan 2
[SW3-Ethernet0/0/1] quit
```

Note: The configuration of SW4 is similar to the configuration of SW3, and is not provided here.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Single-Region Multi-Instance Configuration (4)



VLAN 2 -> MSTI 1
VLAN 3 -> MSTI 2

2. Configure basic MSTP functions.

Configure an MST region and mapping between VLANs and MSTIs on SW1.

```
[SW1] stp region-configuration
[SW1-mst-region] region-name 1
[SW1-mst-region] instance 1 vlan 2
[SW1-mst-region] instance 2 vlan 3
[SW1-mst-region] active region-configuration
[SW1-mst-region] quit
```

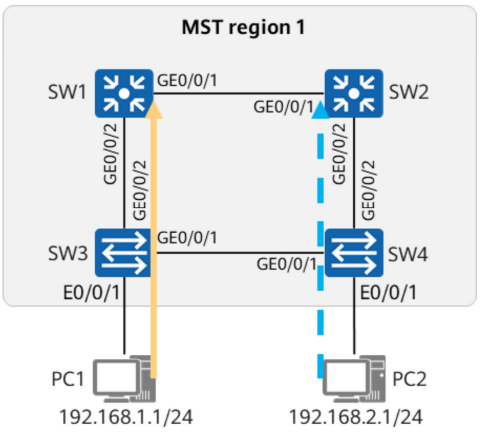
Note: The configurations of SW2, SW3, and SW4 are similar to the configuration of SW1, and are not provided here.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- By default, the MSTP function is enabled on the device. If the MSTP function is not enabled, run the **stp enable** command to enable the MSTP function on the switching device or port.

Case: Single-Region Multi-Instance Configuration (5)



3. **Configure the root bridge and secondary root bridge for MSTI 1 and MSTI 2.**

   Configure SW1 as the root bridge and SW2 as the secondary root bridge in MSTI 1.

```
[SW1] stp instance 1 root primary
[SW2] stp instance 1 root secondary
```

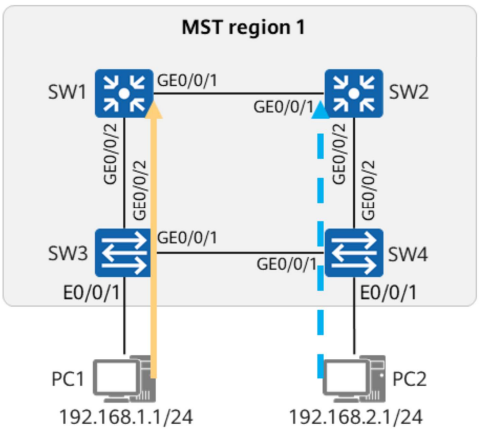
   Configure SW2 as the root bridge and SW1 as the secondary root bridge in MSTI 2.

```
[SW1] stp instance 2 root secondary
[SW2] stp instance 2 root primary
```

Note: The configurations of SW2, SW3, and SW4 are similar to the configuration of SW1, and are not provided here.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Single-Region Multi-Instance Configuration (6)



VLAN 2 -> MSTI 1
VLAN 3 -> MSTI 2

4. **Enable MSTP and configure the port connected to the PC as the edge port.**

   Configure Ethernet0/0/1 on SW3 as an edge port.

```
[SW3] interface Ethernet 0/0/1
[SW3-Ethernet0/0/1] stp edged-port enable
[SW3-Ethernet0/0/1] quit
```

Note: The edge port configuration of SW4 is similar to that of SW3, and is not provided here.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Verifying the Configuration (1)

[SW1] display stp brief

MSTID	Port	Role	STP State	Protection
0	GigabitEthernet0/0/1	DESI	FORWARDING	NONE
0	GigabitEthernet0/0/2	ROOT	FORWARDING	NONE
1	GigabitEthernet0/0/1	DESI	FORWARDING	NONE
1	GigabitEthernet0/0/2	DESI	FORWARDING	NONE
2	GigabitEthernet0/0/1	ROOT	FORWARDING	NONE
2	GigabitEthernet0/0/2	DESI	FORWARDING	NONE

[SW2] display stp brief

MSTID	Port	Role	STP State	Protection
0	GigabitEthernet0/0/1	ROOT	FORWARDING	NONE
0	GigabitEthernet0/0/2	ALTE	DISCARDING	NONE
1	GigabitEthernet0/0/1	ROOT	FORWARDING	NONE
1	GigabitEthernet0/0/2	DESI	FORWARDING	NONE
2	GigabitEthernet0/0/1	DESI	FORWARDING	NONE
2	GigabitEthernet0/0/2	DESI	FORWARDING	NONE

[SW3] display stp brief

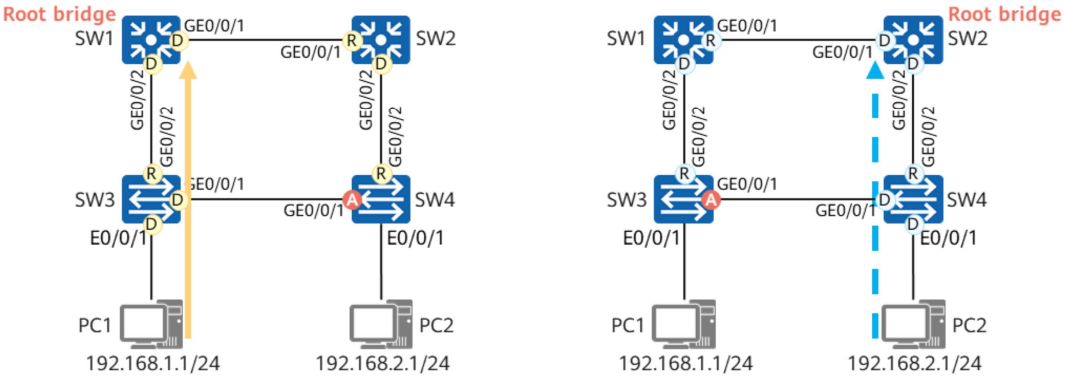
MSTID	Port	Role	STP State	Protection
0	Ethernet0/0/1	DESI	FORWARDING	NONE
0	GigabitEthernet0/0/1	DESI	FORWARDING	NONE
0	GigabitEthernet0/0/2	DESI	FORWARDING	NONE
1	Ethernet0/0/1	DESI	FORWARDING	NONE
1	GigabitEthernet0/0/1	DESI	FORWARDING	NONE
1	GigabitEthernet0/0/2	ROOT	FORWARDING	NONE
2	GigabitEthernet0/0/1	ALTE	DISCARDING	NONE
2	GigabitEthernet0/0/2	ROOT	FORWARDING	NONE

[SW4] display stp brief

MSTID	Port	Role	STP State	Protection
0	Ethernet0/0/1	DESI	FORWARDING	NONE
0	GigabitEthernet0/0/1	ROOT	FORWARDING	NONE
0	GigabitEthernet0/0/2	DESI	FORWARDING	NONE
1	GigabitEthernet0/0/1	ALTE	DISCARDING	NONE
1	GigabitEthernet0/0/2	ROOT	FORWARDING	NONE
2	Ethernet0/0/1	DESI	FORWARDING	NONE
2	GigabitEthernet0/0/1	DESI	FORWARDING	NONE
2	GigabitEthernet0/0/2	ROOT	FORWARDING	NONE

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Verifying the Configuration (2)



Root port
Designated port
Alternate port
VLAN 2 -> MSTI 1
VLAN 3 -> MSTI 2

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Quiz

1. (Single) The following figure shows port roles of a switch running MSTP. What is the state of GigabitEthernet0/0/1 in MSTI 1? ( )

   A. Blocking

   B. Discarding

   C. Forwarding

   D. Learning

[Switch] display stp brief

MSTID	Port	Role
0	Ethernet0/0/1	DESI
0	GigabitEthernet0/0/1	ROOT
0	GigabitEthernet0/0/2	DESI
1	GigabitEthernet0/0/1	ALTE
1	GigabitEthernet0/0/2	ROOT
2	Ethernet0/0/1	DESI
2	GigabitEthernet0/0/1	DESI
2	GigabitEthernet0/0/2	ROOT

2. (TorF) The CIST is a tree that consists of the ISTs and CST. ( )

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. B
2. True

Summary

- On an MSTP network, one or more VLANs can be mapped to an MSTI, and MSTP calculates the spanning tree based on the MSTI. The MSTI is an instance-based spanning tree. MSTP maintains a spanning tree for each independent MSTI. The VLANs mapped to the same MSTI share the same spanning tree.
- MSTP can implement the following functions on the Ethernet:
  - Forms multiple loop-free trees to prevent broadcast storms and implement redundancy.
  - Multiple spanning trees perform load balancing and transmit traffic in different VLANs along different paths.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Stack and CSS Features of Switches

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- As the scale of enterprise networks increase, they pose higher requirements on network reliability, fault recovery time, and device manageability.
- Traditional high reliability techniques on campus networks cannot achieve millisecond-level switchover. One active device and one standby device are typically deployed to ensure reliability; however, this deployment causes a great waste of resources. In addition, network management becomes more complex as more network devices are deployed. To build a reliable, easy-to-manage, highly efficient, and easy-to-expand switching network, the stacking and clustering techniques — stack and Cluster Switch System (CSS) — of switches are introduced.
- This document mainly describes the stack and CSS implementation and networking modes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- On completion of this course, you will be able to:
  - Describe basic concepts involved in stack and CSS.
  - Learn about the process of setting up a stack and CSS.
  - Understand the MAD mechanism of stack and CSS.

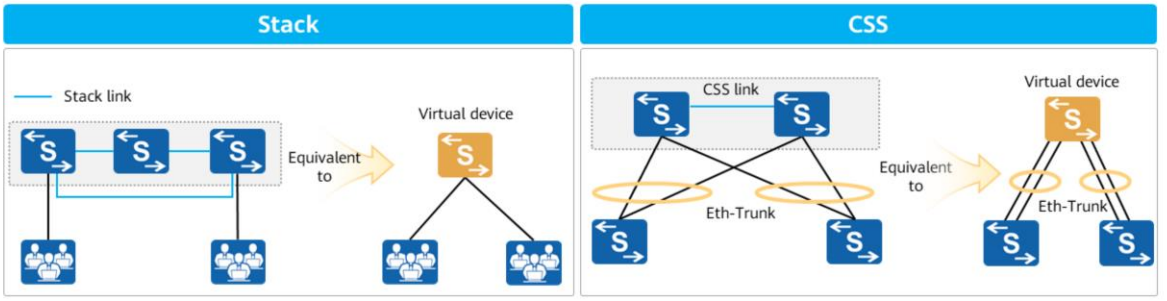
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Overview of Stack and CSS**
2. Implementation of Stack
3. Implementation of CSS
4. Basic Configuration

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

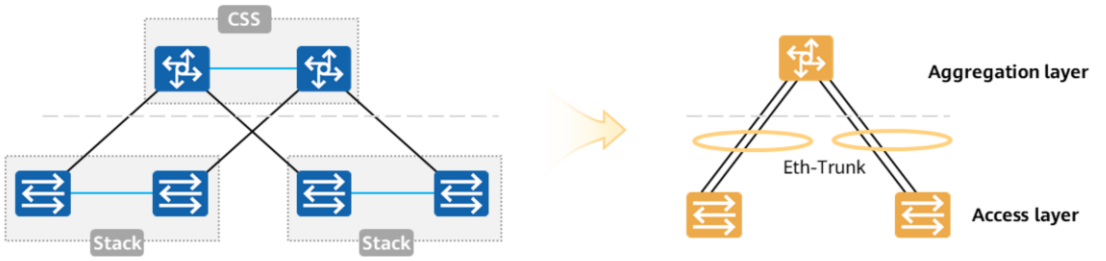
Overview of Stack and CSS



- A stack can have multiple stacking-capable switches connected through stack cables, which appear as one device when forwarding data.
- A CSS combines two clustering-capable switches into a logical switch.
- Generally, modular switches set up a CSS, and fixed switches set up a stack. Only two switches are supported in a CSS. The stack and CSS can be used together with link aggregation to build a highly reliable and loop-free campus network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Stack and CSS Architecture

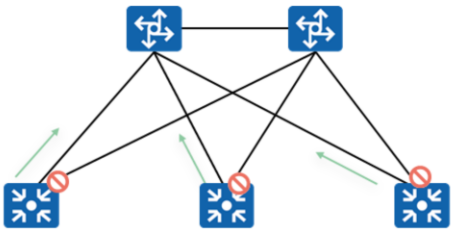
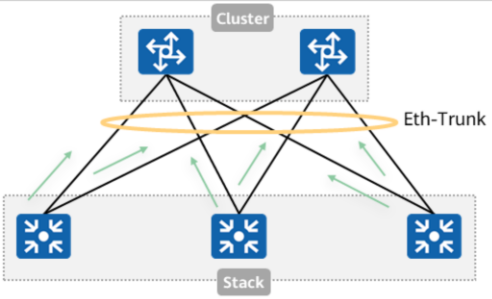


- By leveraging the stack or CSS technique, multiple independent switches are virtualized into a single logical switch. Generally, fixed switches at the access or aggregation layer set up a CSS, and modular switches at the aggregation or core layer set up a CSS.
- Eth-Trunks can be deployed between logical switches to achieve high reliability, without the need to deploy STP or VRRP.
- The use of Eth-Trunks also improves link utilization by allowing links between devices to transmit traffic simultaneously.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Advantages of Stack and CSS (1)

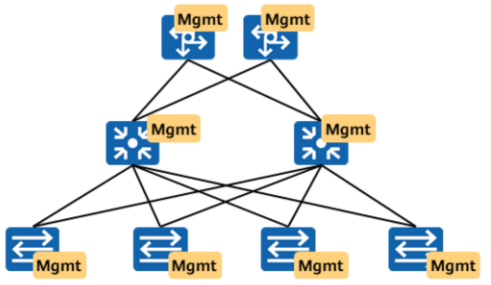
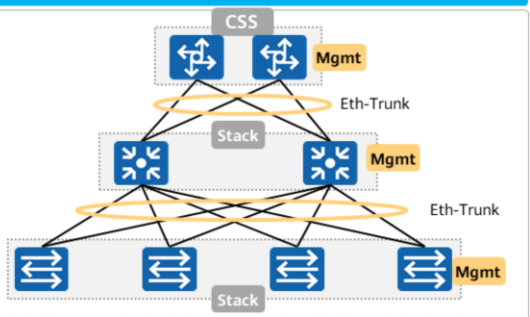
Stack and CSS effectively improve resource utilization, forwarding performance, and link bandwidth.

STP networking	Stack/CSS networking
	
• Core devices work in active/standby mode. • Dual uplinks of each aggregation device work in active/standby mode. • Low device and link utilization	• Core devices are virtualized into a logical device and they can forward traffic simultaneously. • All links between the aggregation and core layers are bundled into Eth-Trunks to forward traffic simultaneously. • High device and link utilization

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Advantages of Stack and CSS (2)

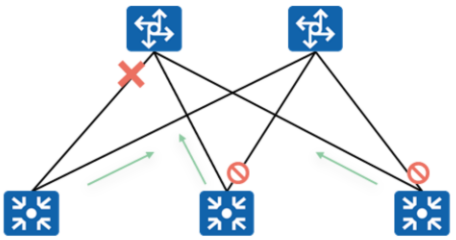
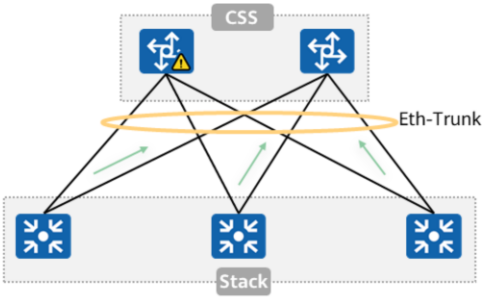
Stack and CSS simplify network planning and management.

STP networking	Stack/CSS networking
	
• It is complex to plan and configure Layer 2 VLANs and STP. • VRRP and STP are used together to implement load balancing between active and standby devices, which are complex to plan and configure. • All devices are managed and configured separately.	• Links between logical devices are bundled into Eth-Trunks, removing the need to deploy STP and VRRP. • This networking simplifies network planning and device configuration. • Only the logical devices need to be managed.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Advantages of Stack and CSS (3)

Stack and CSS greatly reduce the service interruption time.

STP networking	Stack/CSS networking
	
• If a fault occurs, for example, the active device or link is unavailable, an active/standby switchover is performed. • The switchover time depends on the STP convergence time, which is usually several seconds.	• The failure of a single device or link does not affect service forwarding. Traffic will be switched to other available member links of Eth-Trunks. • Users are unaware of the interruption.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

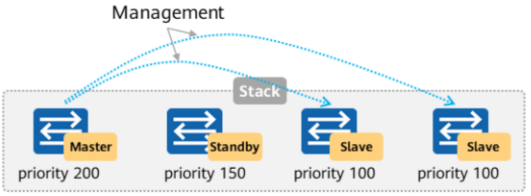
Contents

1. Overview of Stack and CSS
2. **Implementation of Stack**
3. Implementation of CSS
4. Basic Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Concepts of Stack

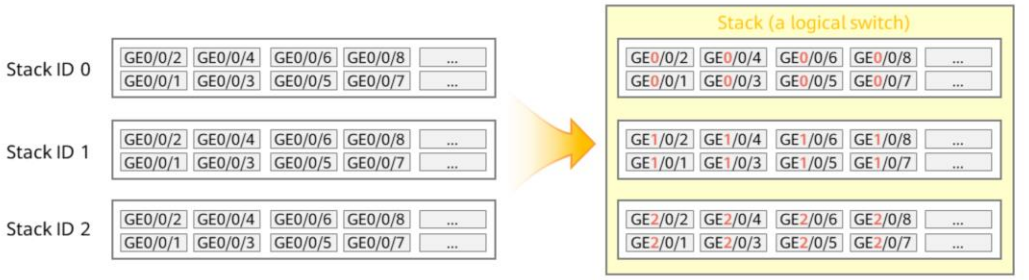
- Switches that join a stack are member switches. In a stack, member switches take on one of the three roles:
  - Master switch: manages the stack. A stack has only one master switch.
  - Standby switch: is the backup of the master switch. A stack has only one standby switch. If the master switch fails, the standby switch takes over all services from the master switch and assumes the master role.
  - Slave switch: forwards service traffic. A stack may have multiple slave switches. The more slave switches in a stack, the higher forwarding performance the stack can provide. Apart from the master and standby switches, all the other member switches are slave switches. If the standby switch becomes unavailable, a slave switch assumes the standby role.
- Stack priority: determines the role of a stack member in role election. A higher priority value for a stack member increases the probability of it being elected the stack master.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Stack ID

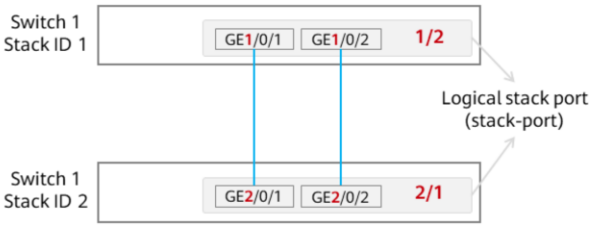
- A stack ID identifies a stack member and is the slot ID of the member. Each stack member must have a unique stack ID in the stack.
- The default stack ID of a device is 0. In a stack, the master switch manages stack IDs of other member switches. When a new member switch joins the stack and its stack ID conflicts with that of another member switch in the stack, the master switch searches for the smallest available stack ID and assigns it to the new member switch.
- You are advised to configure stack IDs for stack members before setting up a stack.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Logical Stack Port

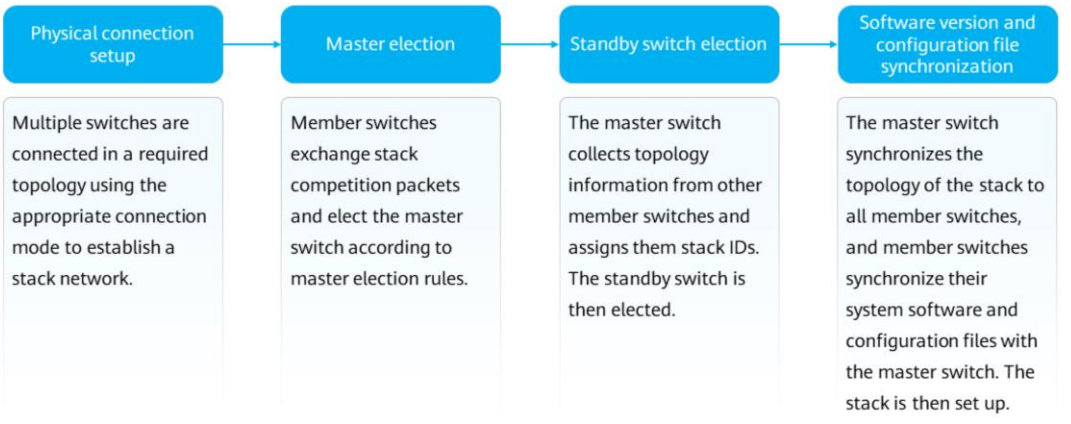
- Logical stack ports are logical ports used to set up a stack between switches. Each switch supports two logical stack ports: stack-port n/1 and stack-port n/2, where n indicates the stack ID of a member switch.
- Multiple stack member ports can be bundled into a logical stack port to improve stack reliability and bandwidth.
- Stack member ports of a logical stack port (stack-port n/1) on one switch must be connected to stack member ports of a logical stack port (stack-port m/2) on another switch.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

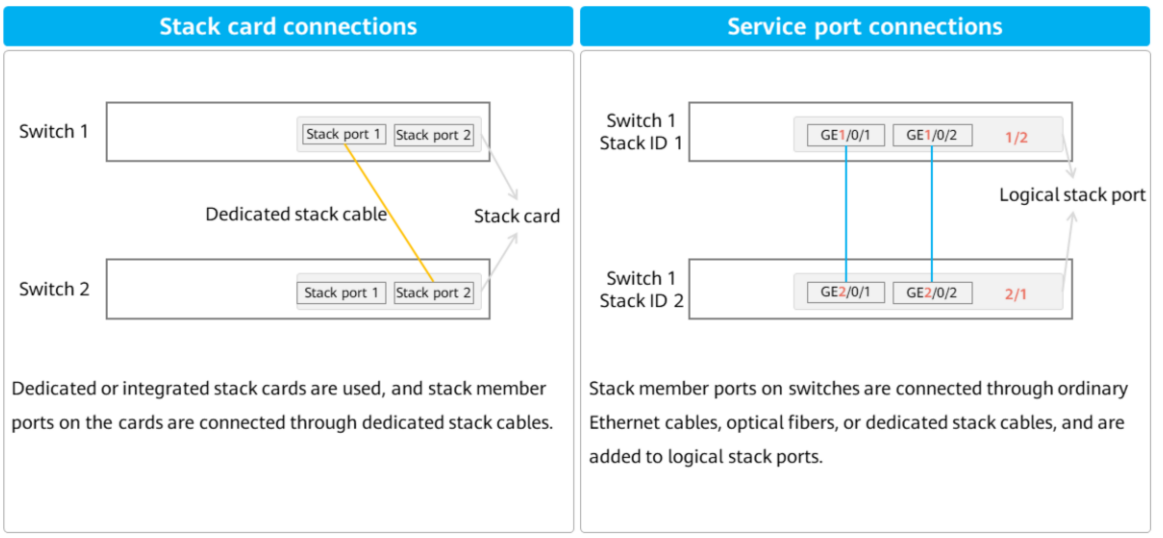
- If the logical stack ports on two ends (stack-port *n*/1 on one switch and stack-port *m*/2 on the other) both contain multiple stack member ports, the stack member ports can be connected in any sequence.

Stack Setup



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Stack Connection Modes



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Each member switch in a stack supports two logical stack ports: stack-port *n*/1 and stack-port *n*/2, where *n* indicates the stack ID of a member switch.

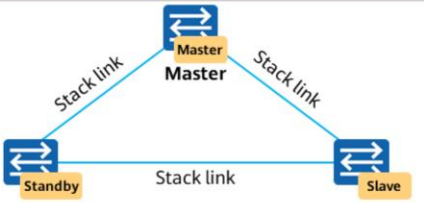
Stack Topologies

Chain topology



- Member switches are far from one another and a ring topology is difficult to deploy.
- Low reliability: If any stack link fails, the stack splits.

Ring topology



- Member switches are located near one another. This topology ensures high reliability and high link bandwidth efficiency.
- High reliability: If a stack link fails, the topology changes from ring to chain, allowing the stack to function normally.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Master Switch Election

After all member switches are connected and powered on, they elect a master switch.

Rules for electing a master switch

1	The switch that starts first becomes the master switch. The master election timeout interval is 20 seconds. The startup process may take different lengths of time on different member switches. When stack member switches are powered on or restart, some member switches may not participate in the first master election.
2	If multiple switches complete startup at the same time, the switch with the highest stack priority becomes the master switch.
3	If multiple switches complete startup at the same time and have the same stack priority, the switch with the smallest MAC address becomes the master switch.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Standby Switch Election

Topology information collection, stack ID assignment, and standby switch election

After the master switch is elected, it collects topology information from member switches and assigns a stack ID to each member switch. The standby switch is then elected to operate as a backup for the master switch.

The first switch that completes startup after the master switch becomes the standby switch.

If multiple switches complete the startup at the same time, the standby switch is elected according to the following rules:

1	The switch with the highest stack priority becomes the standby switch.
2	If the switches have the same stack priority, the one with the smallest MAC address becomes the standby switch.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

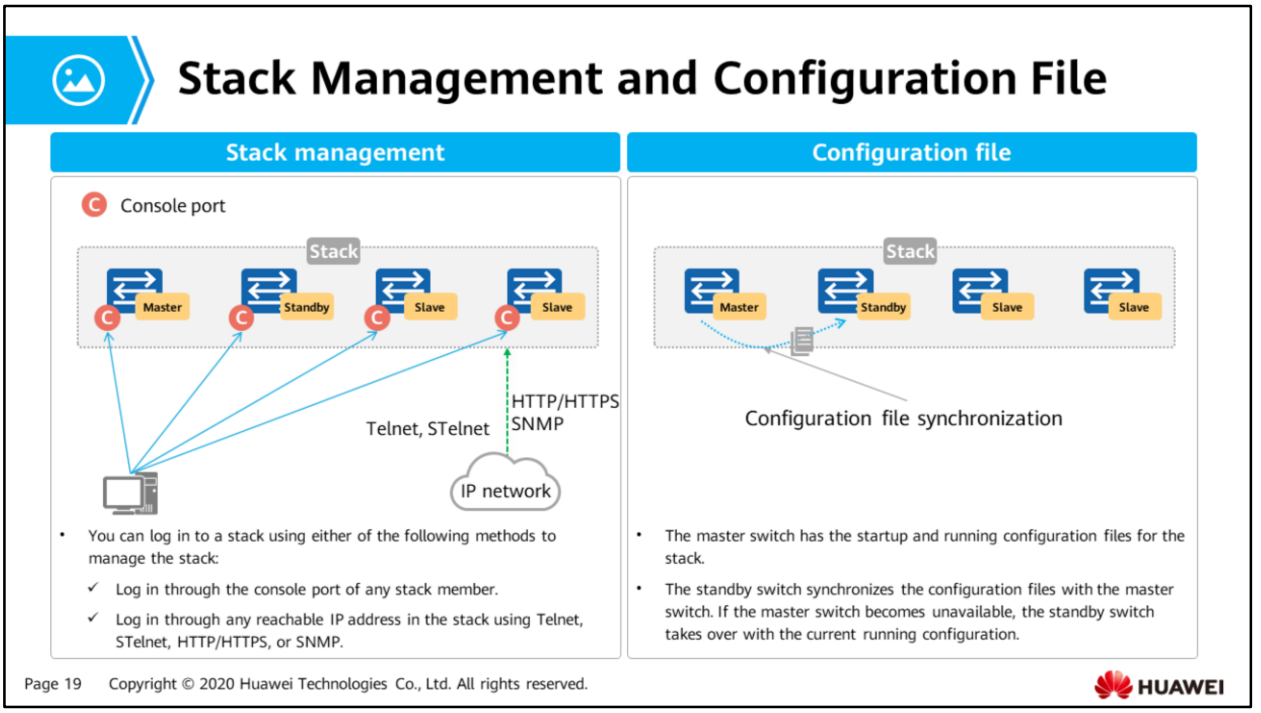
Software Version and Configuration File Synchronization

Software version and configuration file synchronization; stable running

After role election and topology information collection are complete, member switches synchronize their system software and configuration files with the master switch. The stack can now run.

1	Automatic software loading: If the standby and slave switches are running a different software version from that of the master switch, they download the system software of the master switch, restart with the new system software, and rejoin the stack. Member switches must have compatible software versions with one another to set up a stack.
2	Configuration file synchronization: The standby and slave switches download and apply the configuration files of the master switch. This enables member switches to work like a single switch and ensures that other switches continue working normally if the master switch fails.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- The configuration files record the following settings:
  - System-level (global) settings such as IP, STP, VLAN, and SNMP settings that apply to all stack members. A new switch joining a stack uses the system-level settings of that stack. Likewise, if a device is moved to a different stack, that device loses its startup configuration file and uses the system-level configuration of the new stack.
  - Stack member interface-specific settings that are specific for each stack member. The interface-specific configuration of each stack member is associated with its stack ID. If the stack ID changes, the new ID takes effect after that stack member restarts.
    - If an interface-specific configuration does not exist for that new ID, the stack member uses its default interface-specific configuration.
    - If an interface-specific configuration exists for that new ID, the stack member uses the interface-specific configuration associated with that ID.
- A switch will retain its stack configuration after leaving a stack, so it will be elected as the master switch, forming a single-switch stack. To delete the stack configuration, run the **reset stack configuration** command. The cleared configuration includes:
  - Switch slot ID
  - Stack priority
  - Reserved VLAN ID of the stack
  - System MAC address switching delay
  - Logical stack port configuration
  - Logical stack port rate
  - Note that running this command will cause the switch to restart.

Removing a Stack Member

Depending on the role of the member switch that leaves a stack, the stack is affected in the following ways:

- If the master switch leaves the stack, the standby switch becomes the new master switch. It recalculates topology information, synchronizes updated topology information to the other member switches, and selects a new standby switch. Afterwards, the stack begins to run stably.
- If the standby switch leaves the stack, the master switch selects a new standby switch, recalculates the stack topology, and synchronizes updated topology information to the other member switches. Afterwards, the stack begins to run stably.
- If a slave switch leaves the stack, the master switch recalculates the stack topology and synchronizes updated topology information to the other member switches. Afterwards, the stack begins to run stably.



An administrator removes stack cables.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A member switch leaves a stack after you disconnect its stack cables and remove it from the stack. When removing a member switch, pay attention to the following points:
  - After removing a member switch from a ring stack topology, use a stack cable to connect the two ports originally connected to this member switch to ensure network reliability.
  - In a chain topology, removing an intermediate switch will cause the stack to split. Analyze the impact on services before doing so.

Adding a Switch to a Stack

A new member switch joins a running stack after the following sequence of events:

- The new switch is connected to the stack, is powered on, and starts. It is elected as a slave switch, while the roles of other stack members remain unchanged.
- The master switch updates stack topology information, synchronizes the updated topology information to the other member switches, and assigns a stack ID to the new member switch (if the new member switch has no stack ID configured or the configured stack ID conflicts with that of another member switch).
- The new member switch updates its stack ID, synchronizes its configuration files and system software with the master switch, and then runs stably.

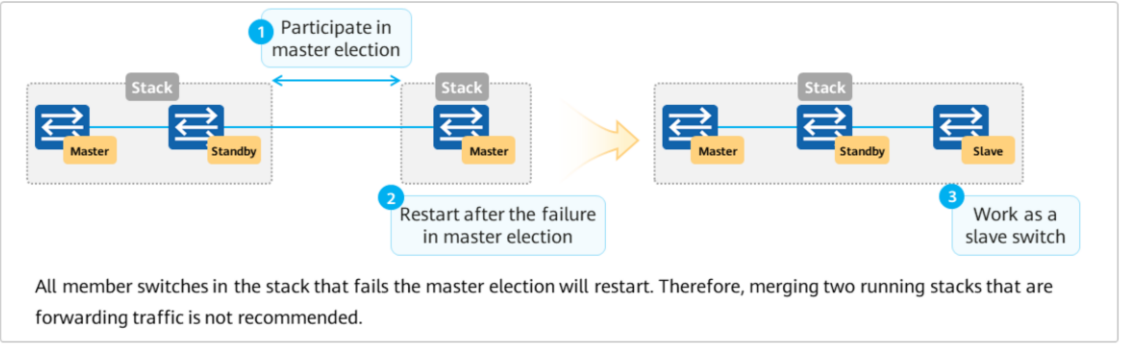


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Note the following when connecting a switch that is powered off to a stack:
  - If the stack has a chain topology, add the new switch to either end of the chain to minimize the impact on running services.
  - If the stack has a ring topology, tear down a physical link to change the ring topology to a chain topology, add the new switch to either end of the chain, and then connect the switches at two ends to form a ring.

Stack Merging

- Two running stacks can merge into a single stack.
- For example, a switch that has the stack configuration and has been powered on (forming a single-switch stack) is connected to a running stack through stack cables. This process is known as stack merging, which is different from the process of adding a member switch to a stack.



All member switches in the stack that fails the master election will restart. Therefore, merging two running stacks that are forwarding traffic is not recommended.

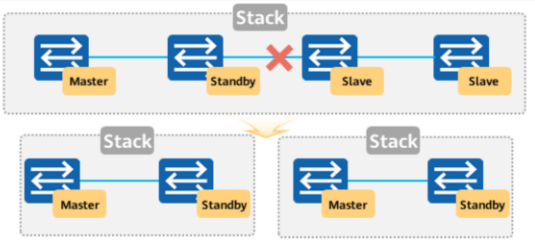
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- A single-switch stack is a standalone switch enabled with the stacking function. There is only one member switch in the stack, which operates as the master switch. Only a switch enabled with the stacking function can join a stack or set up a stack with other switches enabled with the stacking function.
- The stack merging process is as follows:
  - When two stacks merge, both master switches compete to be the master switch of the new stack.
  - After a new master switch is elected, the remaining stack members in the same stack as this new master switch retain their roles and configurations, without affecting services.
  - Switches in the other stack restart and join the new stack as slave switches. The master switch assigns new stack IDs to these switches. Then these switches synchronize their configuration files and system software with the master switch. During this period, services on these switches are interrupted.

Stack Split

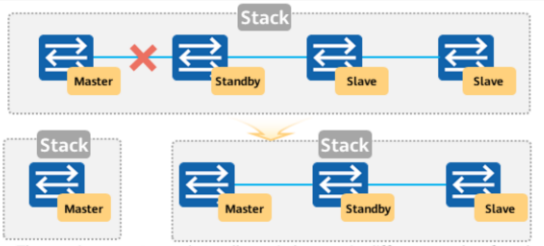
If you remove some member switches from a running stack without powering off the switches or if multiple stack cables fail, the stack splits into multiple stacks.

Scenario 1



- The previous master and standby switches are in the same stack after the stack split.
- When removed slave switches detect that the timeout timer for stack protocol packets has expired, the switches restart and begin a new master election.

Scenario 2

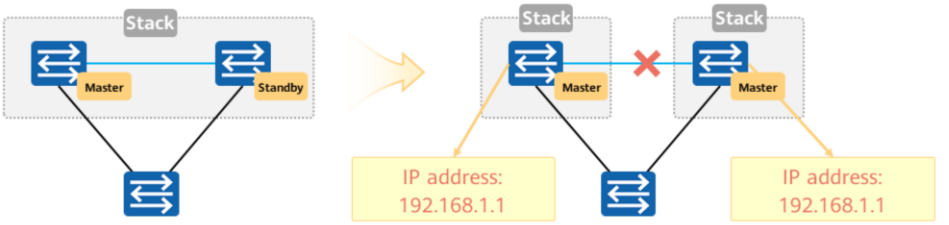


- The previous master and standby switches are in different stacks after the stack split.
- The previous master switch selects a new standby switch in its stack and updates the stack topology information.
- The previous standby switch becomes the new master switch in its stack and selects a new standby switch.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Problems Caused by a Stack Split

All member switches in a stack use the same IP address (VLANIF interface address) and MAC address (stack MAC address). After a stack splits, more than one stack may use the same IP address and MAC address. To prevent this situation, a mechanism is required to check for IP address and MAC address collision after a split.



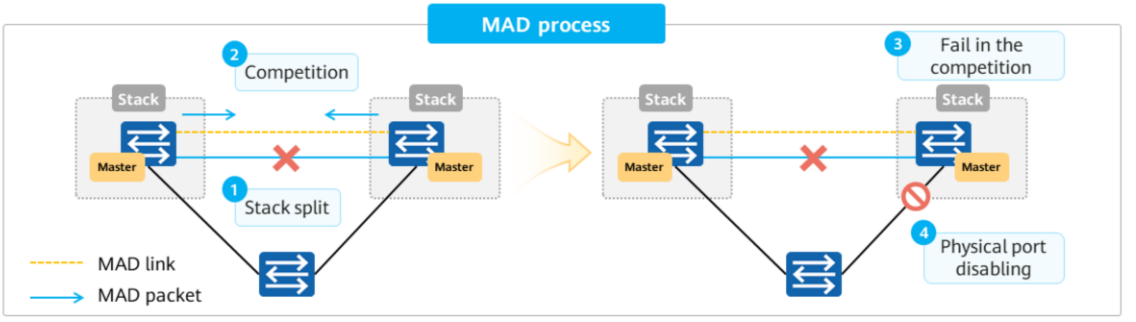
In the original stack, the IP and MAC addresses of VLANIF 1 are 192.168.1.1 and the system MAC address, respectively. After the stack splits, both the new stacks use the IP and MAC addresses of VLANIF 1.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A stack communicates with other network devices as one device using a unique MAC address. This MAC address is known as the stack MAC address.
- Generally, the stack MAC address is the MAC address of the master switch. As such, if the master switch is unavailable or leaves the stack, the stack MAC address will be changed 10 minutes later by default. That is, the MAC addresses of the two new stacks are the same within 10 minutes.

MAD

- Multi-active detection (MAD) is a stack split detection protocol. If a stack splits due to a link failure, MAD provides split detection, multi-active handling, and fault recovery mechanisms to minimize the impact of a stack split on services.
- MAD can be implemented in direct or relay mode. Direct and relay modes cannot be both configured in the same stack.



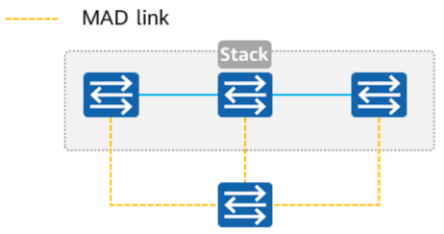
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The new stacks send MAD packets over a MAD link (an ordinary cable, which is manually configured as a MAD link) for competition. The stack that fails in the competition shuts down all physical ports (except the manually configured reserved ports) on its member switches to prevent IP or MAC address conflicts.
- The MAD process is as follows:
  - After a stack splits, the new stacks have the same IP address and MAC address (stack MAC address). As a result, network entries, such as ARP entries and MAC address entries of downstream devices, are incorrect, causing service exceptions.
  - When MAD is enabled, the new stacks compete through the MAD link.
  - The stack that fails in the competition shuts down all physical interfaces on its member switches to prevent IP or MAC address conflicts.

MAD: Direct Mode

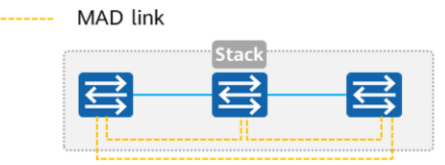
In direct mode, stack members use directly connected ordinary cables as dedicated MAD links. When the stack is running properly, member switches do not send MAD packets. After the stack splits, member switches each send a MAD packet every 1 second over a MAD link to check whether more than one master switch exists.

Directly connected to an intermediate device



All member switches are directly connected to an intermediate switch through ordinary cables.

Fully meshed with each other



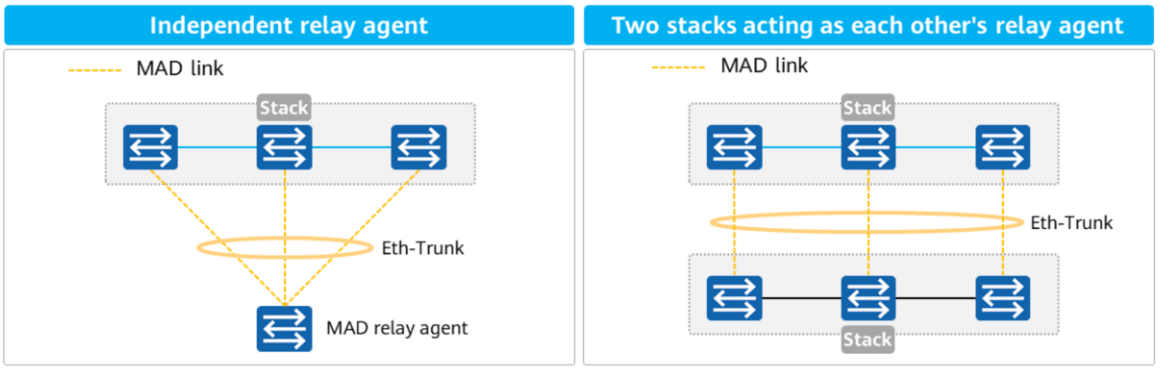
In the full-mesh topology, at least one MAD link exists between any two member switches.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The use of an intermediate device can shorten the MAD links between member switches. This topology applies to stacks with a long distance between member switches.
- The full-mesh topology prevents MAD failures caused by intermediate device failures, but occupies many interfaces on the member switches. Therefore, this topology applies to stacks with only a few member switches.

MAD: Relay Mode

In relay mode, MAD relay detection is configured on an Eth-Trunk interface in the stack, and the MAD detection function is enabled on an agent. Each member switch must have a link to the agent, and these links must be added to the same Eth-Trunk. In contrast to the direct mode, the relay mode does not require additional interfaces because the Eth-Trunk interface can run other services while performing MAD relay detection.



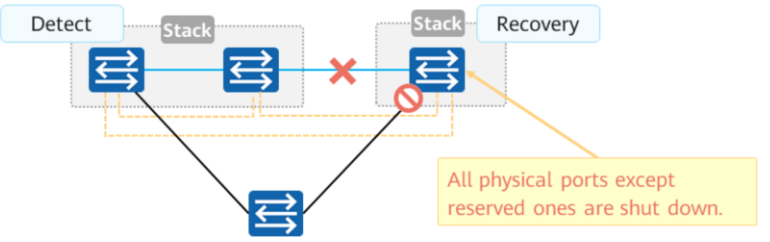
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In relay mode, when the stack is running properly, member switches send MAD packets at an interval of 30 seconds over the MAD links and do not process the received MAD packets. After the stack splits, the new stacks send MAD packets at an interval of 1 second over MAD links to check whether more than one master switch exists.

Multi-Active Handling (1)

After a stack splits, the new stacks send MAD packets to perform MAD competition, and set their states to Detect or Recovery according to the competition result.

- Detect: The competition is successful, and the stack is working properly.
- Recovery: The competition fails. The stack that fails in the competition shuts down all physical ports except the manually configured reserved ports on member switches, and no longer forward service packets.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Multi-Active Handling (2)

MAD competition rules

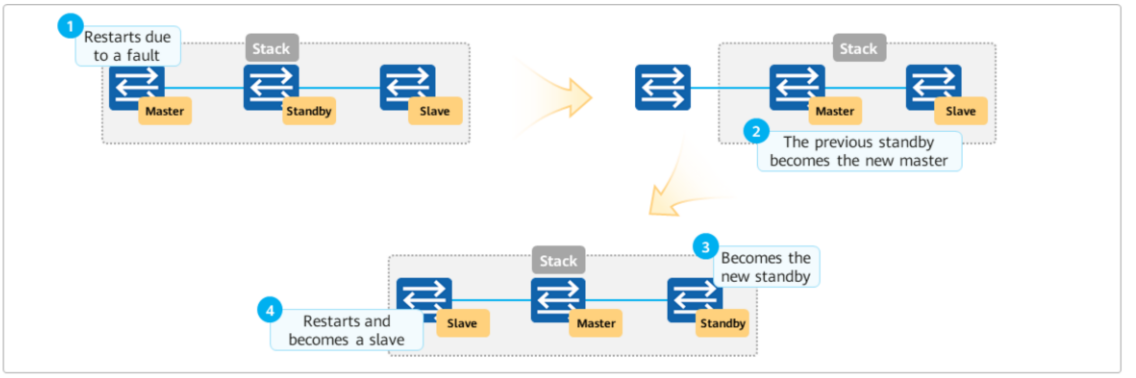
MAD competition follows similar guidelines to master switch competition.

1	The stack that completes its startup first enters the Detect state. If the difference in the time points when multiple stacks complete their startup is within 20 seconds, the stacks are considered to complete their startup at the same time.
2	When the stacks complete their startup at the same time, the stack in which the master switch has the highest stack priority enters the Detect state.
3	When the master switches in the stacks have the same stack priority, the stack with the smallest MAC address enters the Detect state.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Master/Standby Switchover

- If you want another switch in the stack to be the master switch, you can perform a master/standby switchover by running a command to change this switch to the new master switch.
- In addition, a master/standby switchover can be triggered if the master switch restarts due to a fault.



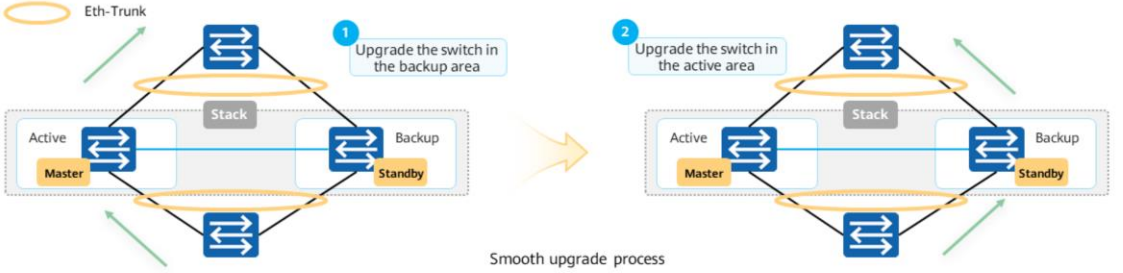
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The previous standby switch becomes the new master switch.
- The new master switch selects a standby switch.
- The previous master switch restarts, rejoins the stack, and becomes a slave switch.

Stack Upgrade

A stack supports three upgrade methods:

- Intelligent upgrade: When a stack is set up or a new switch is added to the stack, standby and slave switches or the switch automatically synchronizes the system software with the master switch.
- Traditional upgrade: similar to an ordinary device upgrade. You need to specify the startup system software on the master switch and restart the entire stack to upgrade the software of member switches. This process will cause lengthy service interruptions.
- Smooth upgrade: A stack is divided into an active area and a backup area. Member switches in the two areas are upgraded in turn. The uplinks and downlinks of the stack work in redundancy mode, and the active and standby links are in the active and backup areas, respectively. In this way, services are not interrupted during the upgrade.



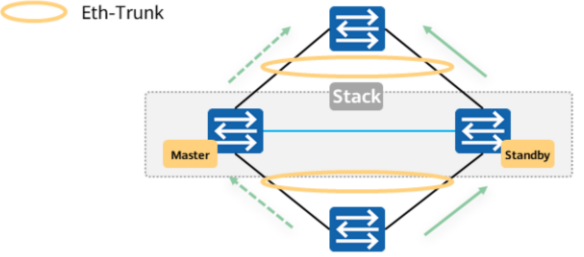
Smooth upgrade process

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Active area: area where the master switch is located.
- A smooth upgrade goes through three phases:
  - The master switch issues the smooth upgrade command to the entire stack. Member switches in the backup area restart with the new system software.
  - Member switches in the backup area set up an independent stack running the new system software and notify member switches in the active area. The master switch in the backup area starts to control the stack, and traffic is transmitted through the backup area. The active area then starts the upgrade.
  - Member switches in the active area restart with the new system software and join the stack set up in the backup area. The master switch in the backup area displays the upgrade result depending on the stack setup result.

Inter-device Link Aggregation (1)

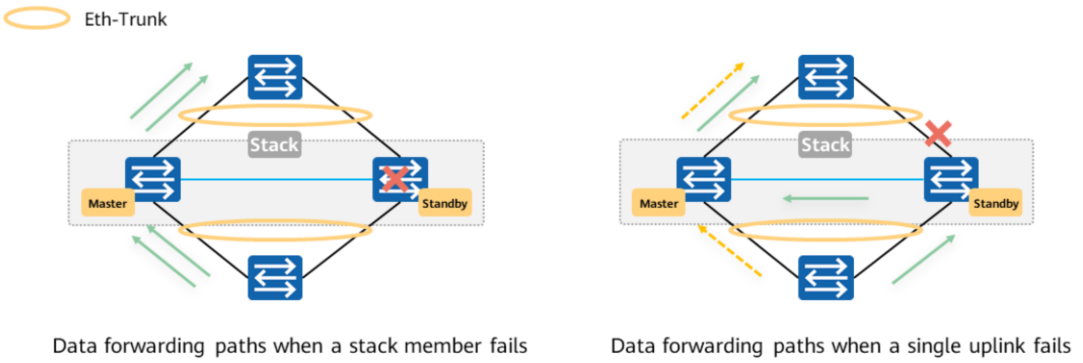
- A stack supports inter-device link aggregation, which allows physical interfaces on different member switches to bundle into an Eth-Trunk interface.
- If a member link of an Eth-Trunk or a member switch fails, traffic on this link or switch can be distributed to other member links through the stack cables connecting member switches. Inter-device link aggregation implements link and device backup and ensures reliable data transmission.



Data forwarding paths when links and stacked devices are operating normally

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Inter-device Link Aggregation (2)



Data forwarding paths when a stack member fails

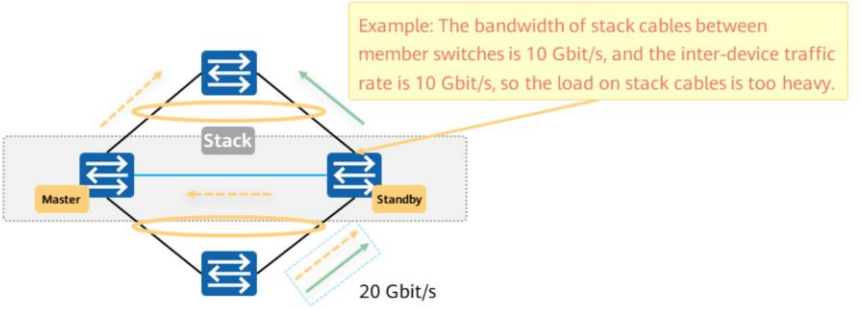
Data forwarding paths when a single uplink fails

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- As shown in the figures, when an uplink or a member switch fails, the inter-device link aggregation technology can load balance traffic to other member interfaces through stack cables connecting member switches, thereby improving network reliability.
- However, heavy inter-device traffic will greatly increase the load on stack cables.

Local Preferential Forwarding (1)

Traffic is distributed to different member links of an Eth-Trunk based on traffic characteristics using the load balancing algorithm of link aggregation. For inter-device link aggregation, the outbound and inbound interfaces of packets may be on different member devices. In this case, traffic is forwarded through stack cables between stack members. This increases the load on stack cables and reduces the forwarding efficiency.



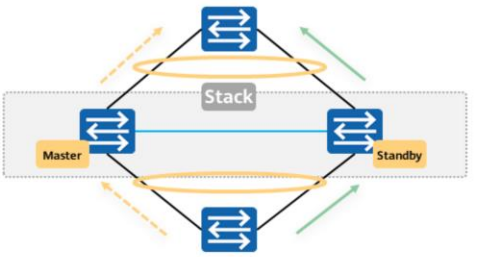
Load balancing among member links of an inter-device link aggregation

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- As shown in the figure, when traffic is load balanced through inter-device link aggregation, some traffic is forwarded through stack cables between member switches. This greatly increases the bandwidth load on stack cables. If the traffic to be forwarded through stack cables exceeds their bandwidth, some packets cannot be transmitted in a timely manner.

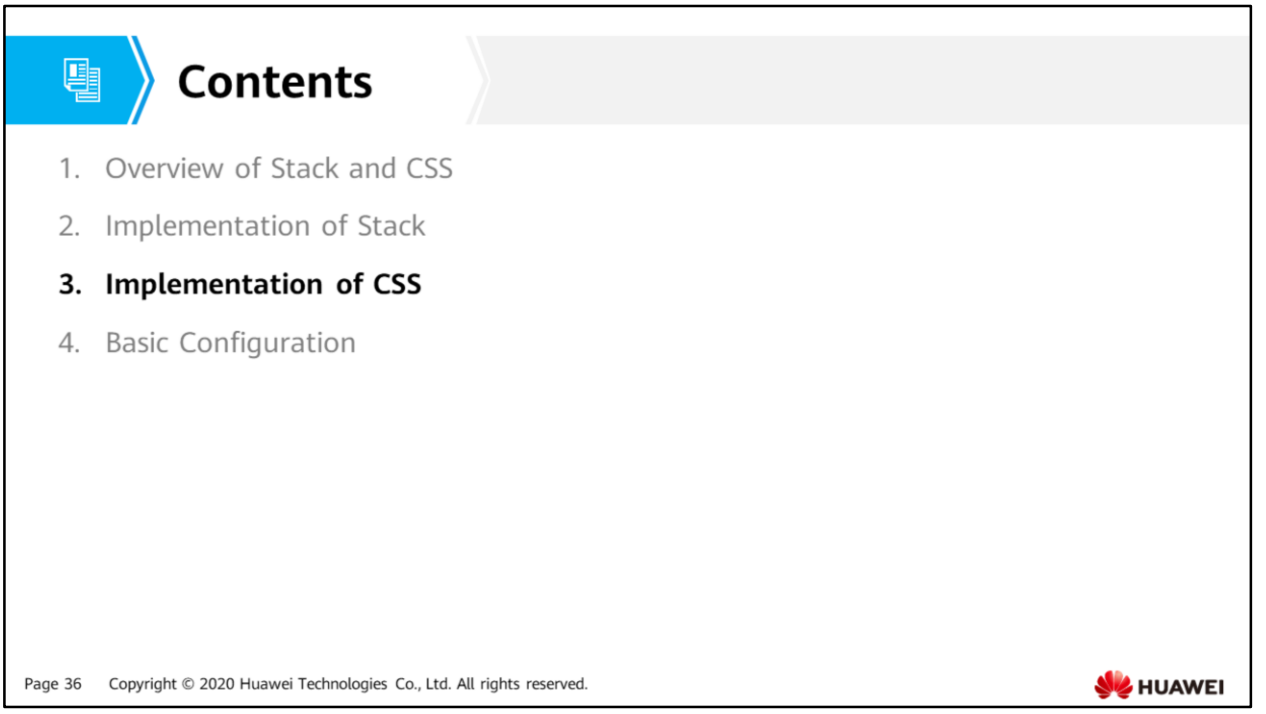
Local Preferential Forwarding (2)

To improve the forwarding efficiency and reduce the load on stack cables, you can enable the local preferential forwarding function on a switch. This function allows traffic reaching the local switch to be preferentially forwarded through a local interface. If the local device has no outbound interface or all outbound interfaces fail, traffic is forwarded through an interface on the other member switch.



Local preferential forwarding

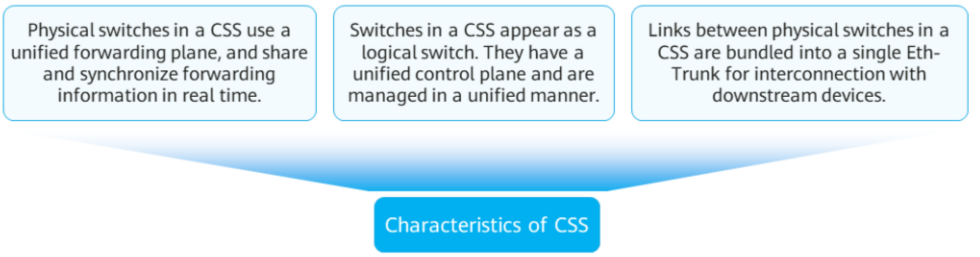
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Implementation

Introduction to CSS

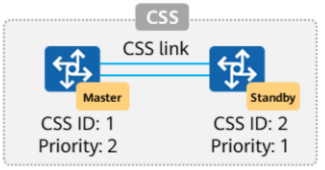
- A cluster switch system (CSS) combines two clustering-capable switches into a single logical switch. A CSS is also called a cluster.
- CSS and stack are different only in their names and implementations, but they are actually the same function. CSS is used on modular switches, whereas stack is used on fixed switches.
- The following mainly describes the peculiarities of CSS compared with stack.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Concepts of CSS

- Switches in a CSS are CSS members. Every CSS member plays one the following roles:
  - Master switch: manages the CSS.
  - Standby switch: acts as a backup to the master switch.
- CSS ID: uniquely identifies a CSS member.
- CSS link: is dedicated to setting up a CSS for data communication between the CSS master and standby switches.
- CSS priority: determines the role of a CSS member in role election. A larger value indicates a higher priority.

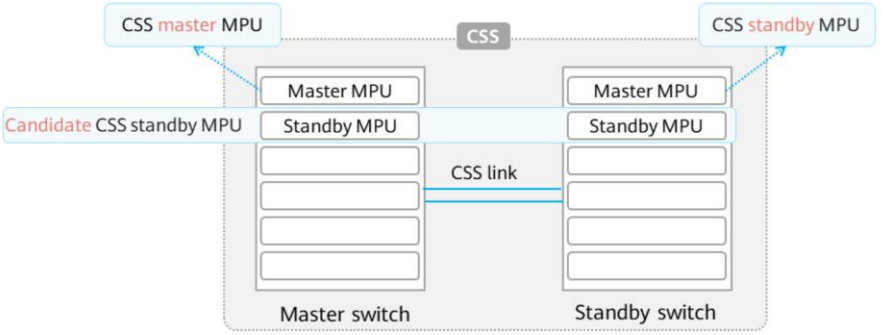


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The standby switch acts as a backup to the master switch. If the master switch fails, the standby switch takes over all services from the master switch and assumes the CSS master role. A CSS has only one standby switch.
- A CSS link can be one link or a bundle of multiple links.
- By default, the CSS ID of a switch is 1. Two switches with the same CSS ID cannot set up a CSS. Before setting up a CSS, you need to manually set the CSS ID of one member switch to 2.
- A switch with a higher CSS priority is more likely to be elected as the master switch.

CSS Control Plane

- A CSS is automatically set up after you connect two switches using CSS cables, enable the CSS function, and complete the CSS configuration on the two switches, and restart them.
- The master MPU on the master switch works as the control plane of the CSS and manages the CSS. The master MPU on the standby switch works as the standby control plane of the CSS. The standby MPUs of the master and standby switches work as candidate CSS standby MPUs.



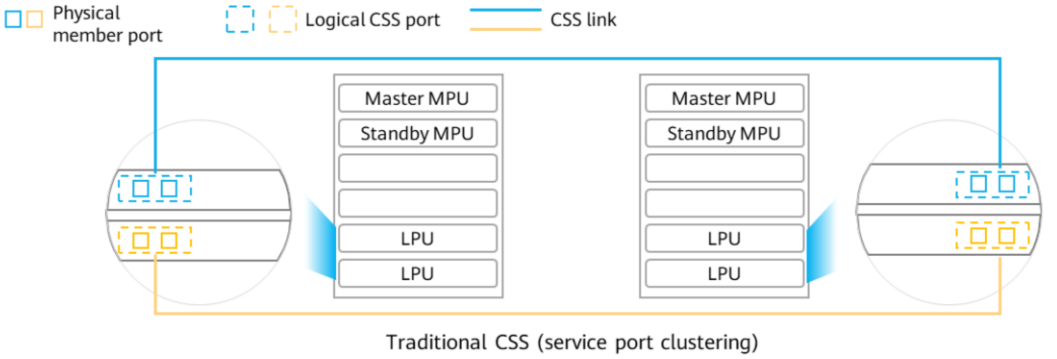
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The master switch in a CSS is elected in the same way as the master switch in a stack, and the other switch is elected as the standby switch.

Physical Connections of a CSS

The physical connections of a CSS vary according to the CSS technology used:

- Traditional CSS: CSS cards on MPUs or service ports on LPUs are connected to set up a CSS.
- Cluster Switch System Generation 2 (CSS2): CSS cards on switch fabric units (SFUs) are connected to set up a CSS.



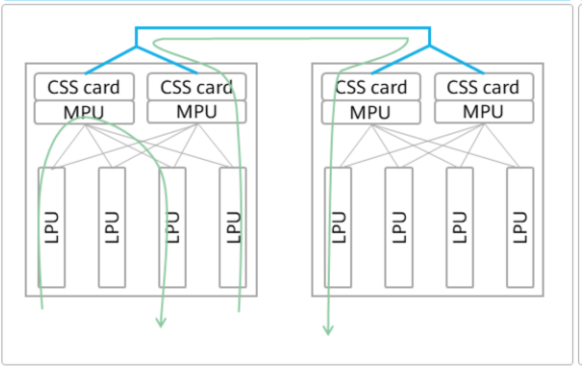
Traditional CSS (service port clustering)

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

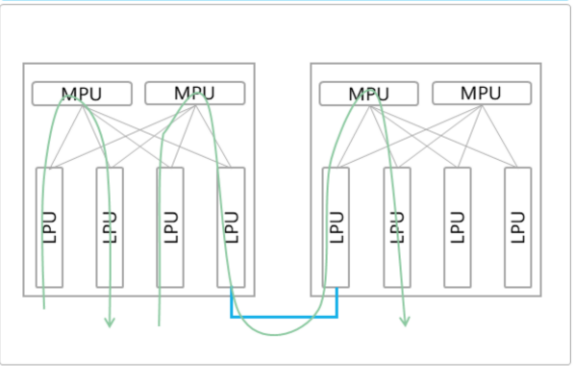
- When service ports are connected to set up a CSS, the number and type of member ports at both ends must be the same, and there is no limitations on the connection sequence.
- Service ports can be connected in either of the following ways according to link distribution:
  - 1+0 networking: Each member switch has one logical CSS port and connects to the other member switch through CSS member ports on one LPU.
  - 1+1 networking: Each member switch has two logical CSS ports and connects to the other member switch through CSS member ports on two LPUs. CSS links on the two LPUs implement link redundancy, as shown in the figure.
- CSS2: CSS cards on SFUs are connected to set up a CSS. In addition to functions supported by traditional CSS, CSS2 supports 1+N backup of MPUs in a CSS.

Traditional CSS

Setting up a CSS through CSS cards on MPUs



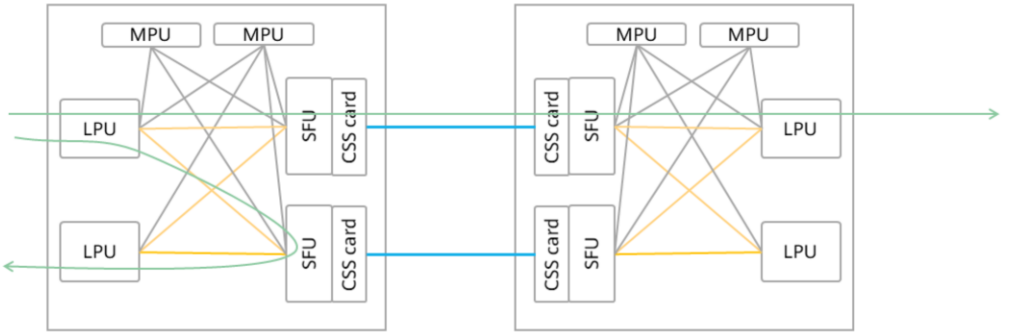
Setting up a CSS through service ports



For modular switches that support only traditional CSS, inter-LPU traffic in a chassis and inter-chassis traffic must pass through MPUs. Therefore, if there is no MPU working normally on a chassis, traffic cannot be forwarded between LPUs on this chassis or from this chassis to another chassis.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

CSS2



- CSS2-capable modular switches adopt the forwarding-control separation architecture, so inter-LPU traffic in a chassis or inter-chassis traffic no longer needs to pass through MPUs. If a chassis in a CSS does not have any MPU that works properly, traffic forwarding on the chassis is not affected.
- CSS2 allows LPUs on two switches in a CSS to forward packets normally as long as one MPU in either switch is running properly. This feature is known as 1+N backup of MPUs in a CSS.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Overview of Stack and CSS
2. Implementation of Stack
3. Implementation of CSS
4. **Basic Configuration**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Stack Configuration (1)

1. Create a logical stack port and add stack member ports to it.

   [Huawei] **interface stack-port** *member-id/port-id*
   [Huawei-stack-port0/1] **port interface** { *interface-type interface-number1* [**to** *interface-type interface-number2*] } &<1-10> **enable**

   member-id specifies the stack ID of a member switch. *port-id* specifies the ID of the local logical stack port, which can be set to 1 or 2.

2. Configure a stack ID for a member switch.

   [Huawei] **stack slot** *slot-id* **renumber** *new-slot-id*

   By default, the stack ID of a member switch is 0. The new stack ID takes effect only after the current configuration is saved and the switch is restarted.

3. Configure a stack priority for a member switch.

   [Huawei] **stack slot** *slot-id* **priority** *priority*

   By default, the stack priority of a member switch is 100.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Stack Configuration (2)

4. Enable MAD in direct mode in the interface view.

```
[Huawei-GigabitEthernet0/0/1] mad detect mode direct
```

5. Configure MAD in relay mode when an independent relay agent is used.

   In the stack, enable MAD on the Eth-Trunk interface connected to the relay agent.

```
[Huawei] interface eth-trunk trunk-id
[Huawei-Eth-Trunk1] mad detect mode relay
```

   On the relay agent, enable MAD on the Eth-Trunk interface connected to the stack.

```
[Huawei] interface eth-trunk trunk-id
[Huawei-Eth-Trunk1] mad relay
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Stack Configuration (3)

6. Configure MAD in relay mode when two stacks function as each other's MAD relay agent.

   Set a unique MAD domain ID for each stack.

   ```
   [Huawei] mad domain domain-id
   ```

   By default, the MAD domain ID of a stack is 0.

   Enable MAD in relay mode, and configure two stacks as each other's relay agent.

   ```
   [Huawei] interface eth-trunk trunk-id
   [Huawei-Eth-Trunk1] mad detect mode relay
   [Huawei-Eth-Trunk1] mad relay
   ```

   MAD needs to be enabled on the Eth-Trunk interface connected to the MAD relay agent.

7. Perform a master/standby switchover.

   ```
   [Huawei] slave switchover
   ```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Stack Configuration (4)

8. Set a period after which a stack changes its system MAC address.

   [Huawei] **stack timer mac-address switch-delay** *delay-time*

   By default, a stack changes the system MAC address after 10 minutes. When this period is set to 0, no stack MAC address switching will be performed. To configure a stack to change its system MAC address immediately, run the **undo stack timer mac-address switch-delay** command.

9. Clear all stack configuration.

   [Huawei] **reset stack configuration**

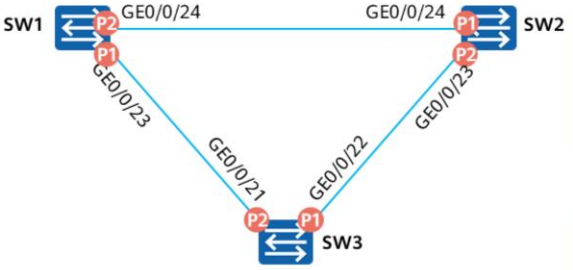
   The following stack configuration will be cleared: switch slot ID, stack priority, stack reserved VLAN, stack MAC address switching delay, logical stack port configuration, and logical stack port rate. For example, when switches are stacked using dedicated stack cables, to ensure that slot IDs are automatically generated for the switches based on the sequence in which dedicated stack cables are connected, you can run this command to clear all stack configuration.
   Running this command will cause the stack to split and member switches to restart.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Example for Configuring a Stack (1)

P1 Stack port 1

P2 Stack port 2



- SW1, SW2, and SW3 are stacked in a ring topology.
- SW1, SW2, and SW3 are the master, standby, and slave switches respectively. Their stack IDs are 0, 1, and 2, and their stack priorities are 200, 100, and 100, respectively.

Configure logical stack ports and add stack member ports to them.

SW1 configuration:

```
[SW1] interface stack-port 0/1
[SW1-stack-port0/1] port interface gigabitethernet 0/0/23 enable
[SW1] interface stack-port 0/2
[SW1-stack-port0/2] port interface gigabitethernet 0/0/24 enable
```

SW2 configuration:

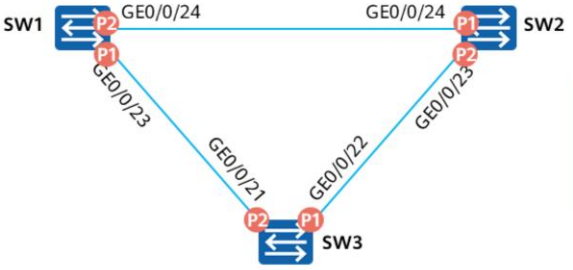
```
[SW2] interface stack-port 0/1
[SW2-stack-port0/1] port interface gigabitethernet 0/0/24 enable
[SW2] interface stack-port 0/2
[SW2-stack-port0/2] port interface gigabitethernet 0/0/23 enable
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Example for Configuring a Stack (2)

P1 Stack port 1

P2 Stack port 2



SW3 configuration:

```
[SW3] interface stack-port 0/1
[SW3-stack-port0/1] port interface gigabitethernet 0/0/22 enable
[SW3] interface stack-port 0/2
[SW3-stack-port0/2] port interface gigabitethernet 0/0/21 enable
```

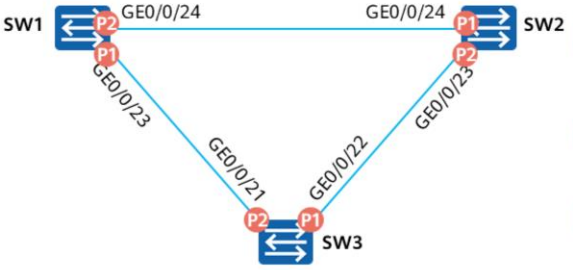
- SW1, SW2, and SW3 are stacked in a ring topology.
- SW1, SW2, and SW3 are the master, standby, and slave switches respectively. Their stack IDs are 0, 1, and 2, and their stack priorities are 200, 100, and 100, respectively.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Example for Configuring a Stack (3)

P1 Stack port 1

P2 Stack port 2



- SW1, SW2, and SW3 are stacked in a ring topology.
- SW1, SW2, and SW3 are the master, standby, and slave switches respectively. Their stack IDs are 0, 1, and 2, and their stack priorities are 200, 100, and 100, respectively.

Configure the stack ID and stack priority.

SW1 configuration:

```
[SW1] stack slot 0 priority 2
```

SW2 configuration:

```
[SW2] stack slot 0 renumber 1
```

SW3 configuration:

```
[SW3] stack slot 0 renumber 2
```

Save the configuration, power off SW1, SW2, and SW3, connect them using SFP+ cables, and then power them on.

Connect stack-port 0/1 on one switch to stack-port 0/2 on another switch. Otherwise, the stack will fail to be set up.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

CSS Configuration (1)

1. Configure a CSS ID for a member switch.

   [Huawei] **set css id** *new-id*

   By default, the CSS ID of a member switch is 1. Two switches with the same CSS ID cannot set up a CSS. Therefore, you need to manually set the CSS ID of a member switch to 2 before setting up a CSS.

2. Create a logical CSS port and add CSS member ports to it.

   [Huawei] **interface css-port** *port-id*
   [Huawei-css-port1] **port interface** { *interface-type interface-number1* [**to** *interface-type interface-number2*] }
   &<1-10> **enable**

   A service port is configured as a CSS member port and added to a logical CSS port.

3. Configure a CSS priority for a member switch.

   [Huawei] **set css priority** *priority*

   By default, the CSS priority of a member switch is 1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

CSS Configuration (2)

4. Enable the CSS function.

   [Huawei] **css enable**

   By default, the CSS function is disabled on a switch. When enabling this function, the system asks you whether to restart the switch immediately to make the configuration take effect.
   The CSS function must be enabled on both member switches.

5. Set the CSS connection mode.

   [Huawei] **set css mode { lpu | css-card }**

   The default CSS connection mode varies according to the device model.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The method and command for configuring the MAD function in a CSS are the same as those in a stack.

Example for Configuring a CSS (1)



- SW1 and SW2 have LPUs installed and are connected through CSS cables. To increase the bandwidth and improve reliability, the two switches are connected through four service ports on two LPUs of each switch.
- Set the CSS connection mode on SW1 and SW2, and set their CSS IDs to 1 and 2, respectively. In addition, set their CSS priorities to 100 and 10, respectively, so that SW1 is more likely to be elected as the master switch.

Configure the CSS connection mode, CSS ID, and CSS priority for each member switch.

SW1 configuration:

```
[SW1] set css mode lpu
[SW1] set css id 1
[SW1] set css priority 100
```

SW2 configuration:

```
[SW2] set css mode lpu
[SW2] set css id 2
[SW2] set css priority 10
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Example for Configuring a CSS (2)



- SW1 and SW2 have LPUs installed and are connected through CSS cables. To increase the bandwidth and improve reliability, the two switches are connected through four service ports on two LPUs of each switch.
- Set the CSS connection mode on SW1 and SW2, and set their CSS IDs to 1 and 2, respectively. In addition, set their CSS priorities to 100 and 10, respectively, so that SW1 is more likely to be elected as the master switch.

Configure logical CSS ports.

SW1 configuration:

```
[SW1] interface css-port 1
[SW1-css-port1] port interface xgigabitethernet 1/0/1 to xgigabitethernet 1/0/2 enable
[SW1-css-port1] quit
[SW1] interface css-port 2
[SW1-css-port2] port interface xgigabitethernet 2/0/1 to xgigabitethernet 2/0/2 enable
[SW1-css-port2] quit
```

Configure service ports XGE1/0/1 and XGE1/0/2 on SW1 as CSS member ports and add them to CSS port 1; configure service ports XGE2/0/1 and XGE2/0/2 on SW2 as CSS member ports and add them to CSS port 2.

The SW2 configuration is similar to the SW1 configuration, and is not mentioned here.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Example for Configuring a CSS (3)

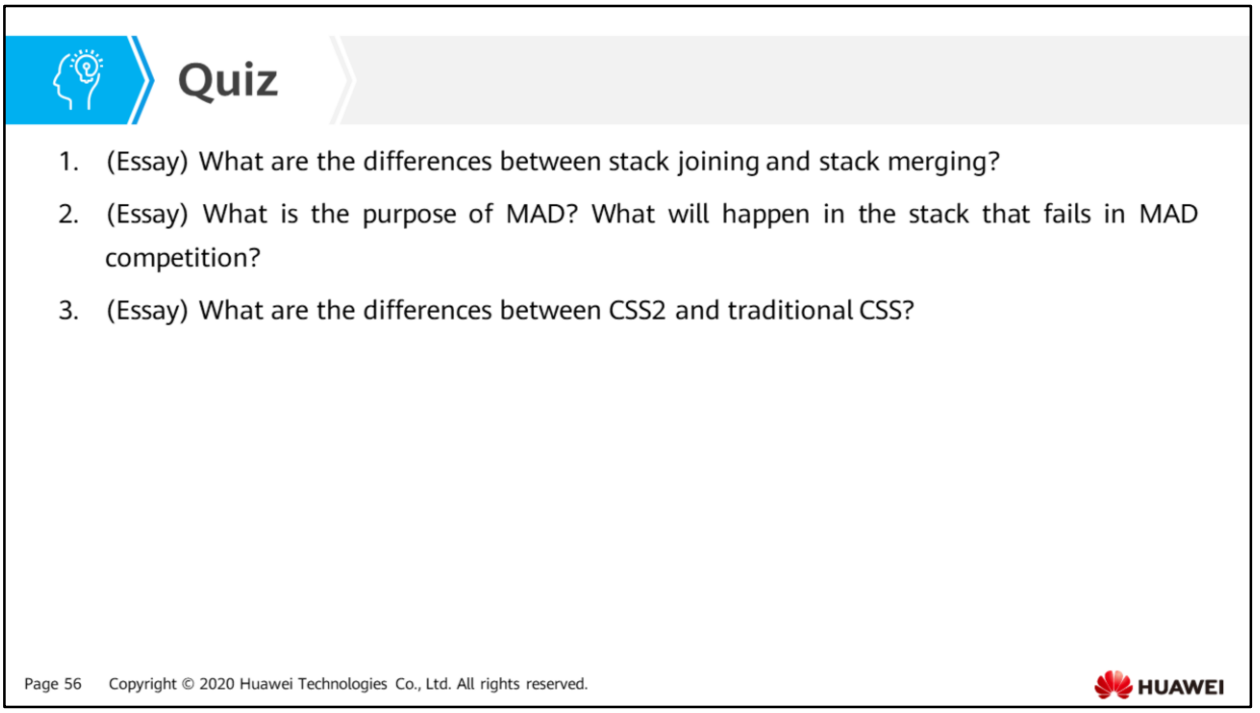


- SW1 and SW2 have LPUs installed and are connected through CSS cables. To increase the bandwidth and improve reliability, the two switches are connected through four service ports on two LPUs of each switch.
- Set the CSS connection mode on SW1 and SW2, and set their CSS IDs to 1 and 2, respectively. In addition, set their CSS priorities to 100 and 10, respectively, so that SW1 is more likely to be elected as the master switch.

Enable the CSS function.

```
[SW1] css enable
Warning: The CSS configuration will take effect only after the system is rebooted.  The next CSS mode is LPU.  Reboot now?  [Y/N]:y
[SW2] css enable
Warning: The CSS configuration will take effect only after the system is rebooted.  The next CSS mode is LPU.  Reboot now?  [Y/N]:y
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Reference answers:
  - In the stack joining scenario, a switch has been connected to a running stack through stack cables before being powered on. After the switch is powered on and starts, it becomes a slave switch since the stack already has a master switch. In the stack merging scenario, two stacks are connected through stack cables, and a new master switch is elected for the new stack and updates topology information.
  - After a stack-enabled switch is powered on, it becomes a single-switch stack, with itself being the master switch. In this case, if this switch is connected to another stack through stack cables, the two stacks merge. This is a typical difference between stack merging and stack joining.
  - If a stack or CSS splits, more than one stack or CSS may use the same IP address and MAC address, which will cause entry conflicts on other network devices. MAD prevents this situation to ensure normal data forwarding. The stack that fails in the MAD competition shuts down all ports except the reserved ones on its member switches. This prevents IP and MAC address conflicts between stacks, thereby preventing entry conflicts on other network.
  - CSS2 supports 1+N backup of MPUs. That is, as long as one MPU on any member switch in a CSS is working and the control plane of the cluster is working normally, the data plane of the cluster can forward packets normally.

Summary

- The stack and CSS techniques greatly simplify network deployment and management, implement fast deployment of highly reliable and loop-free campus networks, and greatly reduce the impact of network faults.
- The MAD function can be configured to prevent forwarding exceptions caused by IP and MAC address conflicts between stacks. MAD can be implemented in direct or relay mode. Member switches in the stack that fails the MAD competition will shut down all physical ports except the reserved ones.
- To ensure packet forwarding efficiency, you are advised to configure local preferential forwarding when deploying Eth-Trunks in a stack or CSS.
- CSS2 technology uses CSS cards on SFUs to set up a CSS, and supports the 1+N backup of MPUs by leveraging the forwarding-control separation architecture of switches.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



IP Multicast Basics

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- There are various services on the network. Generally, services can be classified into the following types based on the traffic model:
  - Point-to-point (P2P) services, such as File Transfer Protocol (FTP) and web services. The key characteristic of P2P services is that different users have different requirements for them. For example, user A needs to download document A, whereas user B needs to download document B. This type of service is usually carried in unicast mode, and the server sends different P2P data flows to different users.
  - Point-to-multipoint (P2MP) services, such as IPTV and video conferencing services. The key characteristic of P2MP services is that users have the same requirements for them. For example, users A, B, C, and D all need to watch video X. P2MP services can be carried in unicast, multicast, or broadcast mode. However, using unicast or broadcast to carry P2MP services encounters problems.
  - Multicast can prevent these problems.
- This course describes the advantages of using multicast to carry P2MP services, basic concepts of multicast, and the basic multicast forwarding process.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe the definition and address structure of multicast.
  - Describe multicast data forwarding implementation.
  - Describe reverse path forwarding (RPF) implementation.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

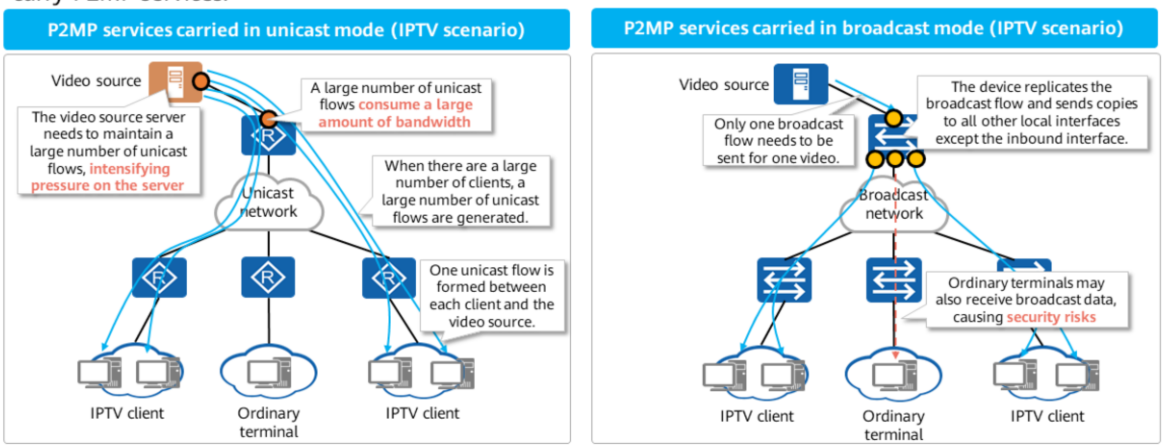
1. **Basic Concepts of Multicast**
2. Multicast Data Forwarding Implementation

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Problems of Using Unicast or Broadcast to Carry P2MP Services

P2MP services can be carried in unicast, multicast, or broadcast mode. There are various implementation modes on the live network. However, some inherent problems occur if unicast or broadcast is used to carry P2MP services.

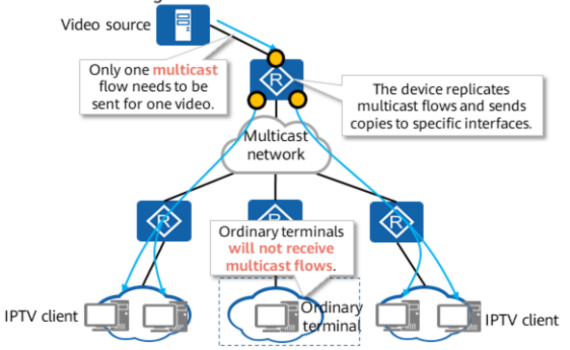


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Unicast transmission is implemented between a source IP host and a destination IP host. Most of data is transmitted in unicast mode on a network. For example, email and online banking applications are implemented in unicast mode.
    - In unicast communication, each data packet has a specific destination IP address. For the same data, if there are multiple receivers, the server needs to send the same number of unicast data packets. If a large number of receivers exist, replication of the same data and transmission of a large number of duplicate copies intensify the pressure on the server, affect device performance, and consume a lot of link bandwidth resources. Therefore, the unicast mode is applicable to networks with only a small number of users. When there are a large number of users, the unicast mode cannot ensure the network transmission quality.
- Broadcast transmission is implemented between a source IP host and all the other IP hosts on the local network. All hosts can receive data from the source host, regardless of whether they require the data.
    - Broadcast data packets are transmitted in a broadcast domain. Once a device sends a broadcast data packet, all other devices in the broadcast domain receive the packet and have to process it, which consumes resources. A large number of broadcast data packets will consume tremendous network bandwidth and device resources. The broadcast mode applies only to shared network segments, and cannot ensure information security and paid services.

Using Multicast to Carry P2MP Services

- In multicast transmission mode, a data flow is transmitted to a group of users along a multicast distribution tree (MDT). Each link transmits only one copy of multicast data flows. Compared with unicast and broadcast, multicast has the following advantages:
  - Compared with unicast, multicast prevents the increase in the number of users from increasing the pressure on the video source or significantly increasing the consumption of network resources.
  - Compared with broadcast, multicast improves transmission security without wasting network resources. In addition, multicast transmission can be implemented across network segments.



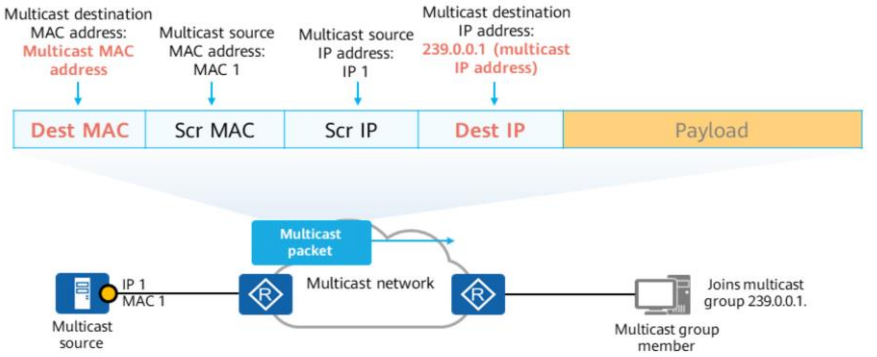
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Multicast transmission is implemented between one source IP host and a group of IP hosts, with transit nodes selectively replicating and forwarding data based on demands of receivers.
- Multicast technologies efficiently implement P2MP service data transmission over an IP network, while conserving network bandwidth and reducing network loads.
- Multicast distribution tree (MDT): a forwarding path of multicast traffic.

Multicast Data Packet Structure

The structure of a multicast data packet is similar to that of a unicast packet, but the destination MAC address and destination IP address of a multicast data packet are different from those of a unicast packet.

- The destination IP address of a multicast packet is a multicast IP address ranging from 224.0.0.0 to 239.255.255.255.
- The destination MAC address of a multicast packet is a multicast MAC address, which is mapped from a multicast IP address.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Multicast IP Addresses

- In the IPv4 address space, class D addresses (224.0.0.0/4) are used for multicast. A multicast address represents a P2MP data flow, for example, IPTV data flow or voice conference data flow.
- In most cases, different services (such as IPTV and voice conference services) on the same multicast network need to use different multicast IP addresses.
- IANA further defines class D addresses. The following table lists main multicast addresses.

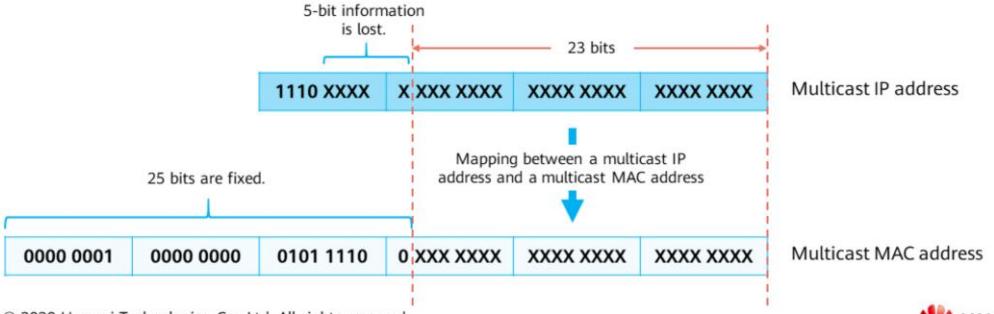
Address Range	Description
224.0.0.0–224.0.0.255	Permanent group addresses reserved for routing protocols
224.0.1.0–231.255.255.255 233.0.0.0–238.255.255.255	Temporary any-source multicast group addresses
232.0.0.0–232.255.255.255	Temporary source-specific multicast group addresses valid on the entire network
239.0.0.0–239.255.255.255	Temporary any-source multicast group addresses valid only in the local administration domain

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- IPv4 multicast addresses:
  - The IPv4 address space is divided into five classes, class A to class E. Class D addresses are IPv4 multicast addresses, ranging from 224.0.0.0 to 239.255.255.255. These addresses identify multicast groups and can only be used as destination addresses of multicast packets, not as source addresses.
  - Source addresses of IPv4 multicast packets are IPv4 unicast addresses, which can be class A, class B, or class C addresses and cannot be class D or class E addresses.
  - All receivers of a multicast group are identified by the same IPv4 multicast group address at the network layer. Once a user joins the multicast group, the user can receive IP multicast packets with the group address as the destination address.

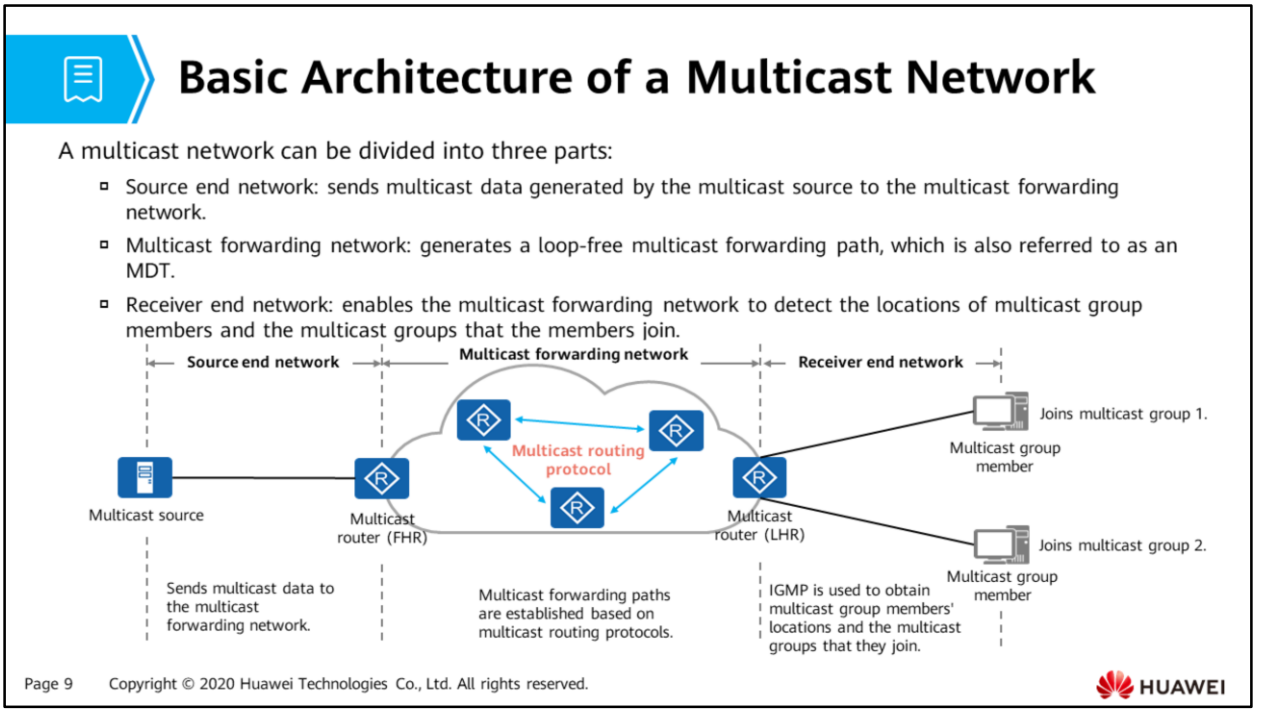
Multicast MAC Addresses

- When unicast IPv4 packets are transmitted on an Ethernet network, MAC addresses of receivers are used as destination MAC addresses. When a multicast packet is transmitted, its destination, however, is a group of unspecific members. Therefore, an IPv4 multicast MAC address is needed.
- As defined by IANA, the most significant 24 bits of an IPv4 multicast MAC address are 0x01005e, the 25th bit is 0, and the least significant 23 bits are the least significant 23 bits of an IPv4 multicast address. For example, multicast group IP address 224.0.1.1 corresponds to multicast MAC address 01-00-5e-00-01-01.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The most significant 4 bits of an IPv4 multicast address are fixed as 1110, mapping the leftmost 25 bits of a multicast MAC address. Among the last 28 bits in the IPv4 address, only 23 bits are mapped to the rest bits in the MAC address, with 5 bits lost. For example, multicast IP addresses 224.0.1.1, 224.128.1.1, 225.0.1.1, and 239.128.1.1 are all mapped to multicast MAC address 01-00-5e-00-01-01. This must be taken into consideration during address assignment.
- IETF believes that this will not cause great impact because there is a very low probability that two or more group addresses in the same LAN will be mapped to the same MAC address.
- A multicast MAC address identifies a group of devices. The least significant bit of the first byte in a multicast MAC address is 1, for example, 0100-5e-00ab.
- The devices identified by the same multicast MAC address are in the same multicast group. These devices listen to the data frames whose destination MAC address is this multicast MAC address. A unicast MAC address can be assigned to an Ethernet interface, whereas a multicast or broadcast MAC address cannot be assigned to any Ethernet interface. In other words, a multicast or broadcast MAC address cannot be used as the source MAC address of a data frame, but can be used as the destination MAC address of a data frame.
- For example, the BPDU payload of the STP protocol is directly encapsulated in the Ethernet data frame, with the destination MAC address being 0180-c200-0000, which is a multicast MAC address. There are many similar examples, which are not listed here. These multicast MAC addresses are not associated with multicast IP addresses.
- In addition, we need to pay special attention to the multicast MAC addresses that map multicast IP addresses. The multicast MAC addresses described in this course are of such a type.



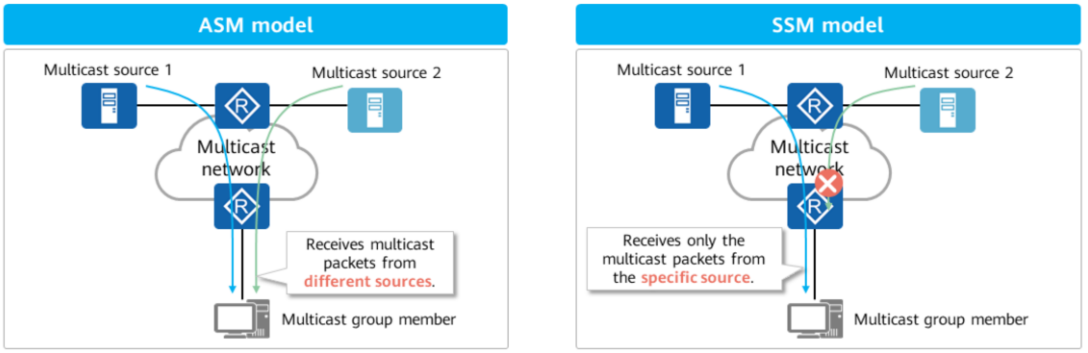
- Multicast source: a sender of multicast traffic, such as a multimedia server. A multicast source does not need to run any multicast protocol. It only needs to send multicast data.
- Multicast receiver: also called a multicast group member, is a device that expects to receive traffic of a specific multicast group, for example, a PC running multimedia live broadcast client software.
- Multicast group: a group of receivers identified by a multicast address. User hosts (or other receiver devices) that have joined a multicast group become members of the group and can identify and receive the IP packets destined for the multicast group address.
- Multicast router: a network device that supports multicast and runs multicast protocols. In addition to routers, switches and firewalls support multicast (depending on device models). Routers are used in this example.
- First-hop router (FHR): a router that directly connects to the multicast source on the multicast forwarding path and is responsible for forwarding multicast data from the multicast source.
- Last-hop router (LHR): a router that directly connects to multicast group members (receivers) on the multicast forwarding path and is responsible for forwarding multicast data to these members.
- The Internet Group Management Protocol (IGMP) is a protocol in the TCP/IP protocol suite and manages group memberships between receiver hosts and immediately

neighboring multicast routers.

Multicast Service Models

When receiving multicast data, multicast group members can select multicast data sources. Therefore, two multicast service models are available: any-source multicast (ASM) and source-specific multicast (SSM).

- ASM: After a member joins a multicast group, the member can receive data sent by any source to the group.
- SSM: After a member joins a multicast group, the member can only receive data sent by the specific source to the group.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- ASM characteristics:
  - In ASM, to improve security, multicast source filter policies can be configured on routers to permit or deny packets from some multicast sources. This filters data sent to receiver hosts.
  - In the ASM model, each group address must be unique on the entire multicast network. That is, an ASM group address can only be used by only one multicast application at a time. If two applications use the same ASM group address to send data, their receiver hosts receive data from two sources, which may cause network traffic congestion and affect the receiver hosts.
- SSM characteristics:
  - The SSM model does not require globally unique group addresses, but the multicast source must be unique to multicast groups. That is, different applications on a source must use different SSM group addresses. Different applications on different sources can share one SSM group addresses because each source-group pair has an (S, G) entry. This model saves multicast group addresses without congesting the network.

Contents

1. Basic Concepts of Multicast
2. **Multicast Data Forwarding Implementation**

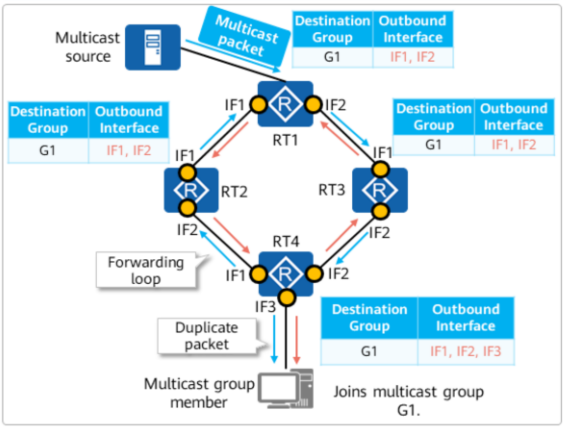
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

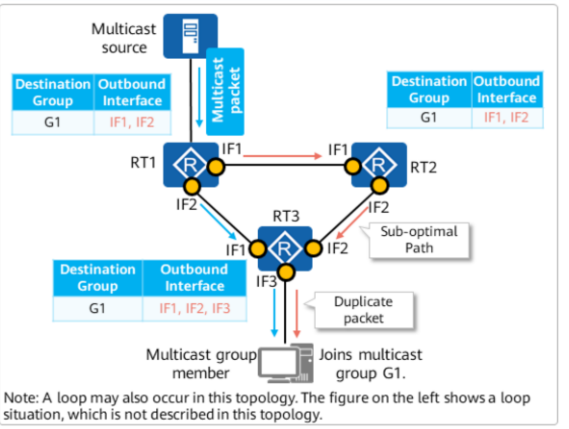
Problems That Accompany Multicast Data Forwarding

Multicast data forwarding depends on routing entries. However, such forwarding encounters the following problems:

Forwarding loops occur, and duplicate packets are received.



Sub-optimal paths exist, and duplicate packets are received.



Note: A loop may also occur in this topology. The figure on the left shows a loop situation, which is not described in this topology.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Multicast Routing and RPF Check

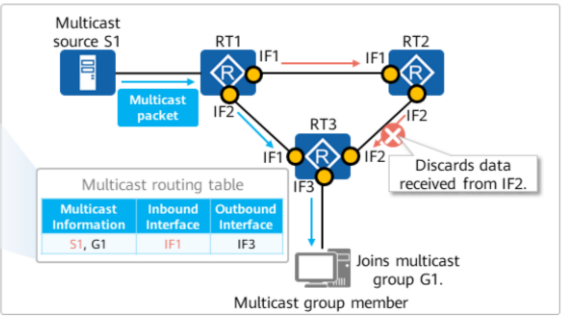
- Loops, sub-optimal routes, and duplicate packets may occur during multicast forwarding. To prevent these problems, in addition to the destination network and outbound interface, multicast source and inbound interface information needs to be added to multicast routing entries. Then a device forwards only the multicast data received from the specified inbound interface, preventing problems such as loops, sub-optimal routes, and duplicate packets (partially solved) during multicast forwarding.
- For the same multicast source, the device can determine the unique inbound interface of multicast traffic through the reverse path forwarding (RPF) check.
- RPF check that is based on multicast routing entries:

Decided by the RPF check

(192.168.0.2, 239.0.0.1)
Upstream Interface: GigabitEthernet1/0/0 //Unique inbound interface
Downstream interfaces
1: GigabitEthernet2/0/0
2: GigabitEthernet3/0/0

Multicast routing entries

RPF check prevents loops, sub-optimal paths, and duplicate packets.

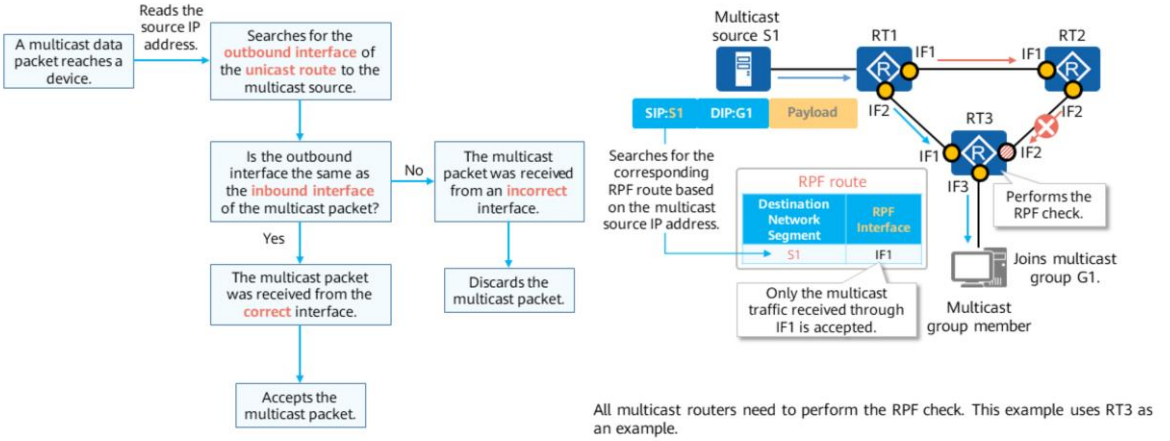


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The outbound interface of a multicast routing entry is usually determined by the multicast routing protocol.
- Multicast routing protocols will be covered in the course of **PIM Implementation and Configuration**.
- A multicast routing entry contains a multicast source and a multicast group. Therefore, it is also called an (S, G) entry.

Implementation of RPF Check

Process of an RPF check:

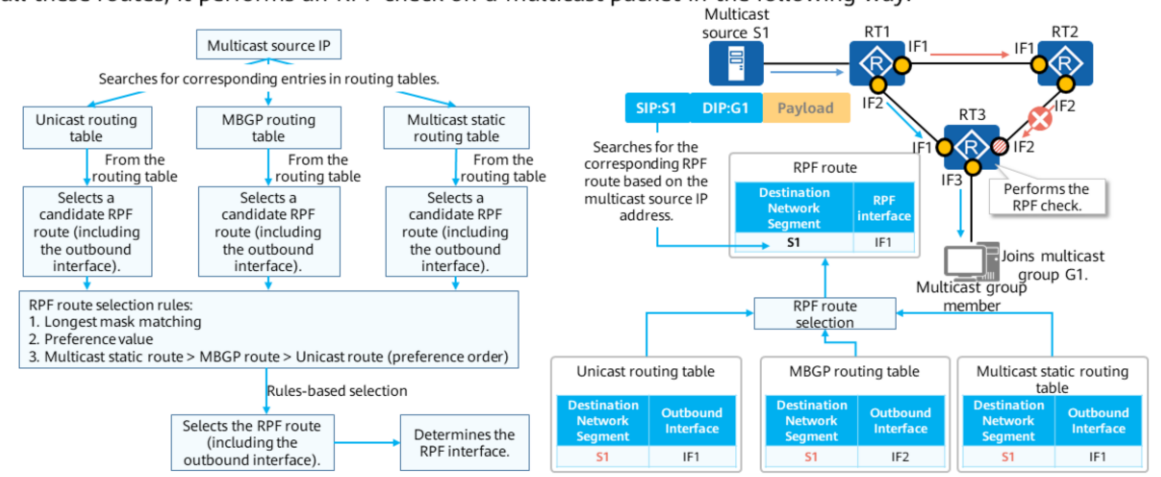


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Each multicast router searches its routing tables (unicast routing table and MBGP routing table or multicast static routing table) for the route to the packet source based on the source address of a received packet. Then, the multicast router checks whether the outbound interface of the route to the packet source is the same as the inbound interface of the received multicast packet. If they are the same, the router considers that the multicast packet was received through the correct interface and accepts it. This ensures the correct forwarding path and allows the router to accept the multicast packet only through one inbound interface. This process is called the RPF check.

RPF Route Selection Rules

RPF routes can be selected among unicast routes, MBGP routes, and multicast static routes. If a router has all these routes, it performs an RPF check on a multicast packet in the following way:

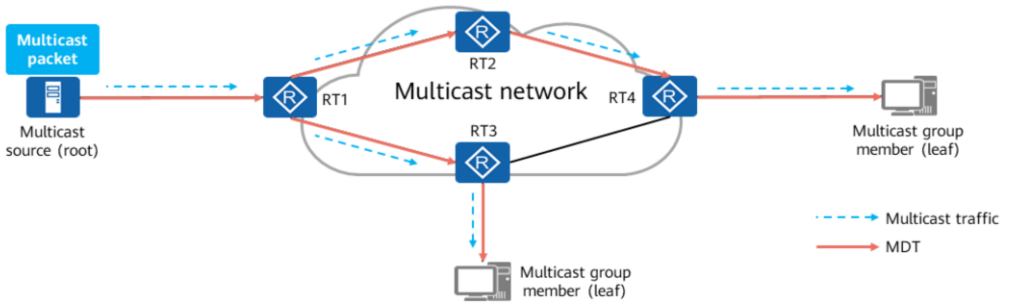


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The router selects one of the three routes as the RPF route according to the following rules:
  - If route selection based on the longest match rule is configured, the router selects the route with the longest matching mask from the three routes.
  - If the masks of the three routes have the same length, the route with the highest preference is selected.
  - If the preferences of the three routes are also the same, the multicast static route, MBGP route, and unicast route are preferred in descending order.
- MBGP:
  - MBGP is used to transmit multicast source-related routing entries.
- Multicast static routing table:
  - It contains routes for which the mapping between the multicast source and the outbound interface is manually configured.

Multicast Distribution Tree

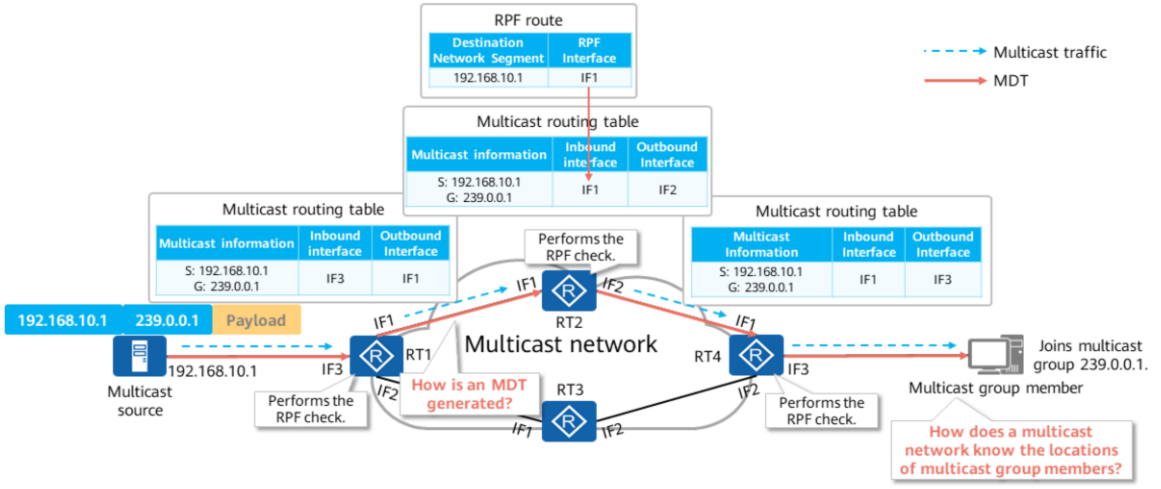
- It is required that no loops, sub-optimal paths, or duplicate packets occur during multicast data forwarding.
- Through the RPF mechanism and multicast routing protocols, a loop-free multicast forwarding path (MDT) without sub-optimal paths or duplicate packets can be established on the multicast network.
- Each MDT takes the multicast source as the root and group members as leaves. Multicast data is forwarded based on the MDT.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Multicast Data Forwarding Process

The multicast data forwarding process is as follows:



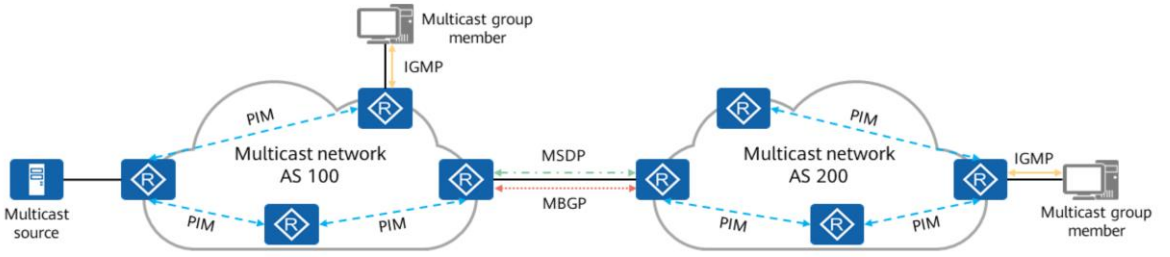
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The outbound interface of a multicast routing entry and multicast forwarding path are determined by a multicast routing protocol.
  - Multicast routing protocols include PIM, MBGP, and Multicast Source Discovery Protocol (MSDP).
  - For details about multicast routing protocols, see the course of **PIM Implementation and Configuration**.
- The locations of multicast group members are advertised through IGMP.
  - For details about IGMP, see the course of **PIM Implementation and Configuration**.

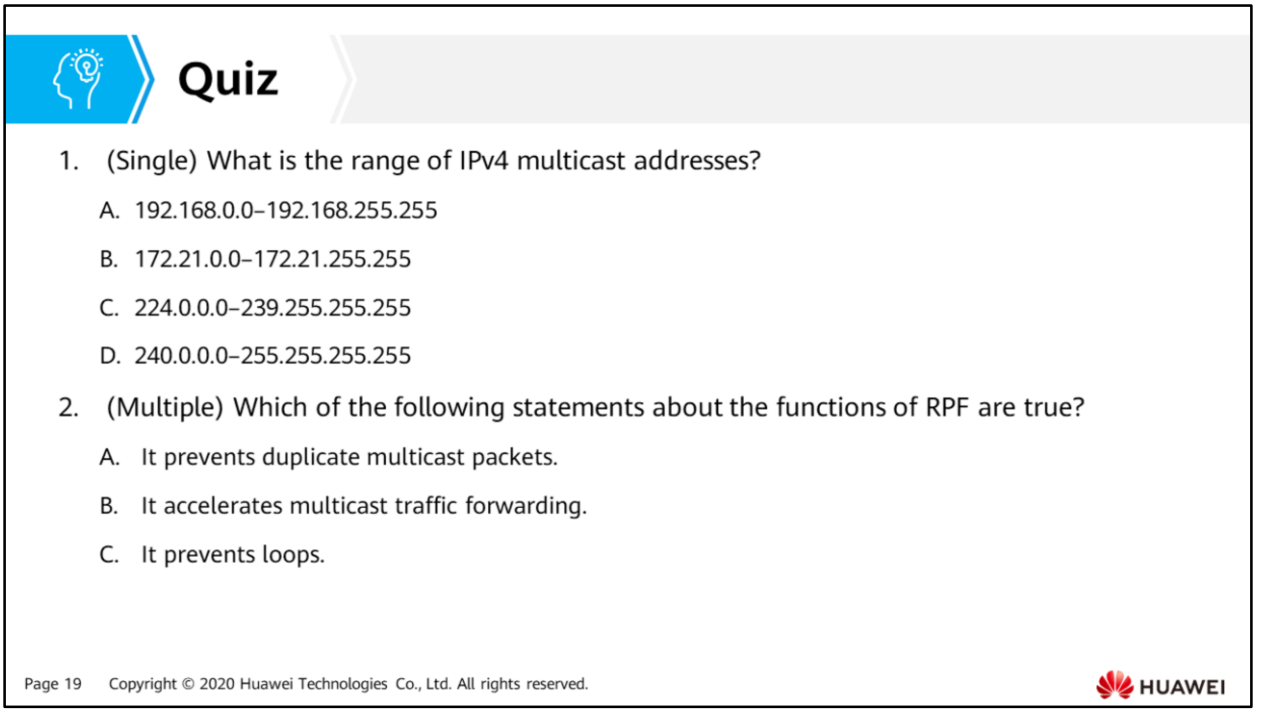
Multicast Protocol Introduction

A multicast network needs to establish forwarding paths based on multiple multicast protocols.

- IGMP runs on the receiver end network and is used to **inform the multicast network of the locations of group members and the multicast groups that the members join**.
- Protocols working on the multicast forwarding network include PIM, MSDP, and MBGP.
  - PIM is mainly used to **generate MDTs in an AS**.
  - MSDP is mainly used to **help generate inter-AS MDTs**.
  - MBGP is used to **help perform RPF check on inter-AS multicast traffic**.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI



1. C
   - The Internet Assigned Numbers Authority (IANA) allocates class D addresses for IPv4 multicast. An IPv4 address is 32 bits long, and the most significant 4 bits of a Class D IP address are 1110. Therefore, multicast IP addresses range from 224.0.0.0 to 239.255.255.255.
2. AC

Summary

- Multicast is mainly used to solve the following problems when P2MP traffic is carried in unicast or broadcast mode:
  - When P2MP traffic is carried in unicast mode, with the increase of P2MP service terminals, the bandwidth consumption or the pressure on the source server may be too high.
  - When P2MP traffic is carried in broadcast mode, although the problems accompanying P2MP traffic transmission in unicast mode do not occur, the security is low.
- A multicast network consists of three parts: source end network, multicast forwarding network, and receiver end network.
  - The multicast forwarding network is responsible for forwarding multicast data between multicast routers, but problems such as loops, sub-optimal paths, and duplicate packets may occur. These problems can be partially or completely solved through the RPF check.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



IGMP Implementation and Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- In multicast communication, a multicast network needs to send multicast data to specific multicast group members. Therefore, the multicast network needs to know the locations of the group members and the multicast groups that the group members join.
- Through **Internet Group Management Protocol (IGMP),** group members send group joining messages to the multicast network, so that the multicast network **obtains the locations of the group members and the multicast groups that the group members join**.
- This course describes the functions and implementation of IGMP, including the implementation and differences of IGMPv1, IGMPv2, and IGMPv3, and the implementation of IGMP snooping, IGMP SSM mapping, as well as IGMP proxy.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Understand IGMP implementation and configurations.
  - Understand the differences between different IGMP versions.
  - Understand the implementation of IGMP snooping.
  - Understand the implementation of IGMP SSM mapping.
  - Understand the implementation of IGMP proxy.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

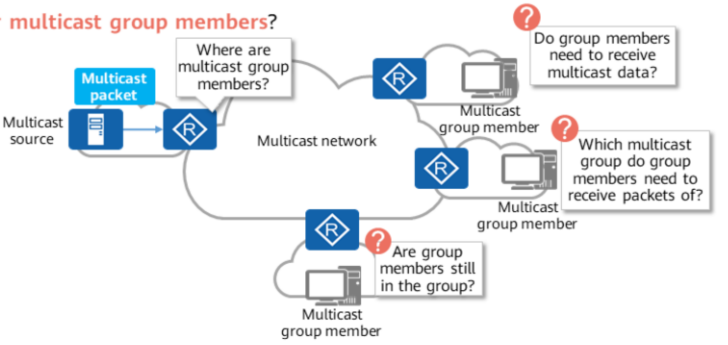
Contents

1. **IGMP Introduction**
   - Overview
     - Implementation of IGMP
2. IGMP Feature Introduction
3. IGMP Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Demands of Forwarding on a Multicast Network

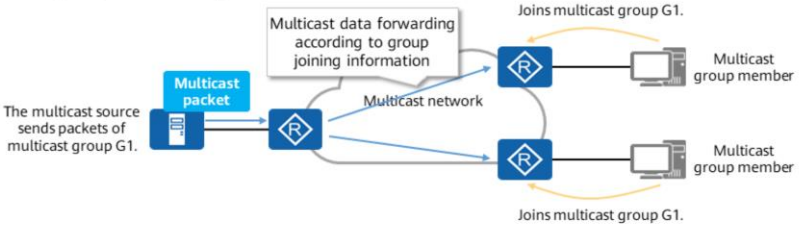
- IP multicast transmits packets from a source to a group of receivers. In the multicast communication model, the multicast source does not need to know the locations of receivers, and multicast data can be sent to multicast group members only through the multicast network.
- The multicast network needs to know the **locations of multicast group members** and the **multicast groups that the members join** so that the multicast network can forward multicast data to the members.
- How does the multicast network **discover multicast group members**?



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Multicast Group Member Discovery

- A multicast network can discover multicast group members in either of the following modes:
  - Manual static configuration: An interface connected to multicast group members is manually specified on a multicast router, and information about the multicast group members is statically configured.
    - The manual static mode is not flexible and requires heavy configuration workload. However, it features high stability and can quickly establish a multicast forwarding path for new group members.
  - Dynamic discovery: The multicast network obtains the interface connected to multicast group members and their group joining information through IGMP messages.
    - Being flexible and easy to configure, the dynamic discovery mode is widely used on the live network.
- After the multicast network obtains the locations of group members and the multicast groups that the group members join, multicast packets can be forwarded accordingly.

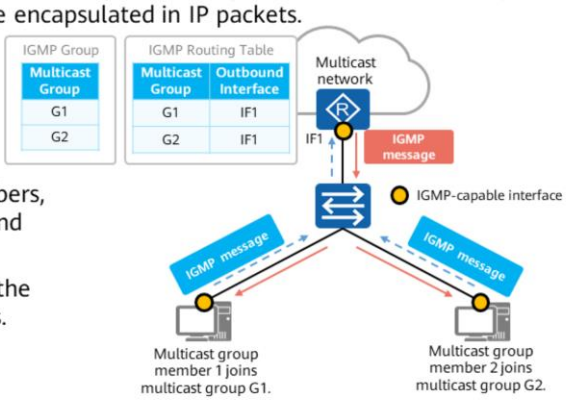


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Multicast packet forwarding on a multicast network depends on an MDT. For details about the MDT, see **PIM Implementation and Configurations**.

IGMP Overview

- In the TCP/IP protocol suite, IGMP is responsible for managing IP multicast members. It **establishes** and **maintains multicast group memberships** between IP hosts and the multicast routers that are directly adjacent to the IP hosts.
- IGMP manages multicast members through IGMP messages exchanged between multicast group members and multicast routers. IGMP messages are encapsulated in IP packets.
- There are three versions of IGMP:
  - IGMPv1
  - IGMPv2
  - IGMPv3
- After exchanging IGMP messages with group members, multicast routers generate **IGMP routing entries** and **IGMP group entries**.
- IGMP routing entries and IGMP group entries help the multicast routers generate multicast routing entries.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

IGMP Group Entries and Routing Entries

- IGMP generates IGMP routing entries and IGMP group entries, based on which multicast routing entries are generated.
  - An **IGMP group entry** is created when an IGMP Report message is received from a host. This entry is used to maintain group joining information and instruct multicast routing protocols (such as PIM) to create a corresponding (*, G) entry.
  - An IGMP group entry is as follows:

```
GigabitEthernet1/0/0(192.168.1.254):
 Total 1 IGMP Group reported
  Group Address   Last Reporter   Uptime     Expires
  239.0.0.1       192.168.1.1     00:02:04   00:01:17
 (Multicast group)      (IP address of a group member)
```

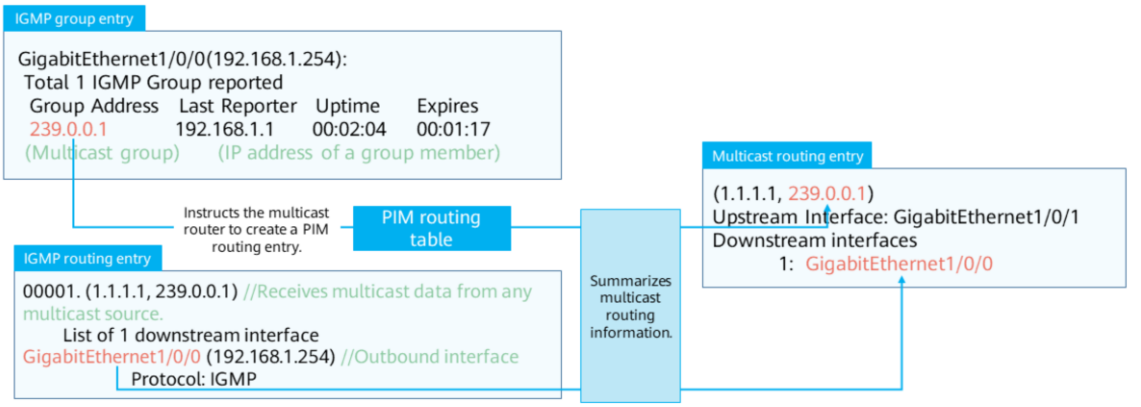
  - **IGMP routing entries** are used to extend the outbound interfaces of multicast routing entries.
  - An IGMP routing entry is as follows:

```
00001. (*, 239.0.0.1)      //Receives multicast data from any multicast source.
      List of 1 downstream interface
       GigabitEthernet1/0/0 (192.168.1.254) //Outbound interface
            Protocol: IGMP
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IGMP Entries and Multicast Routing Entries

- On the **last-hop multicast router (multicast leaf router)**, a multicast routing table is formed by summarizing IGMP routing entries, IGMP group entries, and the multicast protocol routing table (PIM routing table).
- IGMP routing entries and IGMP group entries provide **multicast group addresses** and **outbound interfaces** for the multicast routing table.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- For details about the PIM protocol, see **PIM Implementation and Configurations**.

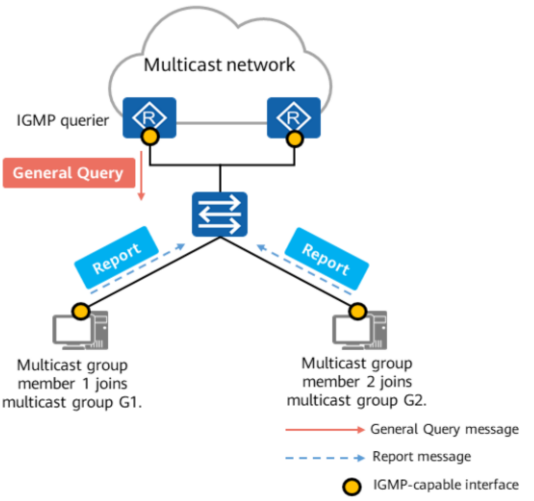
Contents

1. **IGMP Introduction**
   - Overview
   - Implementation of IGMP
2. IGMP Feature Introduction
3. IGMP Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Concepts of IGMPv1

- IGMPv1 uses a query-report mechanism to manage multicast groups.
- The query-report mechanism is implemented through two types of messages:
  - **General Query message**: a message sent by a querier to all hosts and routers on the local shared network to discover the multicast groups that have members.
  - **Report message**: Hosts send Report messages to the querier to **request to join a multicast group** or **respond to Query messages**.
- IGMP messages are multicast. Therefore, a multi-access network requires **only one multicast router to send Query messages**. This multicast router is called an **IGMP querier**.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IGMPv1 Message Format

- Both IGMPv1 General Query messages and Report messages are multicast messages with the **destination address 224.0.0.1**.
- IGMPv1 General Query messages are similar to Report messages in format, including three major fields: Version, Type, and Group Address.

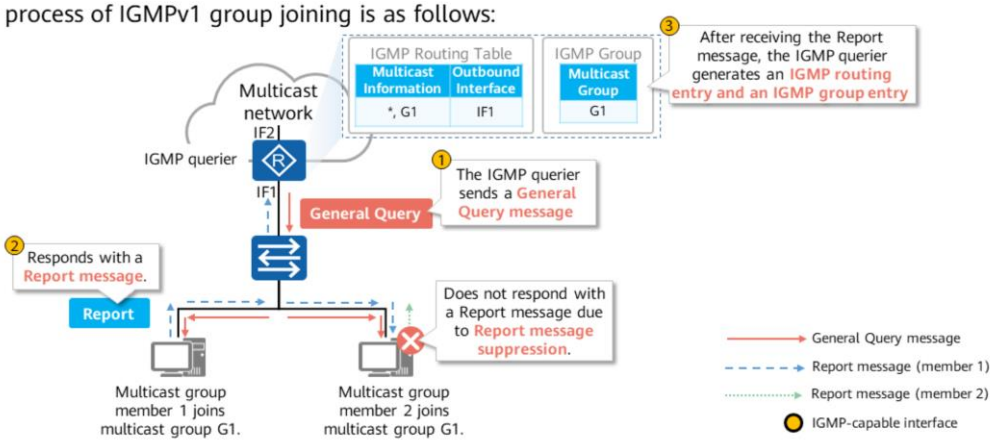


Field	Description
Version	IGMP version. In IGMPv1 messages, this field is 1.
Type	Message type. The following two message types are available: 0x1: General Query message 0x2: Report message
Group Address	Multicast group address. In a General Query message, this field is 0s. In a Report message, this field is set to the address of the multicast group that the member requests to join.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IGMPv1 Group Joining Mechanism

- Through General Query messages and Report messages, the IGMP querier can know which multicast groups have members on the network segment.
- The basic process of IGMPv1 group joining is as follows:



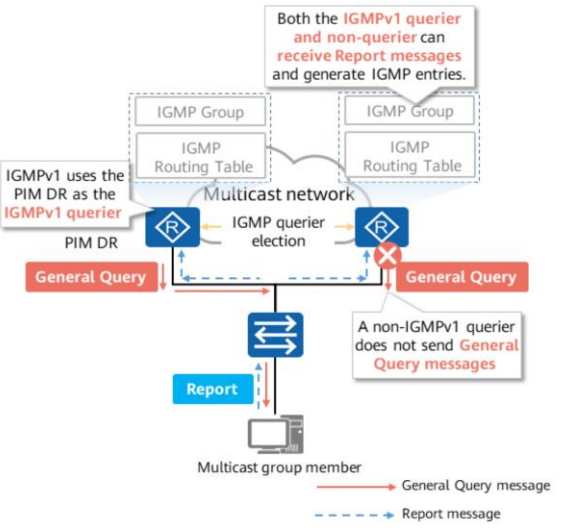
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The General Query and Report process is as follows:
  - The IGMP querier sends a General Query message, with destination address 224.0.0.1 (indicating all hosts and routers on the network segment). All group members start a timer when they receive the General Query message. The IGMP querier sends General Query messages at intervals. The interval is configurable, and the default interval is 60 seconds. Group members 1 and 2 are members of G1, and start Timer-G1 upon reception of the General Query message. By default, the value of the timer is a random value ranging from 0 to 10, in seconds.
  - The group member whose timer expires first sends a Report message for the group.
  - After receiving the Report message from group member 1, the IGMP querier knows that members of G1 exist on the local network segment. Then, the IGMP querier generates an IGMP group entry and (*, G1) IGMP routing entry. The asterisk (*) indicates any multicast source. Once the IGMP querier receives data of G1, it forwards the data to this network segment.
- Report message suppression mechanism:
  - The IGMP querier sends General Query messages at intervals. The interval is configurable, and the default interval is 60 seconds. Group members 1 and 2 are members of G1, and start Timer-G1 upon reception of the General Query message. By default, the value of the timer is a random value ranging from 0 to 10, in seconds.
  - Assuming that Timer-G1 on group member 1 expires first, group member 1 sends a Report message with G1 address as the destination address to the network segment. When group member 2 receives the Report message from group member 1, it stops Timer-G1 and does not send a Report message for G1. This mechanism reduces the number of Report messages transmitted on the network segment.

IGMPv1 IGMPv2 IGMPv3

IGMPv1 Querier Election Mechanism

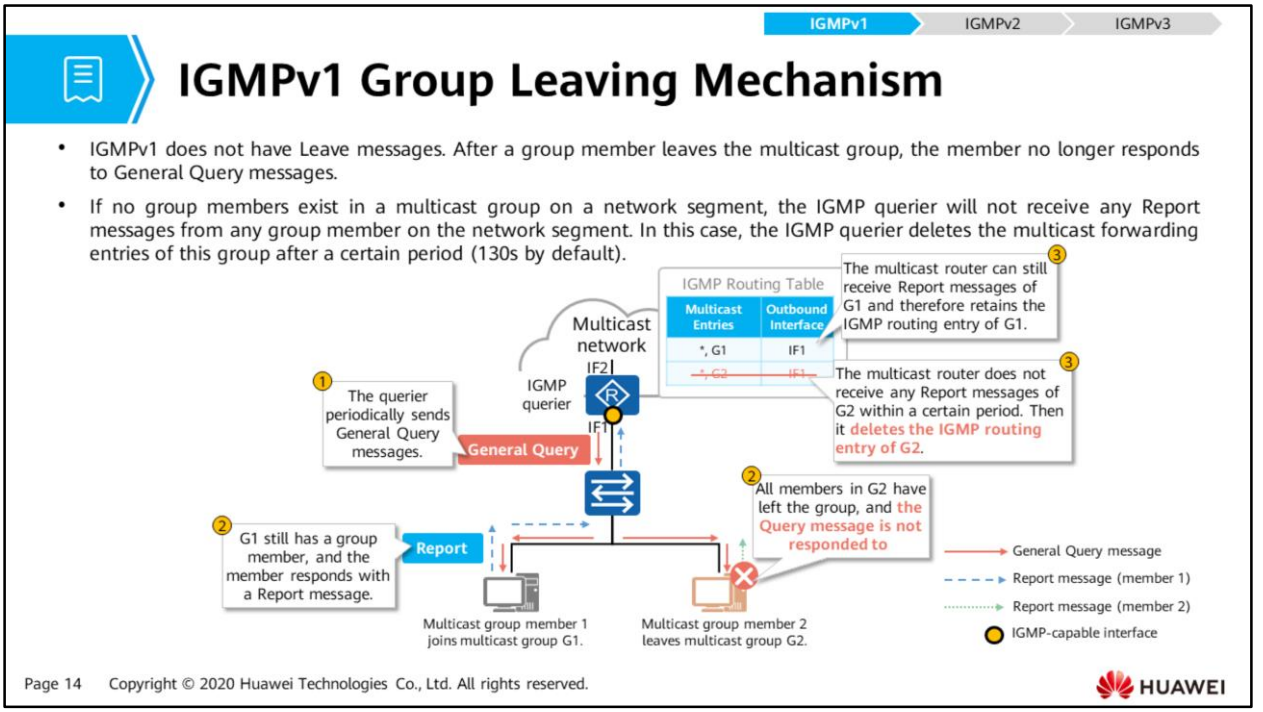
- General Query messages are multicast. Therefore, a network segment requires only one querier to query information about all group members.
- IGMPv1 **does not have its own querier election mechanism**. Instead, it relies on the multicast routing protocol (PIM) to elect an IGMP querier.
- IGMPv1 uses the unique **assert winner or DR elected by PIM as the querier**. The querier is the only device that sends Query messages on the local network segment.
- Both the querier and non-querier can receive Report messages with destination address 224.0.0.1. Therefore, they can generate both IGMP routing entries and IGMP group entries.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

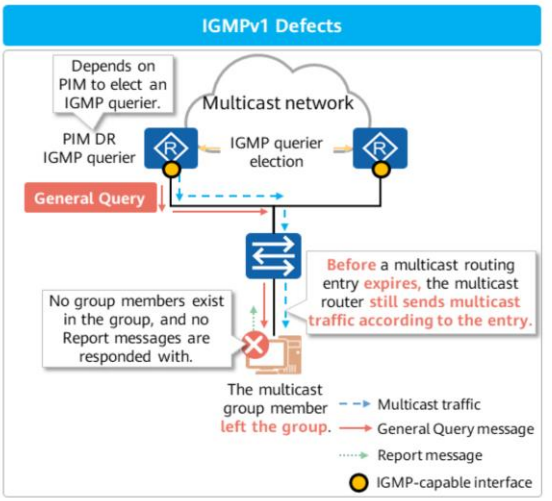
- The assert winner or DR is used to forward multicast traffic.
- Detailed functions of the assert winner or DR will be covered in the course of **PIM Implementation and Configurations**.



- Group Leaving Mechanism
  - Assume that group member 2 wants to leave multicast group G2.
    - When group member 2 receives the General Query messages from the IGMP querier, it does not respond with Report messages for G2. Because G2 no longer has members on this network segment, the IGMP querier will not receive Report messages for G2. After a certain period (130 seconds by default), the IGMP querier **deletes the IGMP routing entry of G2**.
    - When group member 1 receives the General Query messages from the IGMP querier, it responds with Report messages for G1. The IGMP querier then retains the corresponding IGMP routing entry of G1.

IGMPv2

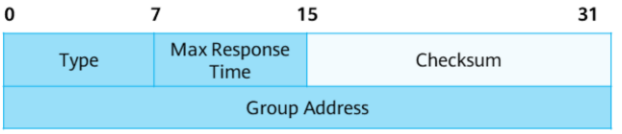
- IGMPv1 has the following defects in the group leaving and querier election mechanisms:
  - IGMPv1 uses the timeout mechanism to monitor group leaving status, and group members leave a group in silent mode. Before the timer expires, multicast traffic is still forwarded by the multicast router.
  - IGMPv1 querier election depends on PIM, which makes querier election inflexible.
- IGMPv2 overcomes the defects of IGMPv1.
  - IGMPv2 and IGMPv1 have similar group joining mechanisms.
  - IGMPv2 supports proactive group leaving.
  - IGMPv2 supports querier election.
- IGMPv2 is compatible with IGMPv1.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

IGMPv2 Message Format

- The following two types of messages are added to IGMPv2 to improve the group leaving mechanism:
  - Leave message: sent by a group member to notify the querier on the local network segment that it has left a group. The destination address of Leave messages is 224.0.0.2.
  - Group-Specific Query message: sent by a querier to a specified group on the local shared network to check whether the group has members. The destination address of Group-Specific Query messages is the address of the queried multicast group.
- A new field, Max Response Time, is added to IGMPv2 General Query messages. This field can be configured to control members' response speed to Query messages.
- The IGMPv2 message format is as follows:

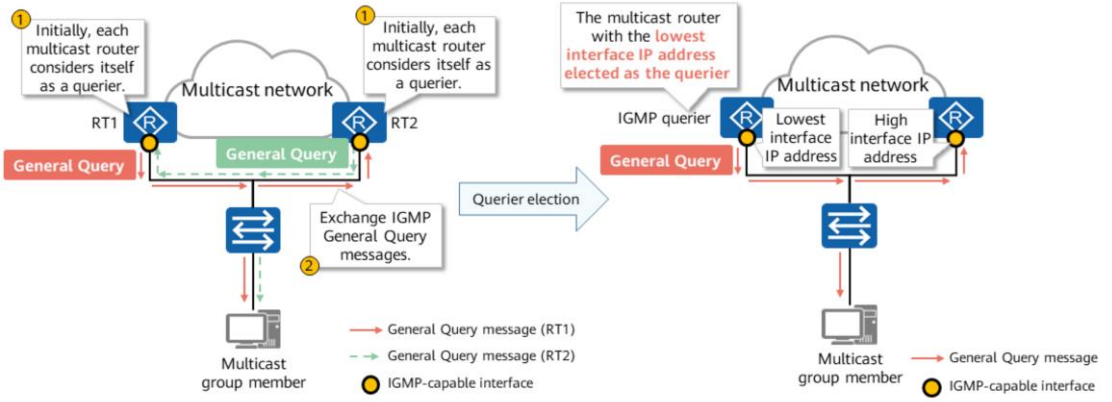


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Fields in an IGMPv2 message:
  - Type
    - Message type. The four message type options are:
    - 0x11: Query message. IGMPv2 Query messages include General Query and Group-Specific Query messages.
    - 0x12: IGMPv1 Report message.
    - 0x16: IGMPv2 Report message.
    - 0x17: Leave message.
  - Max Response Time: maximum time for a member to respond to a Query message with a Report message.
    - For a General Query message, the default maximum response time is 10 seconds.
    - For a Group-Specific Query message, the default maximum response time is 1 second.
  - Group Address:
    - In a General Query message, the group address is set to 0s.
    - In a Group-Specific Query message, the group address is the address of the queried group.
    - In a Report or Leave message, the group address is the address of the group that a member has joined or left.

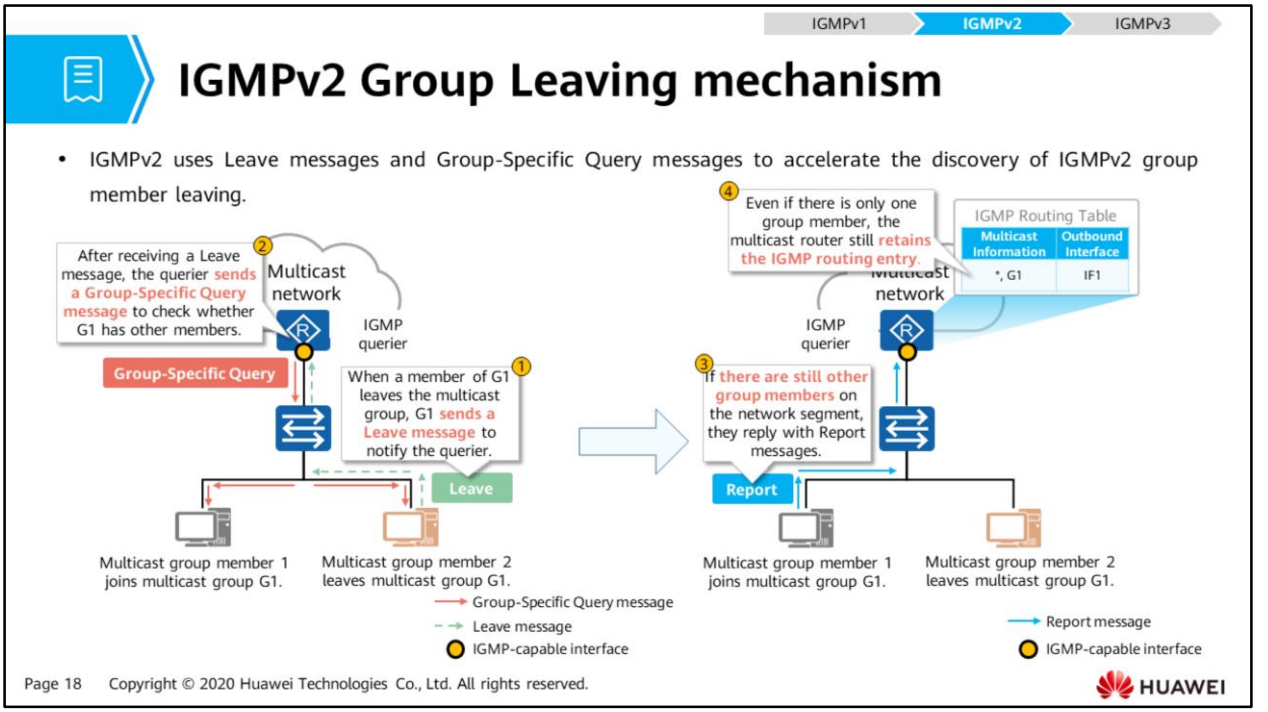
IGMPv2 Querier Election Mechanism

- The IGMPv2 group joining mechanism is the same as the IGMPv1 group joining mechanism, and is not mentioned here.
- The querier election mechanism in IGMPv2 varies greatly from that in IGMPv1. IGMPv2 has an independent querier election mechanism. When multiple multicast routers exist on a shared network segment, the router with the lowest IP address is elected as the querier.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

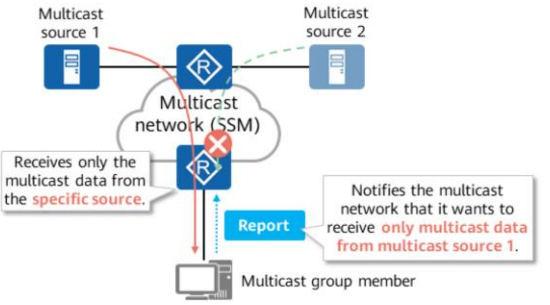
- Each non-querier starts a timer (Other Querier Present Timer). If a non-querier receives a Query message from the querier before the timer expires, it resets the timer; otherwise, it triggers a new round of querier election.



- A member sends a Leave message for G1 to all multicast routers on the local network segment. The destination address of the Leave message is 224.0.0.2.
    - When the querier receives the Leave message, it sends Group-Specific Query messages for G1 to check whether G1 has other members on the network segment. The Group-Specific Query interval and Count are configurable. By default, the querier sends Group-Specific Query messages twice, at an interval of 1s. In addition, the querier starts the group membership timer (Timer-Membership). The value of the timer is the Group-Specific Query interval multiplied by Count.
    - If G1 still has other members on the network segment, when receiving a Group-Specific Query message from the querier, they immediately respond with a Report message for G1. The querier then keeps maintaining the membership of G1 after receiving the Report message.
    - If G1 has no members on the network segment, the querier will not receive any Report message for G1. When the Timer-Membership expires, the querier deletes the (*, G1) entry. Thereafter, if the querier receives multicast data of G1, it does not forward the data downstream.

Demand for the SSM Model

- To ensure security, multicast group members can choose to receive only the multicast data from a specific multicast source. In this case, they need to inform the multicast network of the multicast source from which they want to receive multicast data.
- IGMPv1 and IGMPv2 messages cannot carry multicast source information. As a result, IGMPv1 and IGMPv2 cannot meet the demand for source-specific multicast (SSM), unless SSM mapping is used.
- IGMPv3 was developed to support the SSM model. IGMPv3 messages can carry multicast source information so that hosts can receive only data from a specific source.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In the SSM model, multicast addresses range from 232.0.0.0 to 232.255.255.255.
- For details about SSM mapping, see the chapter of "IGMP Features."



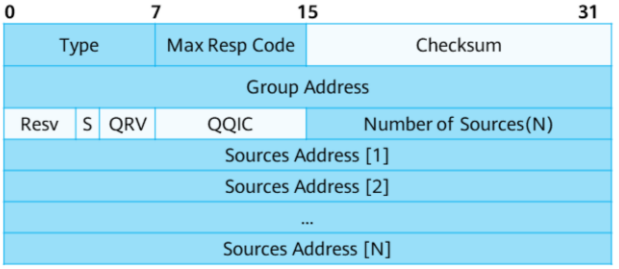
IGMPv3

- Most mechanisms of IGMPv3 are similar to those of IGMPv2.
  - Their querier election mechanisms are the same. The multicast router with the lowest IP address is elected as the querier.
  - General Query messages are used to query group memberships.
  - Group-Specific Query messages are used to query group memberships of a specific multicast group.
- IGMPv3 needs to report multicast source information. Compared with IGMPv2, IGMPv3 has the following changes:
  - In addition to General Query and Group-Specific Query messages, IGMPv3 has a new Query message type: Group-and-Source-Specific Query.
  - Each IGMPv3 Report message contains the group that a host wants to join and the multicast sources from which the host wants to receive data.
  - Different members in the same multicast group may want to receive multicast data from different sources. Therefore, IGMPv3 **does not require the Report message suppression mechanism**.
  - Unlike IGMPv2, **IGMPv3 does not define a Leave message**. Group members send Report messages of a specified type to notify multicast routers that they have left a group.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IGMPv3 Query Message Format

- There are three types of IGMPv3 Query messages:
  - **General Query message**. This type of message has the same function as that in IGMPv1 or IGMPv2.
  - **Group-Specific Query message**. This type of message has the same function as that in IGMPv2.
  - **Group-and-Source-Specific Query message**. This type of message is used to query whether a group member wants to receive data from specific sources. Each such a message carries one or more multicast source addresses.
- The IGMPv3 Query message format is as follows:

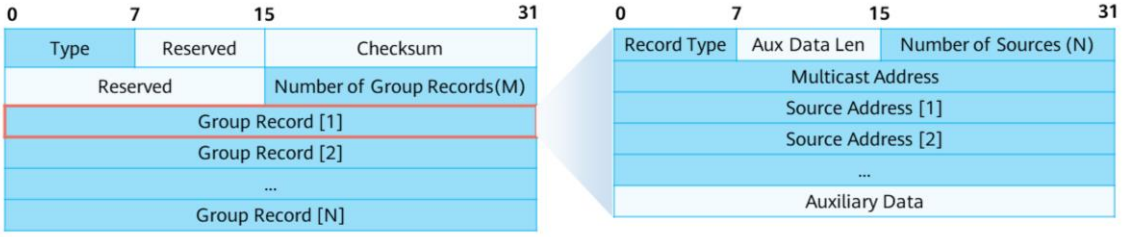


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Key fields in an IGMPv3 Query message:
  - Type: message type. In IGMPv3 Query messages, this field is set to 0x11.
  - Max Response Time: maximum response time. After receiving a General Query message, hosts must respond with a Report message within the maximum response time.
  - Group Address: address of a multicast group. In a General Query message, this field is set to 0. In a Group-Specific Query or Group-and-Source-Specific Query message, this field is set to the IP address of the queried group.
  - Number of Sources: number of multicast sources contained in the message. In a General Query or Group-Specific Query message, this field is set to 0. In a Group-and-Source-Specific Query message, this field is not 0. This number is limited by the maximum transmission unit (MTU) of the network over which the Query message is transmitted.
  - Source Address: address of the multicast source. The value is subject to the Number of Sources field.

IGMPv3 Report Message Format

- IGMPv3 Report messages are used to notify the querier of group memberships and the multicast sources from which the members want to receive multicast data. Multicast source information can be notified in either of the following modes:
  - INCLUDE: The members want to receive multicast data from specified multicast sources.
  - EXCLUDE: The members want to have multicast data from specified multicast sources filtered out.
- The relationship between multicast group information and multicast source information in the Report message is recorded in the Group Record field and sent to the IGMP querier. The destination address of each IGMPv3 Report message is 224.0.0.22. The format of the message is as follows:



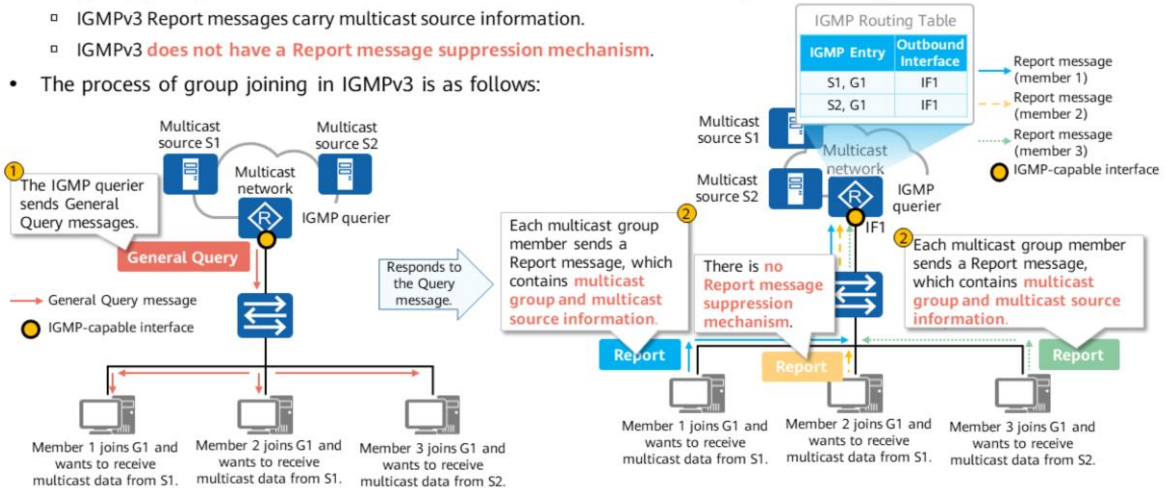
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- An IGMPv3 Report message can carry multiple groups, whereas an IGMPv1 or IGMPv2 Report message can carry only one group. Therefore, the number of IGMPv3 messages needed is greatly reduced.
- The key fields in an IGMPv3 Report message are described as follows:
  - Type: message type. In IGMPv3 Report messages, this field is set to 0x22.
  - Number of Group Records: number of group records contained in a message.
  - Group Record: group record.
- Key fields in Group Record are described as follows:
  - Record Type: type of a group record. There are three group record types.
    - Current-State Record. It is used to respond to Query messages and advertise its current state. There are two types of states. One of the states is MODE_IS_INCLUDE, indicating that the member wants to receive only the multicast data sent from the sources in the source address list to the group. If the specified source address list is empty, the message is invalid. The other state is MODE_IS_EXCLUDE, indicating that the member rejects the multicast data sent from the sources in the source address list to the group.
    - Filter-Mode-Change Record. In the case of a switchover between INCLUDE and EXCLUDE, the querier is notified of the filtering mode change. There are two filtering mode changes. One is CHANGE_TO_INCLUDE_MODE, indicating that the filtering mode is changed from EXCLUDE to INCLUDE; in this case, the member wants to receive the data sent by the multicast sources in the source address list to the multicast group. If the specified source address list is empty, the member will leave the multicast group. The other change is CHANGE_TO_EXCLUDE_MODE, indicating that the filtering mode is changed from INCLUDE to EXCLUDE; in this case, the member rejects the multicast data sent from the multicast sources in the source

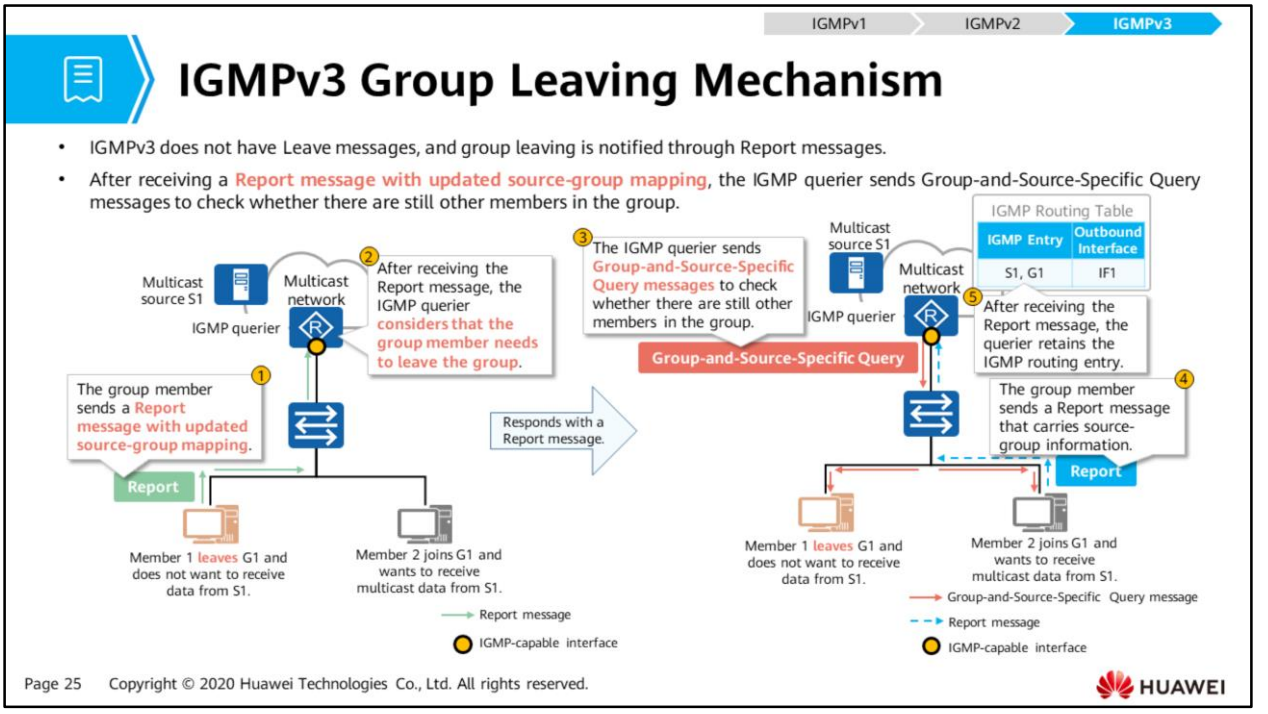
address list to the multicast group.

IGMPv3 Group Joining Mechanism

- The group joining mechanism of IGMPv3 is similar to that of IGMPv2 except the following differences:
  - IGMPv3 Report messages carry multicast source information.
  - IGMPv3 does not have a Report message suppression mechanism.
- The process of group joining in IGMPv3 is as follows:



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI



- Various Report messages can be used to update source-group mapping. For example:
  - A member used to receive multicast data from S1. It can send a (G1, EXCLUDE, S1) or (G1, CHANGE_TO_EXCLUDE_MODE, S1) message to update source-group mapping.

Differences Between IGMP Versions

Mechanism	IGMPv1	IGMPv2	IGMPv3
Querier election	Depends on another protocol.	Uses its own mechanism.	Uses its own mechanism.
Group leaving	Leave silently.	Leave proactively.	Leave proactively.
Group-Specific Query	Not supported	Supported	Supported
Source-and-Group-Specific Query	Not supported	Not supported	Supported
Version mapping		IGMPv1	IGMPv1, IGMPv2

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

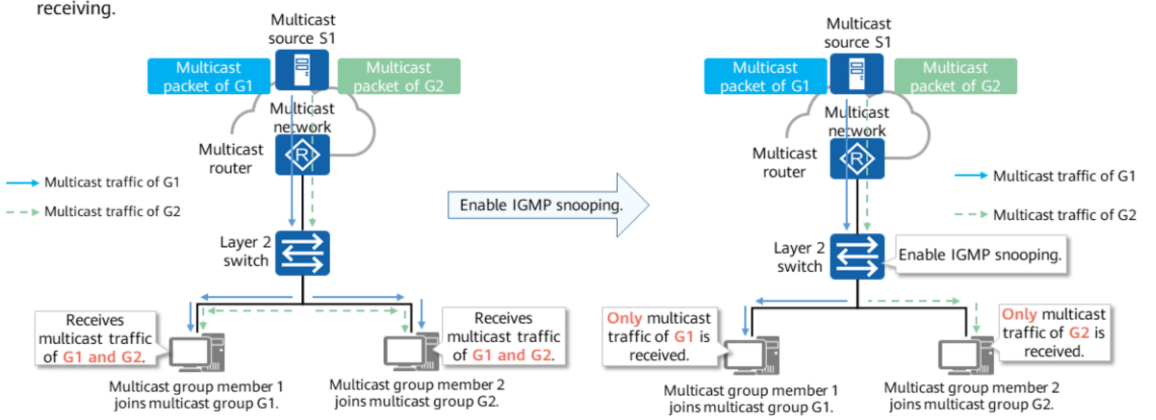
Contents

1. IGMP Introduction
2. **IGMP Feature Introduction**
   - IGMP Snooping Introduction
   - IGMP SSM Mapping Introduction
   - IGMP Proxy Introduction
3. IGMP Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Problem Accompanying Multicast Forwarding on the Ethernet Network

- The multicast data packets sent from the last-hop router to multicast group members usually pass through a switch. The destination MAC addresses of multicast data packets are multicast MAC addresses. By default, the switch floods such data frames. As a result, multicast traffic may be received across groups.
- IGMP snooping controls the flooding scope of multicast traffic on an Ethernet network, which prevents inter-group multicast traffic receiving.

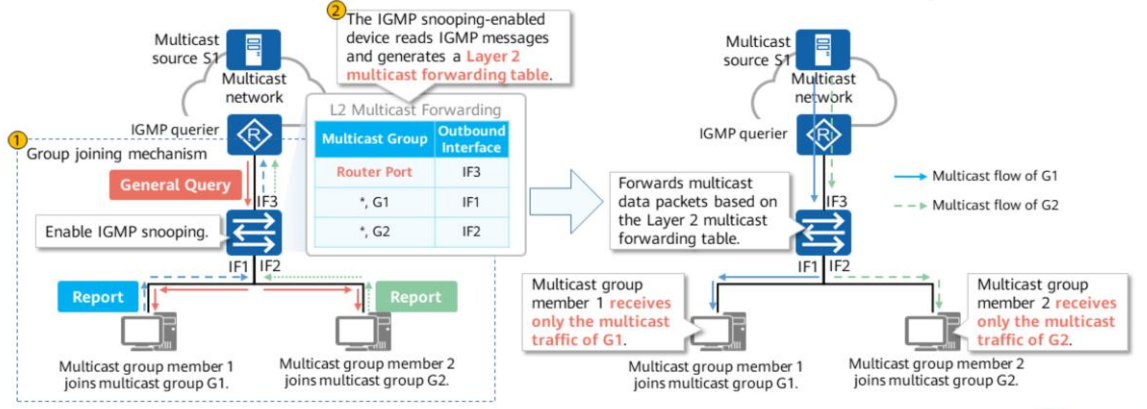


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- After receiving multicast data packets from the router, the switch forwards the packets to the group members. Destination addresses of multicast packets are multicast group addresses and cannot be learned by a Layer 2 switch. Therefore, when a Layer 2 switch receives multicast packets from a router, it broadcasts the packets in the broadcast domain. All hosts in the broadcast domain receive the multicast packets, regardless of whether they are group members. This wastes network bandwidth and poses security risks.
- IGMP snooping solves this preceding problem. With IGMP snooping configured, the Layer 2 multicast switch listens to and analyzes IGMP messages exchanged between multicast users and the upstream router, and creates Layer 2 multicast forwarding entries accordingly. Multicast data packets are then forwarded based on the Layer 2 multicast forwarding entries. This prevents multicast data packets from being broadcast on the Layer 2 network.

IGMP Snooping Introduction

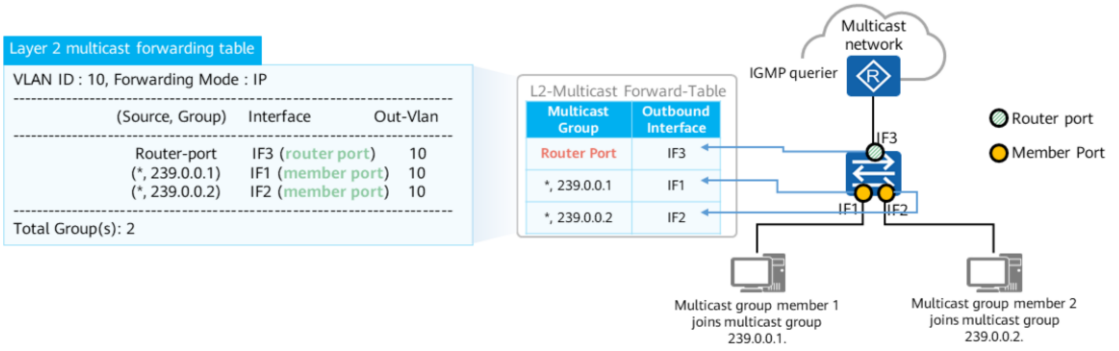
- IGMP snooping implements forwarding and control of multicast data packets at the data link layer.
- IGMP snooping runs on a Layer 2 multicast device and analyzes received IGMP messages exchanged between a Layer 3 device and hosts to create and maintain a Layer 2 multicast forwarding table. Based on this table, the Layer 2 device forwards multicast packets at the data link layer.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

IGMP Snooping Port and Forwarding Table Introduction

- Layer 2 multicast forwarding entries contain the following types of ports:
  - Router port: a port on a Layer 2 multicast device, used to receive multicast packets from a Layer 3 multicast device such as a DR or IGMP querier
  - Member port: a port on a Layer 2 multicast device, used to send multicast data packets to group members.

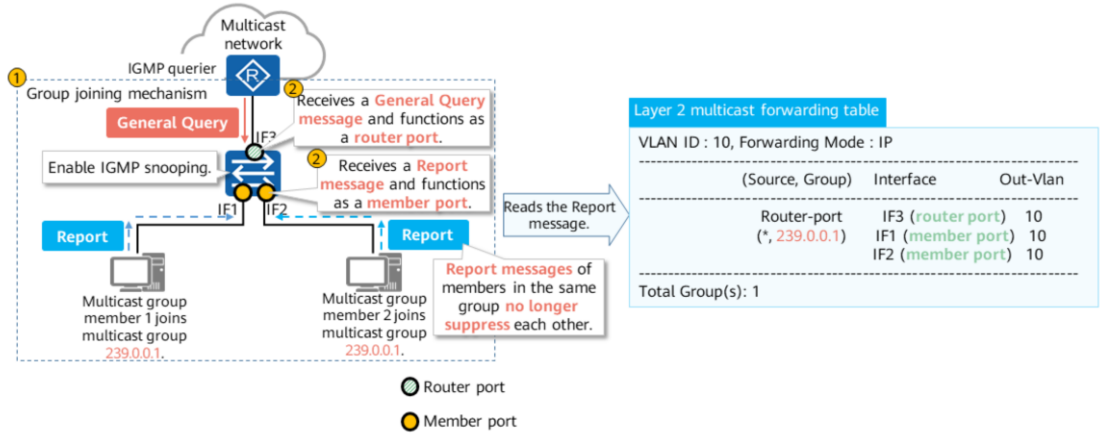


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Router port:
  - A router port generated by a protocol is called a dynamic router port. A port becomes a dynamic router port when it receives an IGMP General Query message or PIM Hello message with any source address except 0.0.0.0. The PIM Hello messages are sent from the PIM port on a Layer 3 multicast device to discover and maintain neighbor relationships.
  - A manually configured router port is called a static router port.
- Member port:
  - A member port generated by a protocol is called a dynamic member port. A Layer 2 multicast device sets a port as a dynamic member port when the port receives an IGMP Report message.
  - A manually configured member port is called a static member port.

Implementation of IGMP Snooping — Forwarding Entry Generation

- An IGMP snooping-enabled device listens for IGMP messages to generate a Layer 2 multicast forwarding table and determines the port type. The detailed process is as follows:

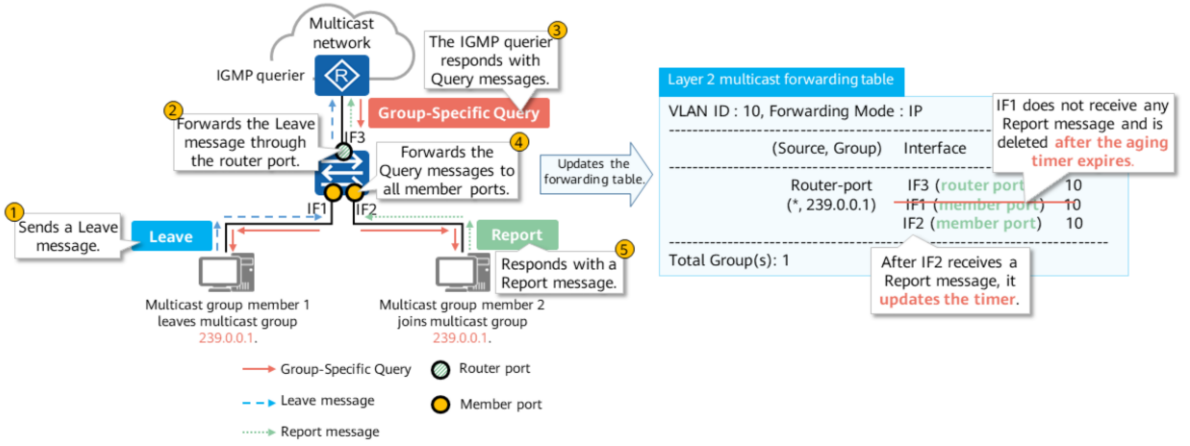


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- When a router port takes effect, an aging timer (180s by default) is started. When the router port receives a new General Query message, it updates the timer.
- When a member port takes effect, an aging timer (180s by default) is started. When the member port receives a new Report message, it updates the timer.
- IGMP snooping no longer uses the Report message suppression mechanism.
  - IGMP snooping needs to listen for IGMP messages to determine the port role and guide packet forwarding. Therefore, all group members need to send Report messages.
  - After receiving a Report message, the IGMP snooping-enabled device forwards the Report message only through the router port. This prevents other group members in this group from receiving the Report message, which will not trigger Report message suppression.

Implementation of IGMP Snooping — Forwarding Entry Maintenance

- An IGMP snooping-enabled device listens for IGMP Leave messages and IGMP Report messages to determine whether a specific interface needs to send multicast data. The detailed process is as follows:



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- After receiving an IGMP Leave message, the switch uses the following formula to calculate the aging timer of member ports: Aging timer = Robustness variable (2 by default) x Group-Specific Query interval (1s by default).

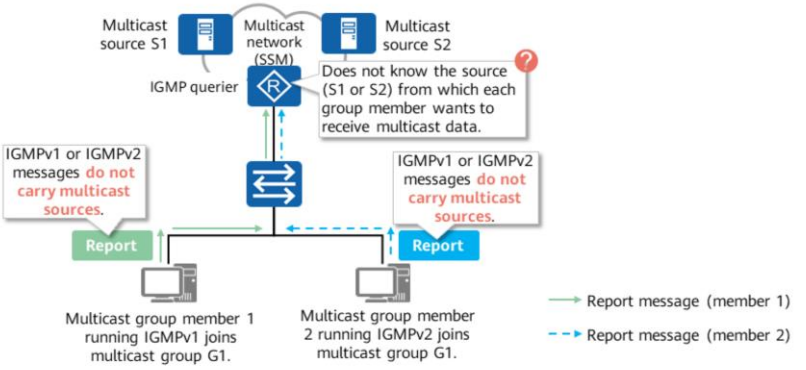
Contents

1. IGMP Introduction
2. **IGMP Feature Introduction**
   - IGMP Snooping Introduction
   - IGMP SSM Mapping Introduction
   - IGMP Proxy Introduction
3. IGMP Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IGMP SSM Mapping Introduction

- If terminals on the live network do not support IGMPv3, multicast in SSM mode cannot be deployed because IGMPv1 or IGMPv2 messages cannot carry multicast source information.
- IGMP SSM mapping allows IGMPv1 and IGMPv2 group members to access the SSM network if a multicast source is manually bound to a multicast group.

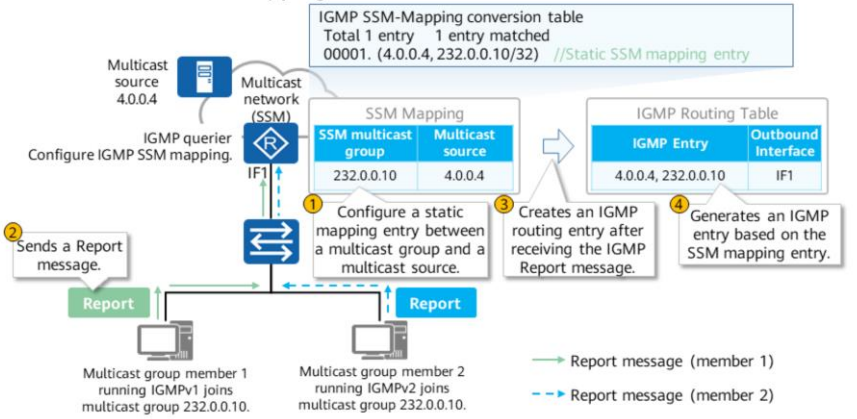


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The SSM group addresses range from 232.0.0.0 to 232.255.255.255, regardless of whether IGMPv1, IGMPv2, or IGMPv3 is used.

Implementation of IGMP SSM Mapping

- Configure static SSM mapping entries on the IGMP querier to map group information in IGMPv1 or IGMPv2 Report messages to source-specific group information.
- The implementation of IGMP SSM mapping is as follows:



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- With SSM mapping entries configured, an IGMP querier checks the group address G in each IGMPv1 or IGMPv2 Report message received, and processes the message based on the check result:
  - If G is in the any-source multicast (ASM) group address range, the router provides the ASM service for the corresponding group member.
  - If G is in the SSM group address range (232.0.0.0 to 232.255.255.255 by default):
    - When the IGMP querier has no SSM mapping entry matching G, it does not provide the SSM service and drops the Report message.
    - If the IGMP querier has an SSM mapping entry matching G, it converts (*, G) information in the Report message into (G, INCLUDE, (S1, S2...)) information and provides the SSM service for the corresponding group member.
- IGMP SSM mapping does not apply to IGMPv3 Report messages. To enable hosts running any IGMP version on a network segment to obtain the SSM service, IGMPv3 must run on interfaces of multicast routers on the network segment.

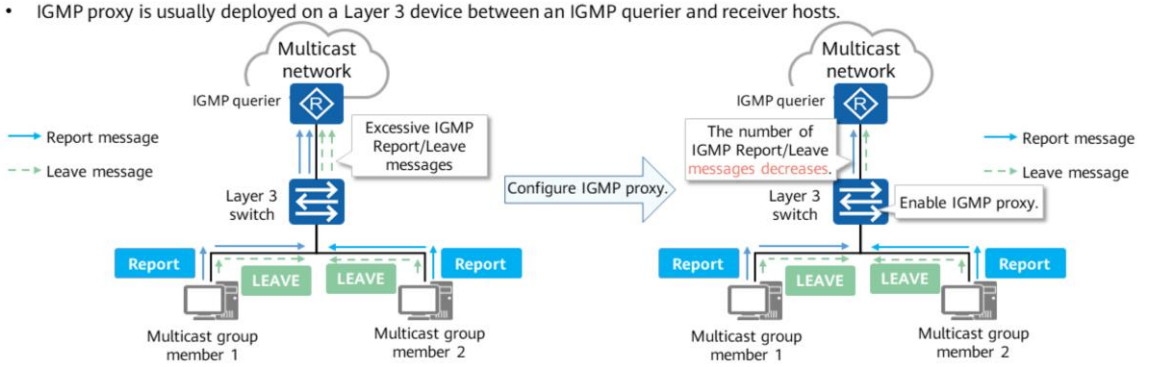
Contents

1. IGMP Introduction
2. **IGMP Feature Introduction**
   - IGMP Snooping Introduction
   - IGMP SSM Mapping Introduction
   - IGMP Proxy Introduction
3. IGMP Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IGMP Proxy Introduction

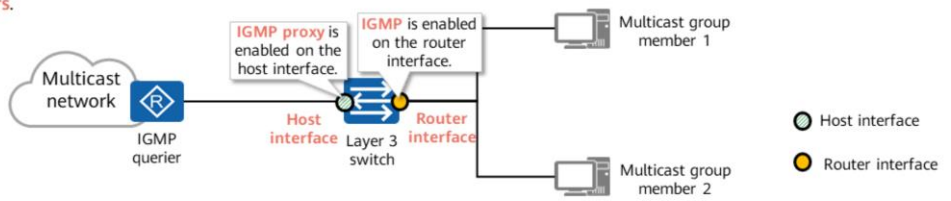
- On a live network, an IGMP querier may need to manage a large number of group members. When a large number of hosts frequently join or leave groups, a large number of IGMP Report/Leave messages are sent to the querier, which burdens the querier.
- IGMP proxy reduces the number of IGMP Report/Leave messages received by the IGMP querier, reducing the burden on the IGMP querier.
- IGMP proxy is usually deployed on a Layer 3 device between an IGMP querier and receiver hosts.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Basic Concepts of IGMP Proxy

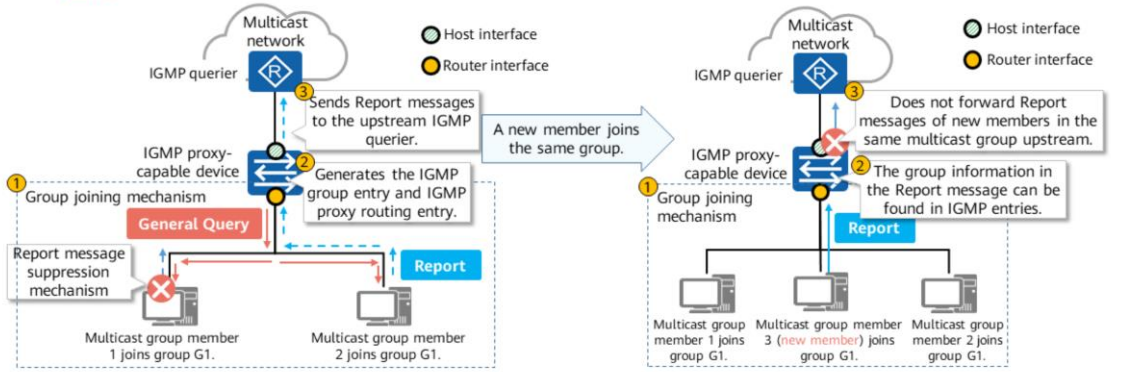
- To relieve the burden on the IGMP querier, the IGMP proxy-capable device aggregates Report/Leave messages before sending them to the IGMP querier.
- An IGMP proxy-capable device can also act as an IGMP querier to send Query messages to receiver hosts, maintain group memberships, and forward multicast data based on group memberships.
- To implement the preceding functions, IGMP proxy defines two types of interfaces:
  - **Host interface**: IGMP proxy-enabled interface on an IGMP proxy-capable device. This interface generally **points to an IGMP querier**.
  - **Router interface**: an interface configured with IGMP on an IGMP proxy-capable device. This interface generally **points to group members**.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Implementation of IGMP Proxy – Group Joining

- IGMP proxy reduces the number of Report messages as follows:
  - A router interface functions as an IGMP interface and is presented as an IGMP querier. The router interface sends Query messages, processes Report messages, generates IGMP entries, and sends Report messages to the upstream IGMP querier through a host interface.
  - When new hosts join the same multicast group, the IGMP proxy-capable device does not send Report messages to the IGMP querier. This reduces the number of Report messages sent to the IGMP querier.

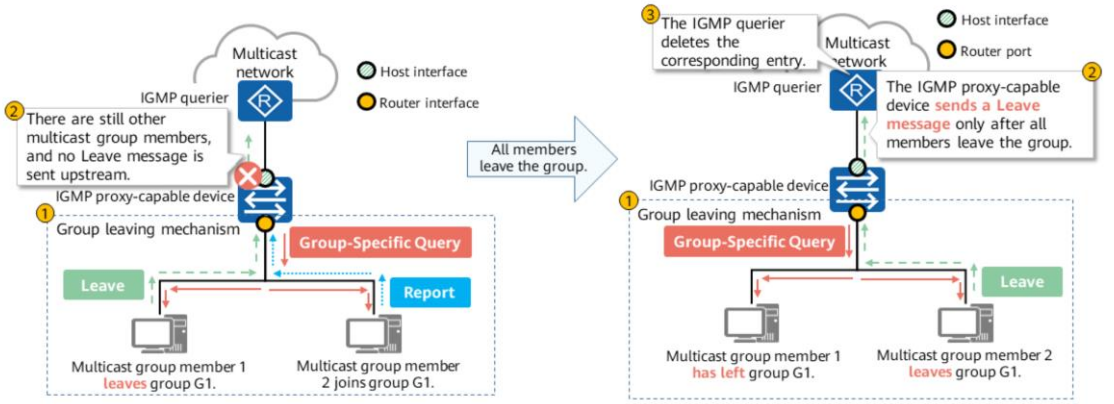


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- When the IGMP proxy-capable device receives a Report message for a group, it searches the multicast forwarding table for the group.
  - If the group is not found in the multicast forwarding table, the IGMP proxy-capable device sends a Report message for the group to the access device and adds the group to the multicast forwarding table.
  - If the group is found in the multicast forwarding table, the IGMP proxy device does not send a Report message to the access device.
- IGMPv1, IGMPv2, and IGMPv3 group joining mechanisms are not described here.

Implementation of IGMP Proxy – Group Leaving

- IGMP proxy reduces the number of Leave messages as follows:
  - When a group member leaves a multicast group, the IGMP proxy-capable device uses the IGMP group leaving mechanism to determine whether there are still other members of the multicast group. If there is no member of the multicast group, the device sends a Leave message to the upstream IGMP querier.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- When the IGMP proxy-capable device receives a Leave message for G1, it sends a Group-Specific Query message through the interface where the Leave message was received, to check whether G1 has other members attached to the interface.
  - If there are no other members of G1 attached to the interface, the IGMP proxy-capable device deletes the interface from the forwarding entry of G1. The IGMP proxy-capable device then checks whether G1 has members on other interfaces.
    - If G1 has no members on other interfaces, the IGMP proxy-capable device sends a Leave message for G1 to the access device.
    - If G1 has members on other interfaces, the IGMP proxy-capable device does not send a Leave message for G1 to the access device.
  - If the group has other members attached to the interface, the IGMP proxy-capable device continues forwarding multicast data to the interface.
- IGMPv1, IGMPv2, and IGMPv3 group leaving mechanisms are not described here.

Contents

1. IGMP Introduction
2. IGMP Feature Introduction
3. **IGMP Configurations**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Configurations of IGMP (1)

1. Enable IGMP.

```
[Huawei - GigabitEthernet1/0/0] igmp enable
```

2. Configure an IGMP version.

```
[Huawei - GigabitEthernet1/0/0] igmp version number
```

3. Enable IGMP snooping globally.

```
[Huawei] igmp-snooping enable
```

4. Enable IGMP proxy in a VLAN.

```
[Huawei-vlan10] igmp-snooping proxy
```

5. Enable IGMP SSM mapping on an interface.

```
[Huawei - GigabitEthernet1/0/0] igmp ssm-mapping enable
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Configurations of IGMP (2)

6. Configure a static SSM mapping entry in the IGMP view.

```
[Huawei-igmp] ssm-mapping group-address group-mask source-address
```

7. Check the IGMP configurations and running information on interfaces.

```
<Huawei>display igmp interface
```

8. Check information about the members of a multicast group.

```
<Huawei>display igmp group
```

9. Check the Layer 2 multicast forwarding table.

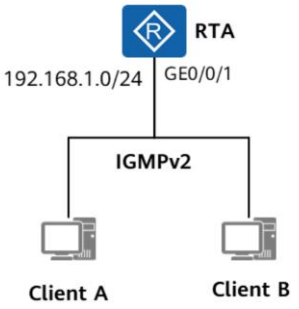
```
<Huawei>display l2-multicast forwarding-table vlan vlan-id
```

10. Check static IGMP SSM mapping information.

```
<Huawei>display igmp ssm-mapping
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IGMP Basis Experiment



Configurations on RTA

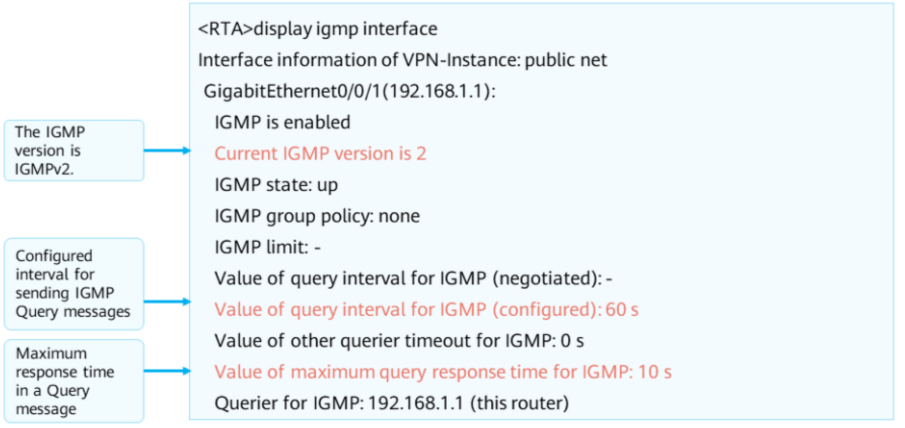
```
[RTA]multicast routing-enable
[RTA]interface g0/0/1
[RTA-GigabitEthernet0/0/1]ip address 192.168.1.1 24
[RTA-GigabitEthernet0/0/1]igmp enable
[RTA-GigabitEthernet0/0/1]igmp version 2
```

Configure IGMP on the router to implement IP multicast member management.
Enable IGMP on GE 0/0/1 of RTA so that RTA functions as a querier to send and receive IGMP messages.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IGMP Configuration Verification

On RTA, check the IGMP configurations and running information on the interface.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Essay) In IGMPv1, when the last member leaves a group, how long will the multicast router wait before deleting the multicast forwarding entry of the group?
2. (Essay) Is the destination IP address of IGMPv2 Group-Specific Query message 224.0.0.1?
3. (Single) Which IGMP versions are available?
   A. IGMPv1, IGMPv2
   B. IGMPv3
   C. IGMPv2, IGMPv3
   D. IGMPv1, IGMPv2, IGMPv3

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. 60 x 2 + 10 = 130s
2. No. The destination IP address of a Group-Specific Query message is the IP address of the group to be queried.
3. D

Summary

- IGMP enables a multicast network to know the locations of multicast group members and the multicast groups that the members join. In addition, IGMP maintains multicast group memberships.
- There are three IGMP versions:
  - IGMPv1: has a basic group joining mechanism, but the group member leaving mechanism is outdated. In addition, IGMPv1 does not have an independent IGMP querier election mechanism.
  - IGMPv2: improves the group leaving mechanism based on IGMPv1 and provides an independent IGMP querier election mechanism.
  - IGMPv3: is an enhancement of IGMPv2 and supports SSM. IGMPv3 can notify a multicast network of the multicast sources from which multicast data needs to be received.
- IGMP provides the following features to ensure efficient running of IGMP on a network:
  - IGMP snooping: prevents multicast data from being flooded across multicast groups on an Ethernet network.
  - IGMP SSM mapping: solves the problem that IGMPv1 and IGMPv2 group members cannot access the SSM multicast network.
  - IGMP proxy: solves the problem that the IGMP querier is overloaded because a large number of multicast group members frequently join and leave a multicast group.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



PIM Implementation and Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- A multicast network consists of multicast sources, multicast group members, and multicast routers.
  - A multicast source is mainly used to send multicast data.
  - Multicast group members receive multicast data. Therefore, IGMP must be used to notify the multicast network of the locations of group members and the groups that the members join.
  - Multicast routers **forward data from a multicast source to multicast group members**. Multicast data forwarding **depends on multicast distribution trees (MDTs)**. Therefore, multicast routers need to use protocols to **establish MDTs**.
- The Protocol Independent Multicast (PIM) protocol is mainly used to **establish MDTs**.
- This course describes basic concepts of PIM and fundamentals of PIM-DM and PIM-SM.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe the basic concepts of PIM.
  - Describe the fundamentals of PIM-DM.
  - Perform basic configurations of PIM-DM.
  - Describe the fundamentals of PIM-SM.
  - Perform basic configurations of PIM-SM.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

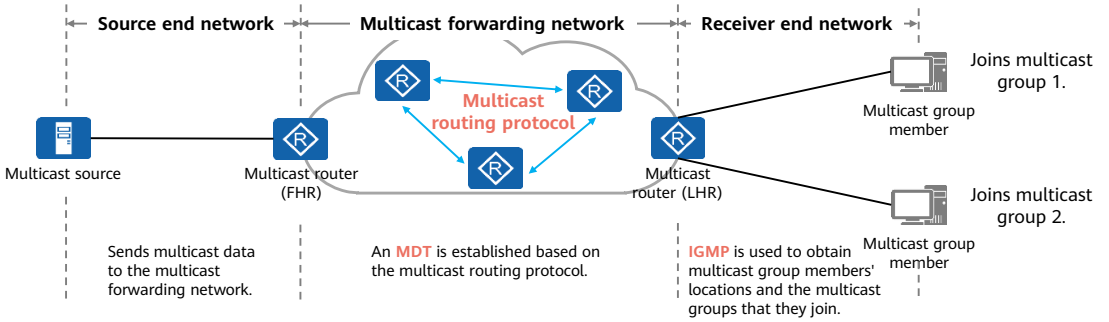
1. **Introduction to PIM**
2. Introduction to PIM-DM
3. Introduction to PM-SM

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Multicast Network Architecture Review

A multicast network can be divided into three parts:

- Source end network: sends multicast data generated by the multicast source to the multicast forwarding network.
- Multicast forwarding network: **generates a loop-free multicast forwarding path**, which is also referred to as an **MDT**.
- Receiver end network: uses **IGMP** to enable the multicast network to **detect the locations of multicast group members** and the multicast groups that the members join.

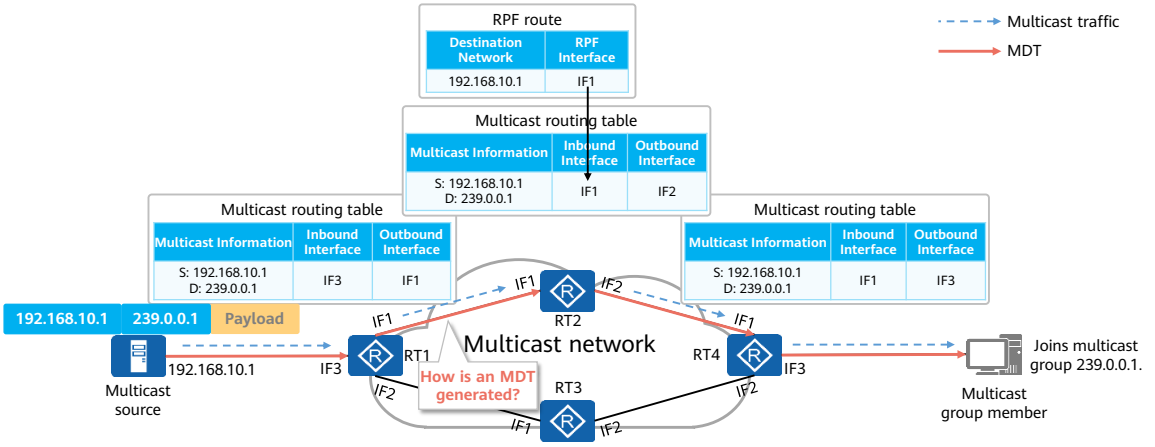


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The receiver end network uses IGMP to enable the multicast network to detect the locations of multicast group members and the multicast groups that the members join.
- MDT establishment on the multicast forwarding network requires a multicast routing protocol.
- There are multiple multicast routing protocols. The most commonly used one is PIM, which is the focus of this course.

Multicast Data Forwarding Process Review

The multicast data forwarding process is as follows:



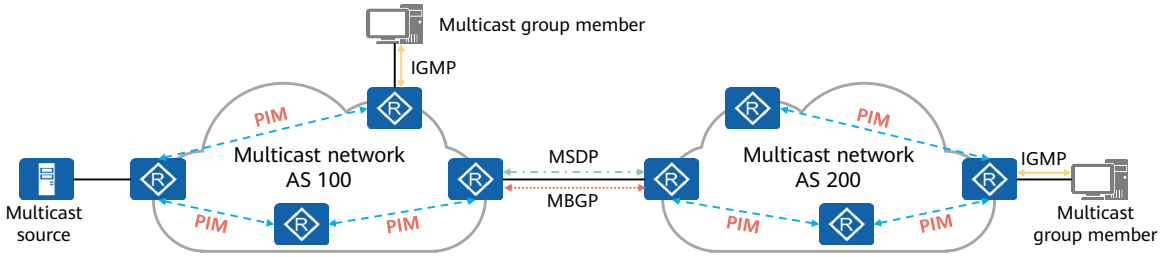
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The course "IGMP Implementation and Configurations" describes how a multicast network discovers the locations of multicast group members and the multicast groups that the members join.
- This course mainly describes how an MDT is established.

Multicast Protocol Review

A multicast network needs to establish forwarding paths based on multiple multicast protocols.

- IGMP runs on the receiver end network and is used to inform the multicast network of the locations of group members and the multicast groups that the members join.
- Protocols working on the multicast forwarding network include PIM, MSDP, and MBGP.
  - **PIM** is mainly used to **generate MDTs in an AS**.
  - MSDP is mainly used to help generate inter-AS MDTs.
  - MBGP is used to help perform RPF check on inter-AS multicast traffic.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- This course mainly describes the implementation of PIM.

PIM Introduction

- PIM is a multicast routing solution that performs a reverse path forwarding (RPF) check on multicast packets based on the unicast routing table, creates multicast routing entries if multicast packets pass the RPF check, and forwards the multicast packets using multicast routing entries. PIM is termed protocol-independent because it is not dependent on any particular unicast routing protocol for topology discovery.
- Currently, there are two PIM modes on the live network:
  - **PIM-Dense Mode (PIM-DM)**
  - **PIM-Sparse Mode (PIM-SM)**: According to the multicast service model, PIM-SM can be classified into the following modes:
    - **PIM-SM (ASM)**: sets up an MDT for any-source multicast (ASM).
    - **PIM-SM (SSM)**: sets up an MDT for source-specific multicast (SSM).

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The commonly used PIM version is PIMv2. PIMv2 messages are encapsulated in IP packets, carrying the protocol ID 103 and group address 224.0.0.13.
- In a PIM domain, a P2MP multicast forwarding path is set up from a multicast source to group members for each multicast group. A multicast forwarding path looks like a tree, so it is also called an MDT.
- Characteristics of an MDT:
  - Each link transmits at most one copy of identical data, regardless of how many group members exist on the network. The multicast data is replicated and distributed on a bifurcating node as far from the source as possible.

Usage Scenarios of PIM-DM and PIM-SM

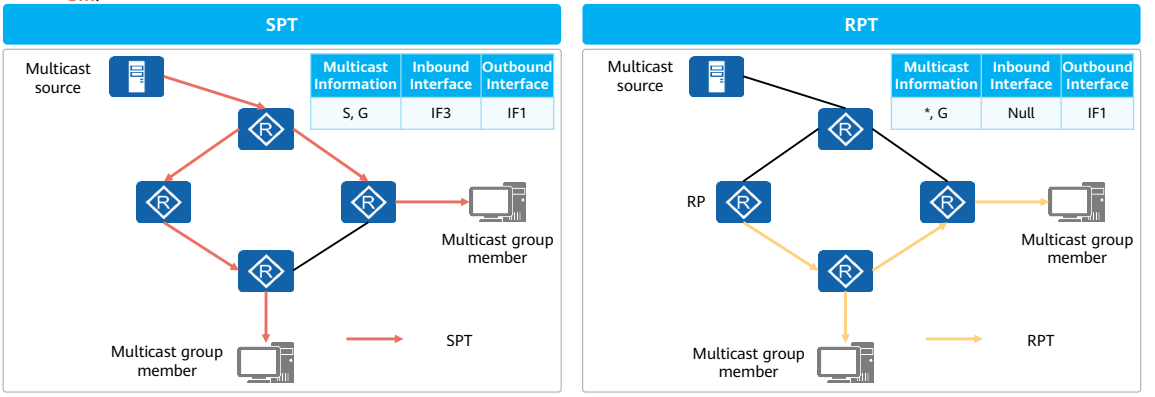
- PIM MDTs can be established in PIM-DM or PIM-SM mode. The two modes are used in different scenarios:
  - The **PIM-DM mode** is mainly used on a multicast network with a **small number of densely distributed group members**. The basic idea of establishing an MDT in this mode is "flooding-prune". That is, multicast traffic is flooded on the entire network, and then the paths without group members are pruned to form an MDT.
  - The **PIM-SM mode** is mainly used on a multicast network with a **large number of sparsely distributed group members**. To establish an MDT in this mode, information about group members needs to be collected first. The PIM-SM mode does not flood multicast packets on the entire network and has little impact on the live network. Therefore, **the PIM-SM mode is usually used** on the live network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Classification of MDTs

MDTs established using PIM are classified into the following types:

- An MDT with the multicast source as the root and group members as leaves is referred to as a shortest path tree (SPT). SPTs are used in both PIM-DM and PIM-SM.
- An MDT with a rendezvous point (RP) as the root and group members as leaves is referred to as an RPT. RPTs are used in PIM-SM.

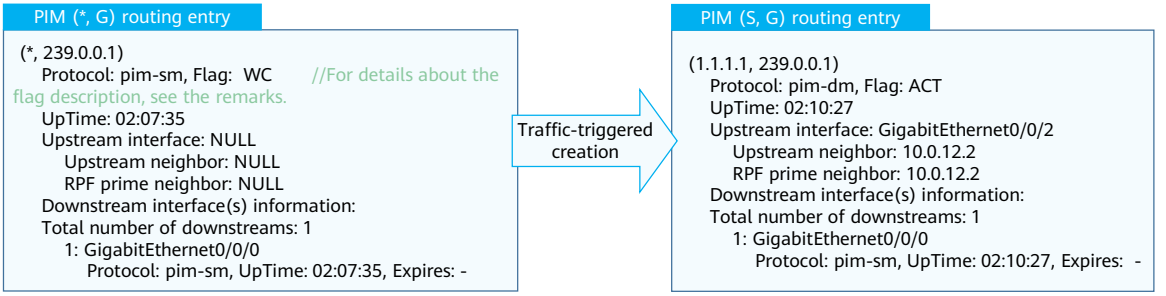


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- SPTs are also called source trees and are used on both PIM-DM and PIM-SM networks.
- RPTs are mainly used on PIM-SM networks.
- For details about RP functions, see the chapter "PIM-SM".

PIM Routing Entries

- PIM routing entries are created using PIM to guide multicast forwarding.
- There are two types of routing entries on a PIM network:
  - (S, G) routing entries are mainly used to establish **SPTs** on a PIM-DM or PIM-SM network.
  - (*, G) routing entries are mainly used to establish **RPTs** on a PIM-SM network.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- S indicates a specific multicast source, G indicates a specific multicast group, and * indicates any multicast source.
- A PIM router may have both (S, G) and (*, G) entries. When the router receives a multicast packet with the source address S and the group address G, the router forwards the packet according to the following rules if the packet passes the RPF check:
  - If a matching (S, G) entry exists, the router forwards the packet according to the (S, G) entry.
  - If no matching (S, G) entry exists but a matching (*, G) entry exists, the router creates an (S, G) entry based on the (*, G) entry, and forwards the packet according to the (S, G) entry.
- For details about the flag description, see the next slide.

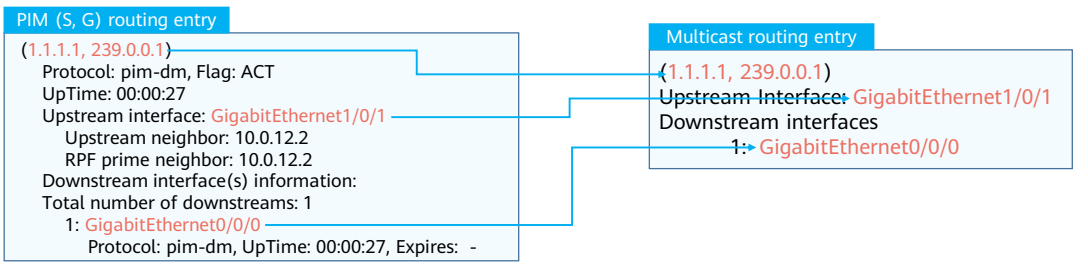
- The Flag field in the PIM routing table is described as follows:

Flag Value	Description
2msdp	The RP recently received a Registration message, learned (S, G) information from the message, and is about to notify MSDP that the next SA message will contain the (S, G) entry.
act	Multicast data corresponding to the multicast routing entry has arrived.
del	Multicast routing entry to be deleted.
exprune	The entry on the RPT is pruned, and no receiver on the RPT requests the information sent by the source.
ext	The routing entry contains an outbound interface provided by another multicast routing protocol.
loc	The routing entry is on the device directly connected to the network segment of the multicast source.
msdp	The routing entry was learned by the RP from a received MSDP SA message.
niif	The routing entry contains an unknown inbound interface.
nonbr	The routing entry has no upstream neighbor address (link-local address) to the RP or source.
none	The routing entry has no flag.
rpt	The routing entry is on the RPT but does not use RPT data.
sg_rcvr	The router has receivers of the source specified in the (S, G) entry, and PIM is the owner of the downstream interface.
sgjoin	The router has receivers of the source specified in the (S, G) entry, but PIM is not the owner of the downstream interface.
spt	The routing entry is on the SPT.
swt	The routing entry is generated during the RPT-to-SPT switchover.
upchg	Route change flag, indicating that the current entry uses the original upstream device to forward data and is waiting for data from a new interface.
wc	(*, G) entry.

PIM Routing Entries and Multicast Routing Entries

- On different multicast routers, multicast routing entries are summarized based on different entries.
  - Multicast routing entries of the last-hop router are mainly generated based on PIM routing entries, IGMP group entries, and IGMP routing entries.

Generation of a multicast routing entry based on a PIM routing entry



- Multicast routing entries can be generated only based on PIM (S, G) routing entries. PIM (*, G) entries do not contain inbound interface information, and as a result, no multicast routing entries can be generated accordingly.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- For details about multicast routing entry generation on the last-hop router, see **IGMP Implementation and Configurations**.

Contents

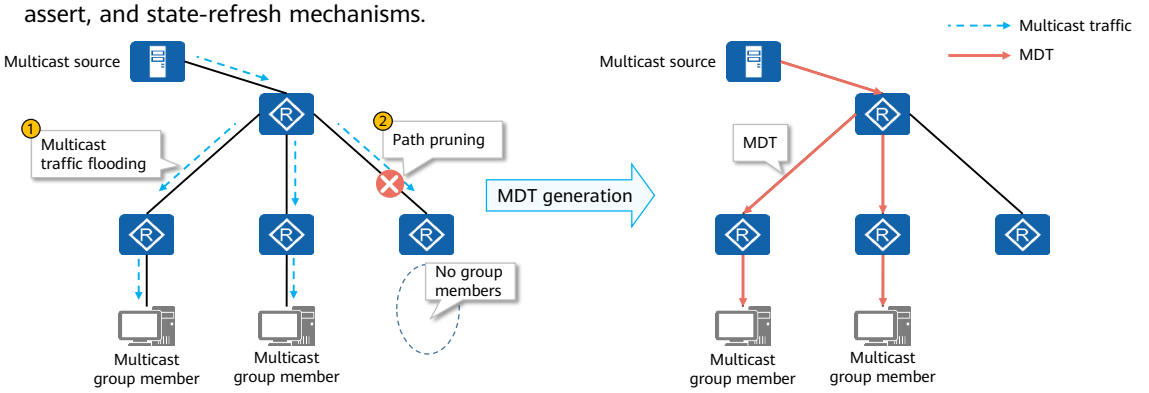
1. Introduction to PIM
2. **Introduction to PIM-DM**
   - Implementation of PIM-DM
   - Basic PIM-DM Configuration
3. Introduction to PM-SM

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Basic Concepts of PIM-DM

- PIM-DM is mainly used on a network with a small number of densely distributed group members. PIM-DM uses the **flooding-prune** mechanism to establish an MDT.
- In addition to the flooding and prune mechanisms, PIM-DM also involves the neighbor discovery, graft, assert, and state-refresh mechanisms.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

PIM-DM Messages

- PIM messages are encapsulated in IP packets, with **destination address 224.0.0.13 and IP protocol number 103**.
- PIM-DM and PIM-SM use different types of messages.
- PIM-DM uses the following types of messages:

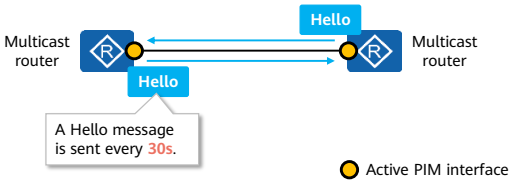
Message Type	Message Function
Hello	Used for PIM neighbor discovery, protocol parameter negotiation, and PIM neighbor relationship maintenance.
Join/Prune	Join messages are used to add nodes to the MDT, whereas Prune messages are used to remove nodes from the MDT. Join and Prune messages are in the same format in PIM, but they have different field contents in the payload.
Graft	Used to graft a node to the MDT.
Graft-ACK	Used to acknowledge the Graft messages sent by neighbors.
Assert	Used in the Assert mechanism.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Neighbor Discovery

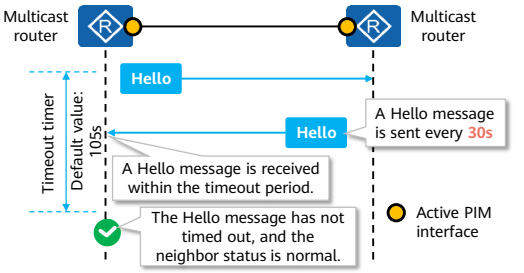
- **An MDT can be established only between PIM neighbors**. Therefore, neighbor discovery is the prerequisite for MDT establishment.
- Neighbor discovery is implemented through PIM Hello messages.

Neighbor discovery and maintenance



- After PIM is enabled on an interface of a router, the interface periodically sends PIM Hello messages with destination address 224.0.0.13. After exchanging Hello messages, multicast routers can learn neighbor information and establish PIM neighbor relationships.

Neighbor relationship maintenance



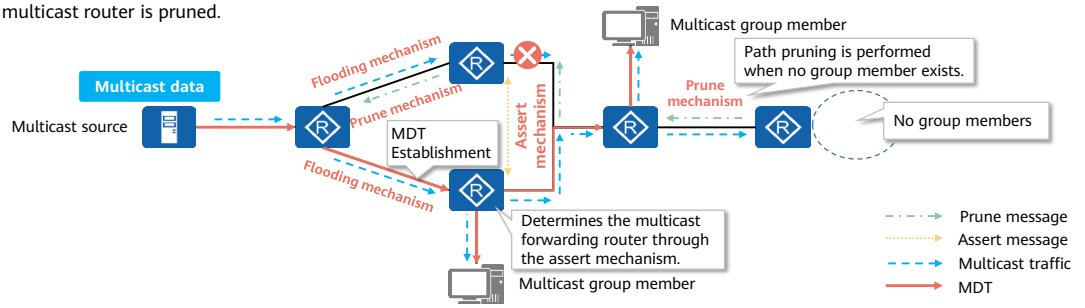
- PIM neighbor relationships are maintained through Hello messages. **The default timeout period of a neighbor relationship is 105s.** If no Hello message is received from a neighbor within the timeout period, the neighbor relationship is deleted.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A Hello message carries the following PIM protocol parameters to control PIM message exchanges between PIM neighbors:
  - DR_Priority: indicates the priority used for DR election among interfaces. The interface with the highest priority becomes the DR.
  - Holdtime: timeout period during which the neighbor remains in the reachable state. If a router does not receive any Hello message from its PIM neighbor within the timeout period, the router considers the neighbor unreachable.
  - LAN_Delay: indicates the delay in transmitting Prune messages on a shared network segment.
  - Neighbor-Tracking: indicates the neighbor tracking function.
  - Override-Interval: indicates the interval for overriding the prune action.

MDT Establishment for the First Time

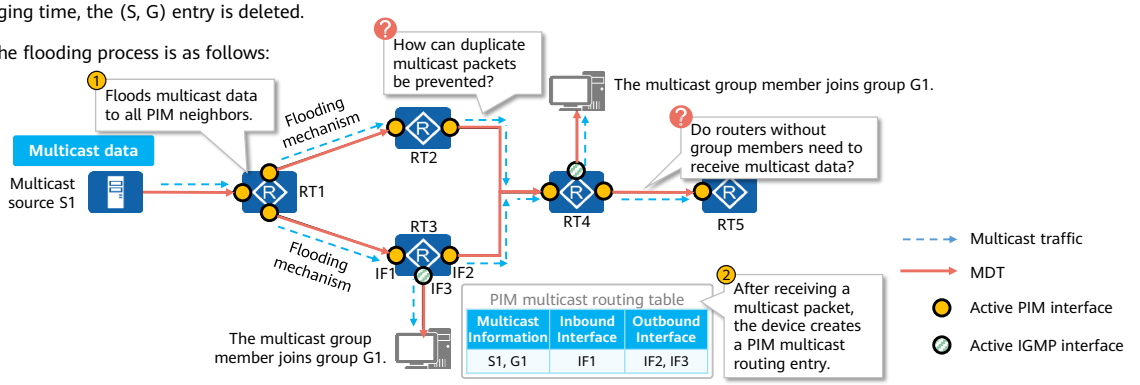
- In PIM-DM mode, an MDT is established for the first time based on the flooding, prune, assert, and DR election mechanisms.
  - Flooding mechanism: Multicast packets are flooded to all PIM neighbors, and multicast routers generate multicast routing entries.
  - Assert mechanism: When a multi-access network exists during multicast traffic forwarding, a multicast forwarding router needs to be elected to prevent duplicate multicast packets.
  - Prune mechanism: If a multicast router has no multicast receivers, the multicast forwarding path from the source to the multicast router is pruned.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Flooding Mechanism

- The multicast packets sent by the multicast source are flooded on the entire network. When a PIM router receives a multicast packet, it performs the **RPF check** on the packet. If the packet passes the RPF check, the router **creates an (S, G) entry** and sends the packet to all PIM neighbors.
- The (S, G) entry generated by PIM-DM has an aging time (210s by default). If no new multicast packet is received throughout the aging time, the (S, G) entry is deleted.
- The flooding process is as follows:

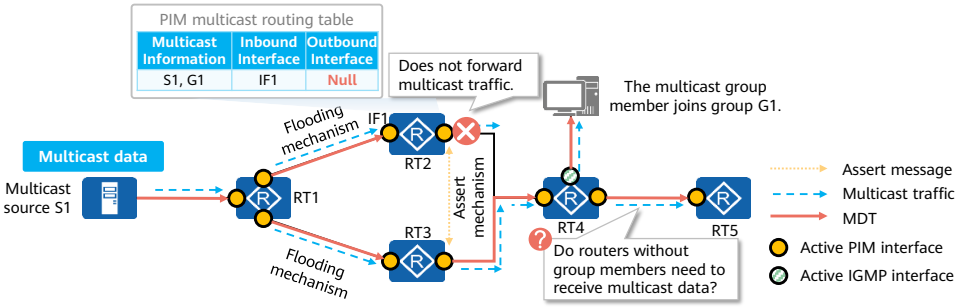


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- According to the flooding mechanism, multicast data is flooded on the entire network periodically (180s by default). The main purpose of periodic flooding is to detect whether new members join a group. However, the flooding of multicast data on the entire network wastes a large amount of bandwidth. Therefore, the state refresh mechanism plus graft mechanism is generally used to listen for new members periodically on the entire network.

Assert Mechanism

- When multiple PIM routers are connected on a network segment, the assert mechanism is used to select **only one of them to forward multicast packets on the network segment**.
- The election rules of the assert mechanism determine the forwarding behavior of each multicast router.
  - If a router wins the assert election, its downstream interface becomes the **assert winner** and is responsible for forwarding multicast packets to the shared network segment.
  - If a router loses the assert election, its downstream interface becomes an **assert loser**, is **deleted from the downstream interface list** of the (S, G) entry, and **no longer forwards** multicast packets to the shared network segment.

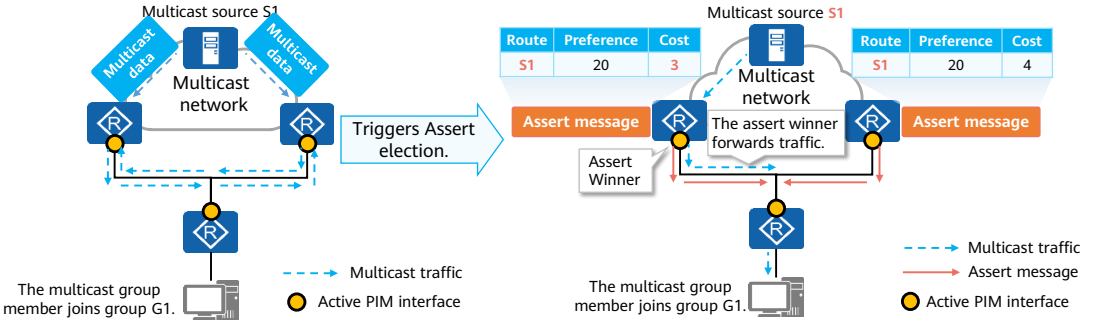


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The assert mechanism is triggered by multicast data.

Election Rules in the Assert Mechanism

- After a PIM router **receives** the **same multicast packet** from its neighbor, it **sends an Assert message** to the network segment for assert election. The Assert message carries the prefix, preference, and cost of the unicast route to the multicast source. Election rules are as follows:
  - If these routers have different unicast route preferences, the router with the highest unicast route preference wins.
  - If these routers have same unicast route preference, the router whose route to the multicast source has the smallest cost wins.
  - If these routers have same unicast route preference and the same route cost to the multicast source, the router whose downstream interface has the highest IP address wins.

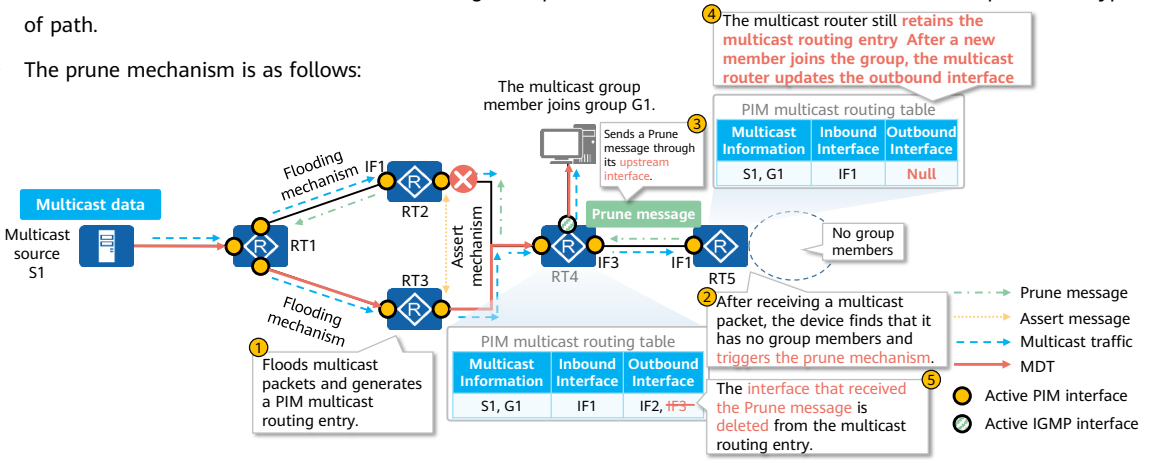


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Assert losers suppress multicast data forwarding and retain the Assert state for a period of time (180s by default).
- After the assert timer expires, the assert losers trigger a new round of election.

Prune Mechanism

- For a multicast router that is not connected to any group member, the multicast network does not need to send multicast traffic to the multicast router. Through the prune mechanism, the multicast network can prune this type of path.
- The prune mechanism is as follows:

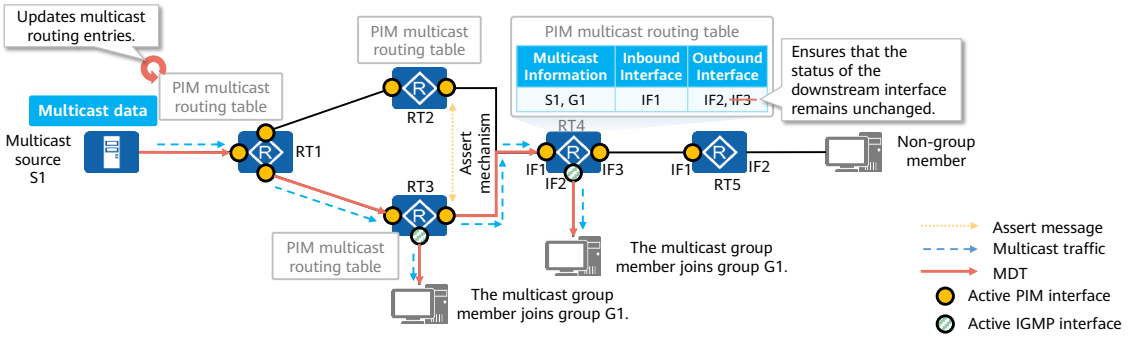


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The pruned downstream interface on a PIM router starts a prune timer (210s by default) and resumes multicast forwarding after the timer expires. Subsequently, multicast packets are flooded throughout the entire network and new group members can receive multicast packets from the interface. If a leaf router connected to a network segment that has no group members receives the flooded multicast packets, the leaf router initiates the prune mechanism. PIM-DM updates the SPT periodically through the process of periodic flooding and prune.
- After a downstream interface of a leaf router is pruned, the leaf router will initiate either the graft or state refresh mechanism:
  - When new members join a multicast group on the network segment connected to the leaf router and want to receive multicast packets before the prune timer expires, the leaf router initiates the graft mechanism.
  - When no member joins a multicast group on the network segment connected to the leaf router and the downstream interface still needs to be suppressed, the leaf router initiates the state refresh mechanism.

MDT Maintenance

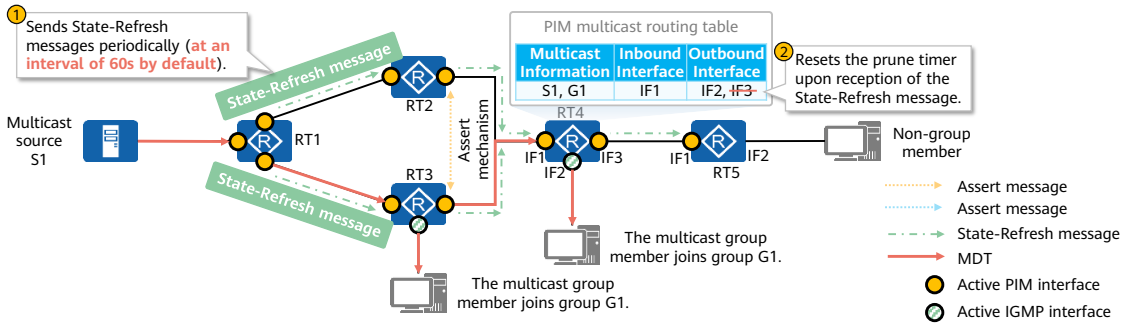
- An MDT may be updated or even removed.
- When PIM neighbor relationships are stable and group memberships remain unchanged, MDTs can be maintained in either of the following ways:
  - Multicast packets are continuously sent to ensure that multicast routing entries always exist.
  - State-Refresh messages are sent to ensure that the downstream interface status in multicast routing entries remains unchanged.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

State-Refresh Mechanism

- To prevent a pruned interface from resuming multicast packet forwarding on a PIM-DM network after the prune timer expires, the first-hop router nearest to the multicast source periodically sends State-Refresh messages throughout the entire PIM-DM network.
- PIM routers receiving the State-Refresh messages reset the prune timer. If the pruned interface on a downstream leaf router keeps having no group members, the interface remains suppressed.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

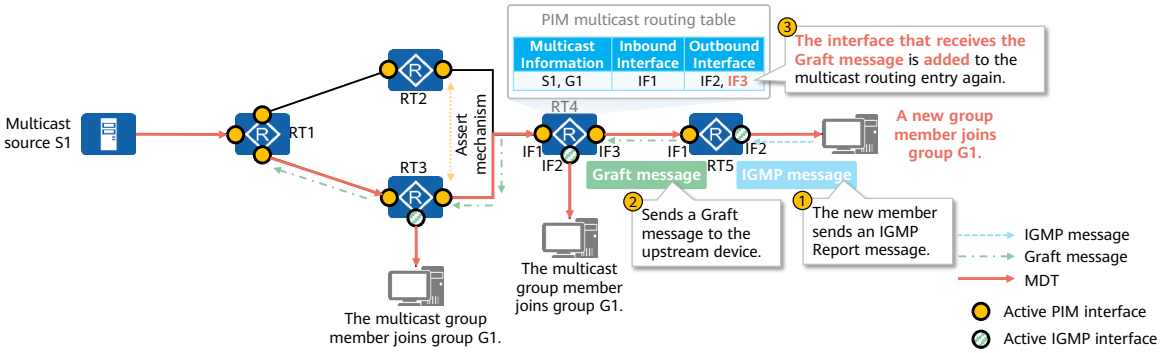
New Members Joining a Group

- After new members join a multicast group, the multicast network needs to **update the MDT** so that the new members can receive multicast data. In PIM-DM mode, after an MDT is established in flooding-prune mode, the **downstream interface cannot be restored automatically once being suppressed** through the **state-refresh mechanism**.
- Therefore, some mechanisms are required to update MDTs. In PIM-DM mode, MDTs can be updated in either of the following methods:
  - Wait for multicast routing entries to time out, and then packets are flooded on the entire network again. This method is uncontrollable and cannot be implemented on the live network.
  - Use the graft mechanism. With the graft mechanism, when a new member joins a group, the leaf PE that receives the Report message of this member **proactively sends a Graft message upstream to establish a reverse MDT**. On the live network, the graft mechanism is often used in this case.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Graft Mechanism

- PIM-DM uses the graft mechanism to enable new group members on a pruned network segment to rapidly obtain multicast data.
- IGMP helps a leaf router learn whether new group members have joined a multicast group on the connected network segment. If a new group member joins the multicast group, the leaf router sends a Graft message to the upstream router **according to the local multicast routing table**, requesting the upstream router to **restore the forwarding capability of the corresponding outbound interface** and add the outbound interface to the downstream interface list of the (S, G) entry.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Contents

1. Introduction to PIM
2. **Introduction to PIM-DM**
   - Implementation of PIM-DM
   - Basic PIM-DM Configuration
3. Introduction to PM-SM

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic PIM-DM Configurations

1. Enable multicast routing on a router.

```
[Huawei] multicast routing-enable
```

2. Enable PIM-DM on an interface.

```
[Huawei - GigabitEthernet1/0/0] pim dm
```

3. Check PIM neighbor parameters.

```
<Huawei>display pim neighbor
```

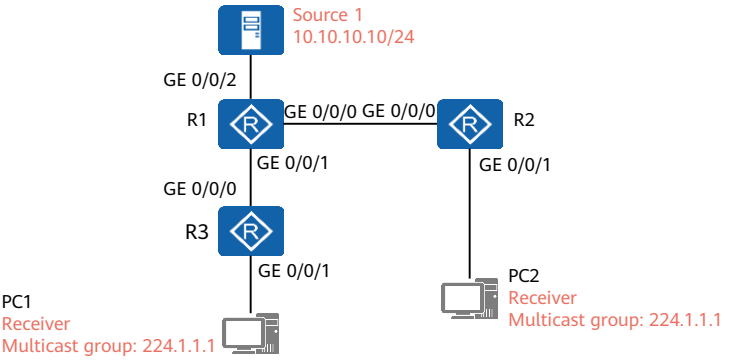
4. Check the PIM routing table.

```
<Huawei>display pim routing-table
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

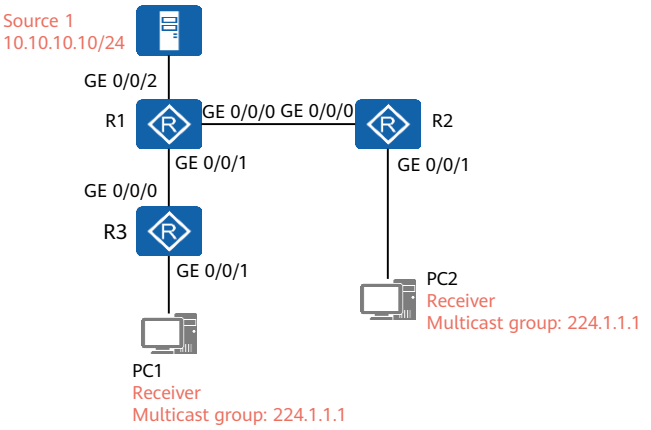
PIM-DM Basics Experiment

- Experiment requirements:
  - Configure PIM-DM to allow PC1 and PC2 to receive data packets from the multicast source.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PIM-DM Configurations (1)



(Configurations of interface IP addresses and OSPF are not provided here.)

Configurations on R1:

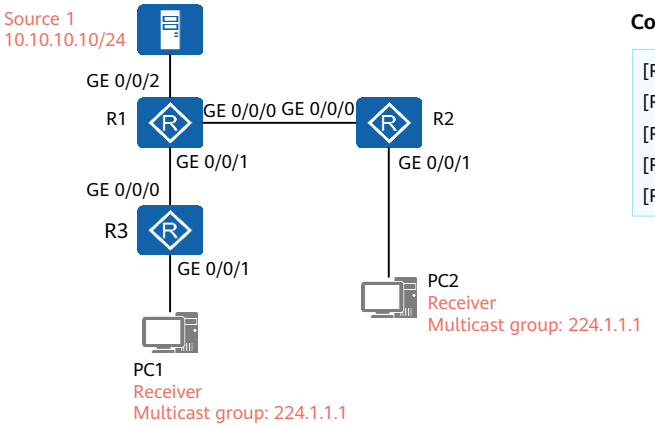
```
[R1]multicast routing-enable
[R1]interface g0/0/2
[R1-GigabitEthernet0/0/2]pim dm
[R1-GigabitEthernet0/0/2]interface g0/0/0
[R1-GigabitEthernet0/0/0]pim dm
[R1-GigabitEthernet0/0/0]interface g0/0/1
[R1-GigabitEthernet0/0/1]pim dm
```

Configurations on R2:

```
[R2]multicast routing-enable
[R2]interface g0/0/0
[R2-GigabitEthernet0/0/0]pim dm
[R2-GigabitEthernet0/0/0]interface g0/0/1
[R2-GigabitEthernet0/0/1]igmp enable
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PIM-DM Configurations (2)



(Configurations of interface IP addresses and OSPF are not provided here.)

Configurations on R3:

```
[R3]multicast routing-enable
[R3]interface g0/0/0
[R3-GigabitEthernet0/0/0]pim dm
[R3-GigabitEthernet0/0/0]interface g0/0/1
[R3-GigabitEthernet0/0/1]igmp enable
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Checking the PIM Routing Table

```
<R1>display pim routing-table
 VPN-Instance: public net
 Total 0 (*, G) entry; 1 (S, G) entry

 (10.10.10.10, 224.1.1.1)
     Protocol: pim-dm, Flag: LOC ACT
     UpTime: 00:00:34
     Upstream interface: GigabitEthernet0/0/2
         Upstream neighbor: NULL
         RPF prime neighbor: NULL
     Downstream interface(s) information:
     Total number of downstreams: 2
         1: GigabitEthernet0/0/0
             Protocol: pim-dm, UpTime: 00:00:34, Expires: never
         2: GigabitEthernet0/0/1
             Protocol: pim-dm, UpTime: 00:00:34, Expires: never
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

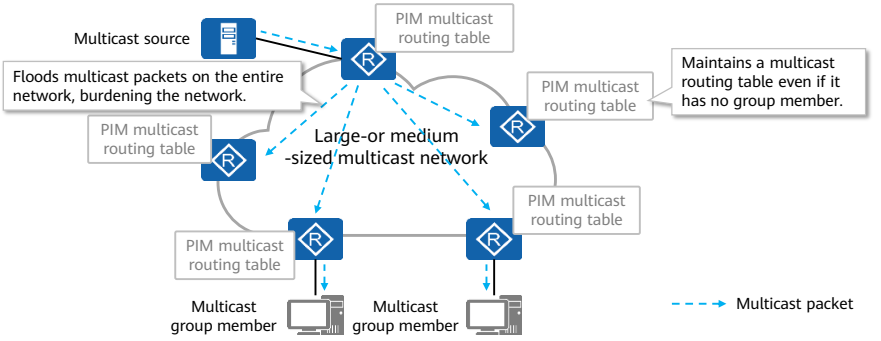
Contents

1. Introduction to PIM
2. Introduction to PIM-DM
3. **Introduction to PM-SM**
   - Implementation of PIM-SM (ASM)
   - Implementation of PIM-SM (SSM)
   - Basic PIM-SM Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PIM-DM Limitations

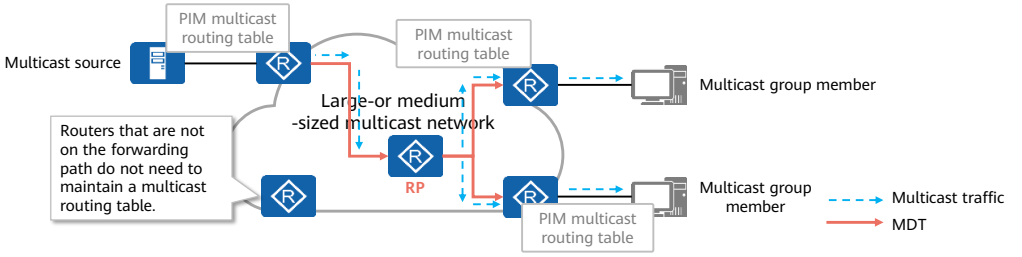
- If PIM-DM is used on a large-or medium-sized multicast network, the following problems may occur:
  - The flooding-prune mode is used, and multicast packets are flooded on the entire network, which **burdens the network**.
  - All multicast routers need to **maintain a multicast routing table** even if they do not need to forward multicast data.
  - On a multicast network with sparsely distributed group members, the **efficiency** of MDT establishment using flooding-prune mechanism is **low**.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Introduction to PIM-SM (ASM)

- The PIM-DM model uses flooding-prune to establish an MDT because most multicast routers on the multicast network do not know the locations of group members.
- In the PIM-SM (ASM) model, an MDT is established as follows:
  - A rendezvous point (RP) is informed of the locations of group members in advance so that an **RPT** is established.
  - After receiving multicast data from the multicast source, the multicast network **sends the multicast data to the RP**, which then forwards the multicast data to group members.
  - If sub-optimal multicast forwarding paths exist, PIM-SM (ASM) **automatically uses the SPT**.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- MDT establishment using the PIM-SM (ASM) model has the following advantages:
  - Only the multicast routers on the multicast forwarding path need to maintain the multicast routing table.
  - The RP enables all multicast routers to know the locations of group members.
  - The flooding-prune mechanism is avoided, which improves MDT establishment efficiency.

PIM-SM (ASM) Messages

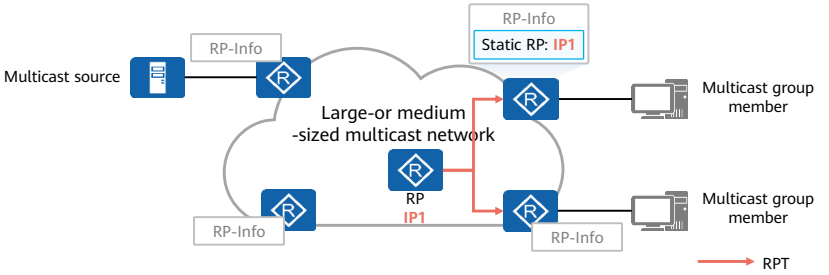
- PIM messages are encapsulated in IP packets, with destination address 224.0.0.13 and IP protocol number 103.
- PIM-SM uses the following types of messages:

Message Type	Message Function
Hello	Used for PIM neighbor discovery, protocol parameter negotiation, and PIM neighbor relationship maintenance.
Register	Unicast message used for source registration. During the source registration process, the first-hop router encapsulates multicast data into a unicast Register message and sends the message to the RP.
Register-Stop	Used by the RP to instruct the first-hop router to stop sending multicast traffic through Register messages.
Join/Prune	Join messages are used to add nodes to the MDT, whereas Prune messages are used to remove nodes from the MDT.
Assert	Used in the Assert mechanism.
Bootstrap	Used for BSR election. In addition, the BSR also uses Bootstrap messages to flood the summary information about candidate-RPs (C-RPs) on the network.
Candidate-RP-Advertisement	Sent by a C-RP to the BSR, carrying information about the C-RP, such as the IP address and priority.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

RP Introduction

- An RP is an important PIM router on the network. It processes the Register messages from the source DR as well as the Report messages from group members. **All PIM routers on the network must know the address of the RP.** An RP is similar to a convergence center of demand and supply information.
- An RP can be configured manually or elected dynamically.
  - **Static RP**: The same RP address is configured for all PIM routers on the network.
  - **Dynamic RP**: An RP is elected among multiple C-RPs through an election mechanism.

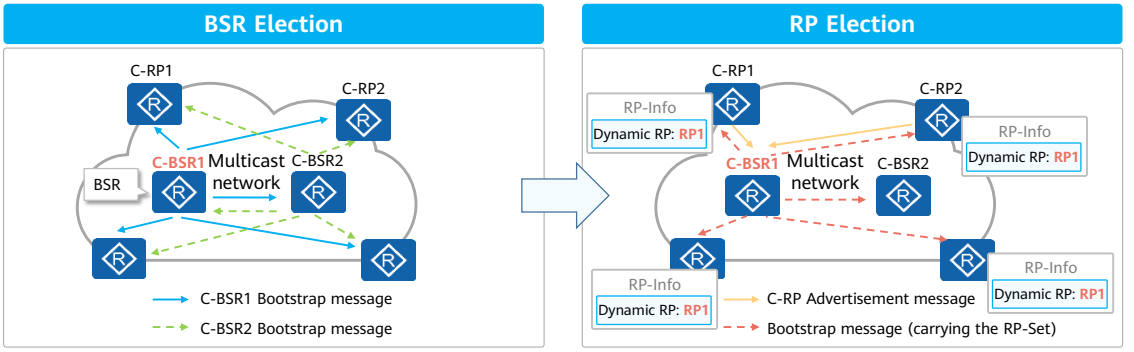


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- You can specify the multicast groups for which a static or dynamic RP provides services.

Dynamic RP Election

- Dynamic RP election involves two roles: candidate-bootstrap router (C-BSR) and C-RP.
  - A unique BSR is elected from C-BSRs.
  - The BSR collects C-RP information to form an RP-Set and floods the RP-Set to all other PIM routers through PIM messages.
  - After receiving the message with the RP-Set, these PIM routers elect an RP according to RP election rules.

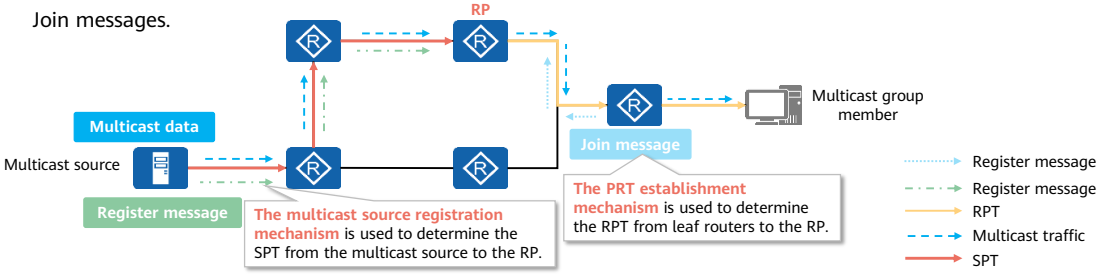


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The BSR election rules are as follows:
  - If the C-BSRs have different priorities, the C-BSR with the highest priority (largest priority value) is elected as the BSR.
  - If the C-BSRs have the same priority, the C-BSR with the highest IP address is elected as the BSR.
- The RP election rules are as follows:
  - The C-RP with the longest mask length of the served group address range matching the specific multicast group wins.
  - If an RP cannot be elected based on the preceding rule, the C-RP with the highest priority (smallest priority value) wins.
  - If an RP cannot be elected based on the preceding rules, the hash function is executed. The C-RP with the largest calculation result wins.
  - If an RP cannot be elected based on the preceding rules, the C-RP with the highest IP address wins.

MDT Establishment for the First Time

- In PIM-SM (ASM) model, MDT establishment for the first time depends on the RPT establishment mechanism, multicast source registration mechanism, and DR election mechanism.
  - RPT establishment mechanism: A multicast leaf router proactively establishes an MDT (RPT) to the RP.
  - Multicast source registration mechanism: used to establish an MDT (SPT) from the multicast source to the RP.
  - DR election mechanism: DRs are responsible for receiving and sending multicast messages from the source or group members to prevent duplicate multicast packets. In addition, the receiver DR is responsible for sending Join messages.

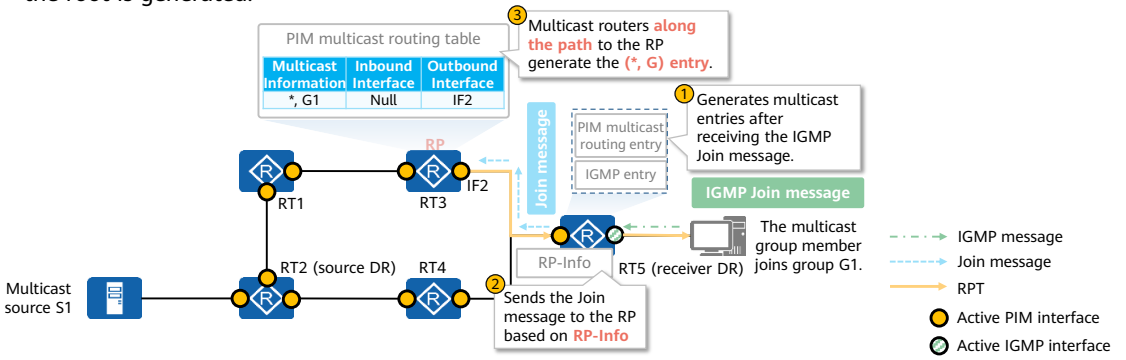


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The DR election mechanism in PIM-DM is similar to that in PIM-DM, and is not mentioned here.

RPT Establishment

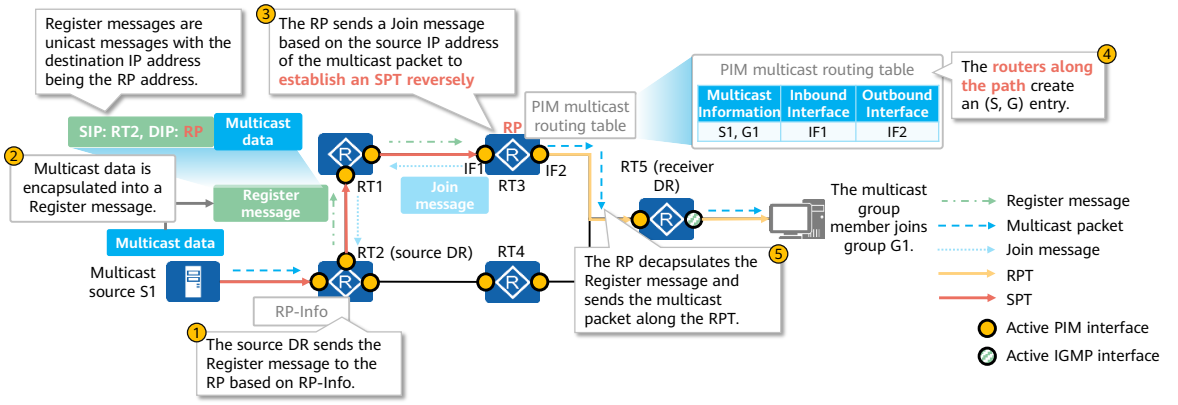
- An RPT is an MDT with the **RP as the root**, and **PIM routers with group members as leaves**.
- When a new group member appears on the network (an IGMP entry is generated), the receiver DR sends a Join message to the RP. The **(*, G) entry** is created hop by hop, and then an RPT with the RP as the root is generated.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Multicast Source Registration Mechanism — SPT Generation

- In the PIM-SM (ASM) model, the MDT from the source DR to the RP cannot be established through Join messages. The **multicast source registration mechanism** is required to help establish the SPT from the source DR to the RP.
- SPT establishment depends on Register and Join messages. The detailed process is as follows:



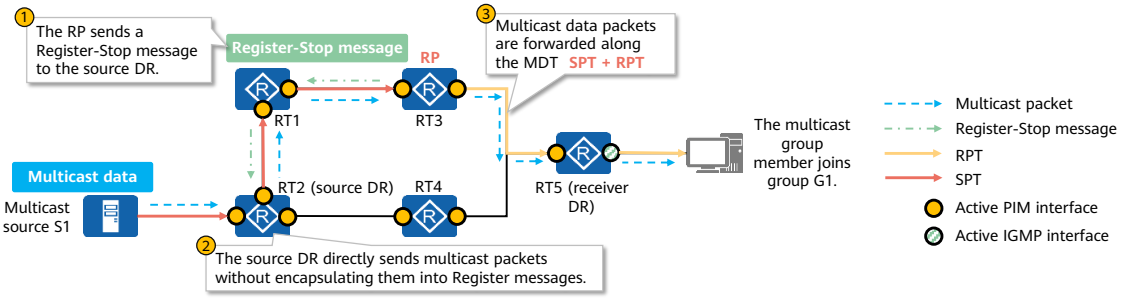
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- IGMP does not run between the multicast source and the source DR. As a result, no PIM (*, G) entries can be generated through IGMP, and Join messages cannot be sent to establish an MDT.

Multicast Source Registration Mechanism — Multicast Data Forwarding

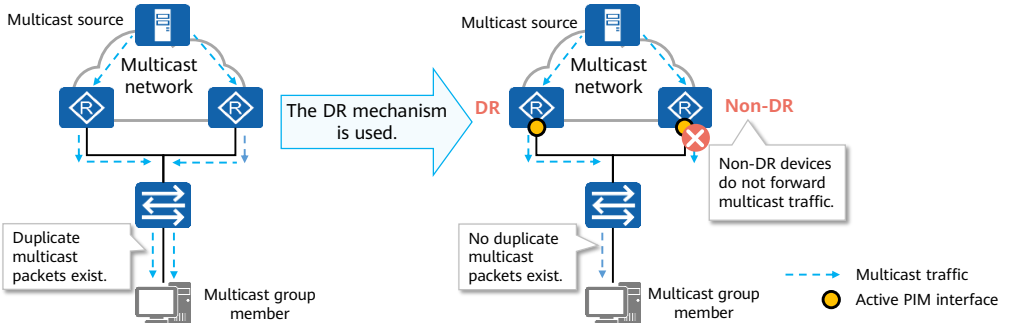
- After multicast source information is registered with the RP, an SPT from the multicast source to the RP is established. However, the source DR still encapsulates multicast data packets into Register messages, which may cause the following problems:
  - The source DR sends unicast Register messages, which **increases the workload of both the source DR and RP**.
  - After an SPT from the source DR to the RP is established, the source DR **sends both unicast Register messages and multicast packets**, causing duplicate multicast packets.
- After the SPT is established, the RP sends Register-Stop messages to instruct the source DR to send subsequent packets in multicast mode.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Duplicate Multicast Packets on the Source End or Receiver End Network

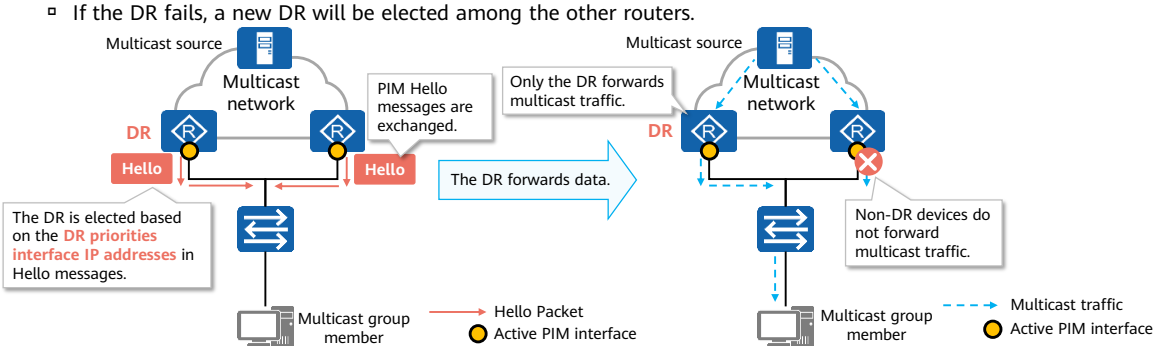
- On the **source end or receiver end network**, multiple multicast routers may forward multicast traffic, causing **duplicate multicast packets**.
- **Only the PIM DR forwards multicast packets** on the source end or receiver end network, which prevents duplicate multicast packets.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PIM DR Election

- PIM DR election:
  - In PIM-SM (ASM), routers compare the **priorities** and **IP addresses** carried in **Hello messages** to elect a DR for a multi-access network.
  - The router with the **highest DR priority** becomes the DR on the multi-access network. If multiple routers have the same DR priority, the router with the **highest interface IP address** becomes the DR.
  - If the DR fails, a new DR will be elected among the other routers.

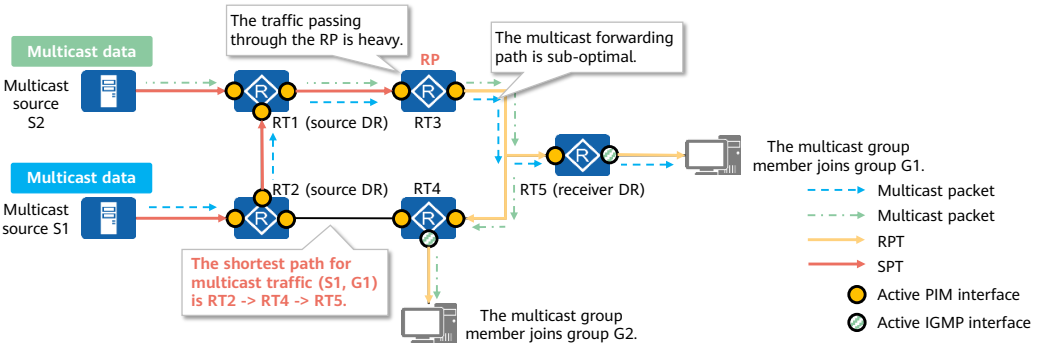


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The default DR priority is 1. A larger value indicates a higher priority.
- If multiple multicast routers are deployed on the receiver end network, both IGMP and PIM must be enabled on the downstream interfaces of the multicast routers.
- The DR can also act as an IGMPv1 querier.

RPT Being a Sub-optimal Path

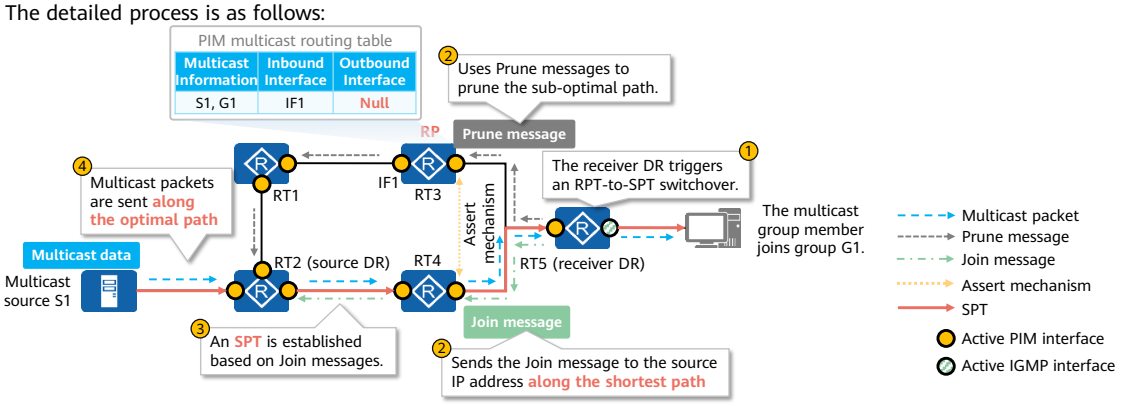
- On a PIM-SM network, a multicast group corresponds to only one RP. Therefore, multicast data is sent to and forwarded by the RP, causing the following problems:
  - A large volume of multicast traffic causes a heavy burden on the RP.
  - The multicast forwarding path may be a sub-optimal one.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

RPT-to-SPT Switchover Mechanism

- After receiving multicast data, the RP sends the data to the receiver DR along the RPT. To prevent the potentially sub-optimal path (RPT) from being used to forward multicast packets, the **receiver DR** establishes an SPT from itself to the source based on the source IP address in the multicast data packets.
- The detailed process is as follows:

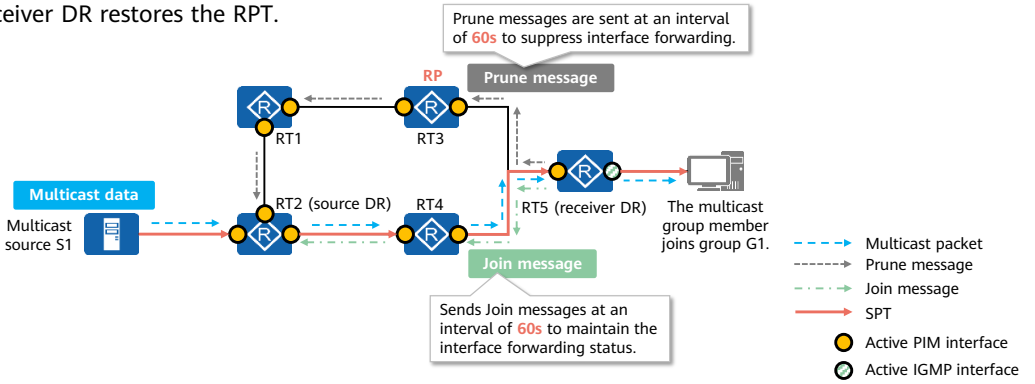


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A device sends Join messages along the shortest path through the upstream interface which is determined based on RPF election rules.
- On a multi-access network, duplicate packets may exist during the RPT-to-SPT switchover. The assert mechanism needs to be used to quickly select a downstream interface.
- Conditions for triggering the RPT-to-SPT switchover: When the forwarding rate exceeds the switchover threshold, the receiver's DR sends a Join messages to the source, triggering the SPT switchover.

MDT Maintenance

- When an MDT (SPT or RPT) becomes stable, the receiver DR periodically sends Join/Prune messages to maintain the MDT.
- If there is no multicast traffic for a period of time (210s by default), the SPT disappears, and the receiver DR restores the RPT.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

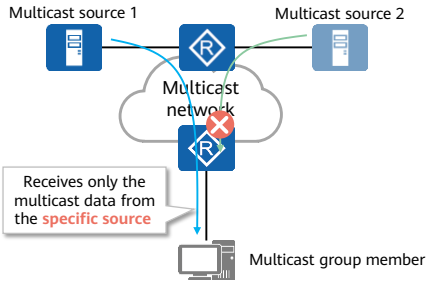
Contents

1. Introduction to PIM
2. Introduction to PIM-DM
3. **Introduction to PM-SM**
   - Implementation of PIM-SM (ASM)
   - Implementation of PIM-SM (SSM)
   - Basic PIM-SM Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SSM Concept Review

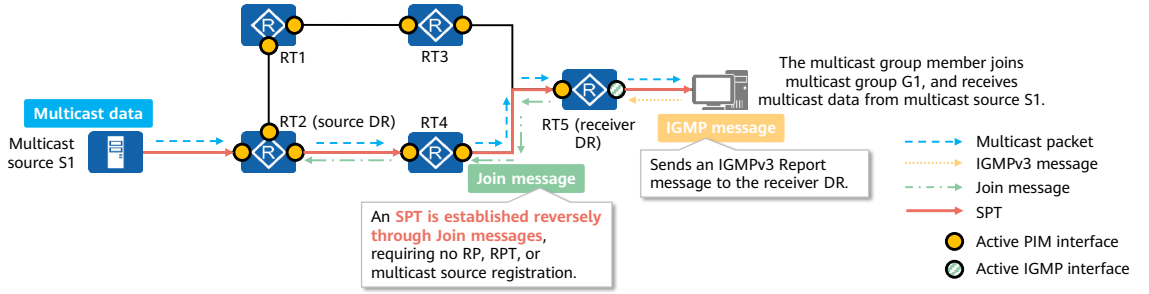
- The SSM model provides services for the data flows from a specific source to a specific group. When a receiver host joins a multicast group, the receiver host can specify the **sources from which it wants to receive multicast data or the sources from which it rejects multicast data**. After joining a group, the hosts receive **only the data sent from the specified sources** to the group.
- **The SSM model does not require globally unique group addresses, but the multicast source must be unique.** Different applications on a source must use different SSM group addresses. Different applications on different sources can reuse SSM group addresses because each source-group pair has an (S, G) entry. This model saves multicast group addresses without congesting the network.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PIM-SM (SSM) Overview

- Because the SSM model predefines a multicast source address, PIM-SM (SSM) allows the receiver DR to establish a reverse SPT based on the multicast source address.
- PIM-SM (SSM) does not require an RP, RPT, or multicast source registration. Instead, PIM-SM (SSM) can directly establish an SPT between a multicast source and group members.
- The key mechanisms of PIM-SM (SSM) include neighbor discovery, DR election, and SPT establishment.

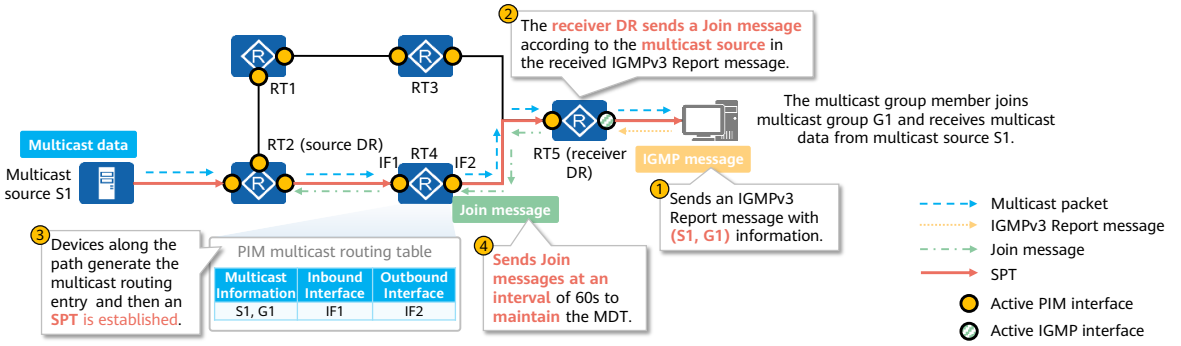


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- PIM-SM (SSM) does not require the Assert mechanism.

MDT Generation and Maintenance

- The PIM-SM (SSM) model uses IGMPv3 messages and Join messages to establish an MDT.
- The **MDT** established in the PIM-SM (SSM) model **always exists** and will not disappear even if no multicast traffic is transmitted.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The DR and neighbor discovery mechanisms in PIM-SM (SSM) are the same as those in PIM-DM, and are not mentioned here.

PIM Model Comparison

Protocol	Model	Usage Scenario	Implementation
PIM-DM	ASM model	Small-scale LANs where multicast group members are distributed densely	Using the flooding-prune mechanism, PIM-DM creates and maintains a unidirectional and loop-free SPT connecting a multicast source and group members.
PIM-SM	ASM model	Large-scale network where multicast group members are distributed sparsely	An MDT is established through proactive notification of receivers' group joining information. PIM-SM requires RP maintenance, RPT establishment, and multicast source registration.
	SSM model	Scenarios where user hosts know the exact locations of multicast sources in advance and can specify the sources from which they want to receive data before they join multicast groups	In PIM-SSM, an SPT is directly established between the multicast source and group members. PIM-SSM does not require RP maintenance, RPT establishment, or multicast source registration.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Introduction to PIM
2. Introduction to PIM-DM
3. **Introduction to PM-SM**
   - Implementation of PIM-SM (ASM)
   - Implementation of PIM-SM (SSM)
   - Basic PIM-SM Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic PIM-SM Configurations

1. Enable multicast routing on a router.

   [Huawei] **multicast routing-enable**

2. Enable PIM-SM on an interface.

   [Huawei - GigabitEthernet1/0/0] **pim sm**

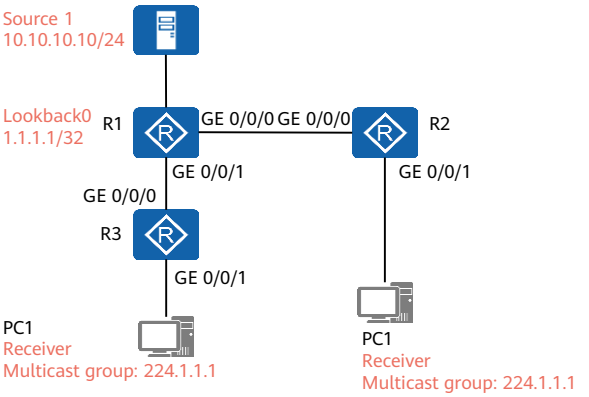
3. Configure a static RP.

   [Huawei -pim] **static-rp** *rp-address*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

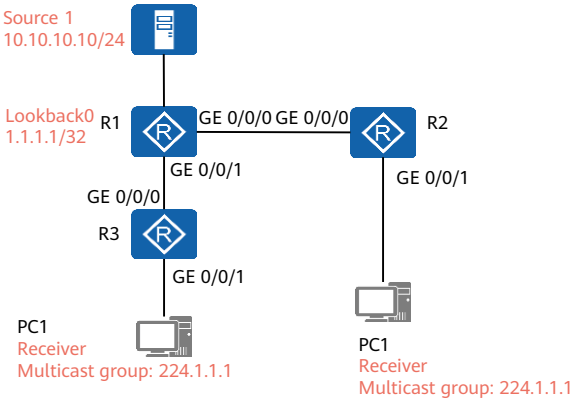
PIM-SM Basics Experiment

- Experiment requirements:
- Configure PIM-SM to allow PC1 and PC2 to receive data packets from the multicast source.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PIM-SM Configurations (1)



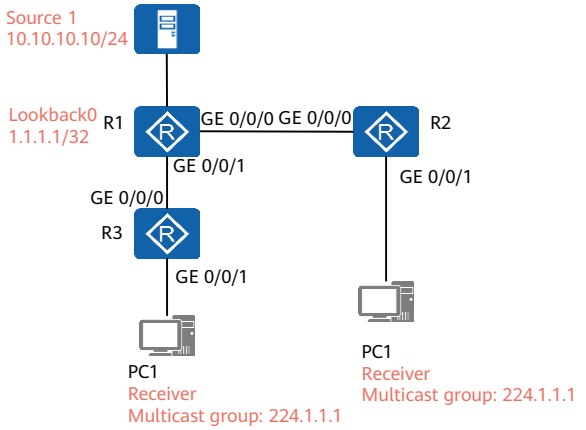
(Configurations of interface IP addresses and OSPF are not provided here.)

Configurations on R1:

[R1]**multicast routing-enable**
[R1]**interface g0/0/2**
[R1-GigabitEthernet0/0/2]**pim sm**
[R1-GigabitEthernet0/0/2]**interface g0/0/0**
[R1-GigabitEthernet0/0/0]**pim sm**
[R1-GigabitEthernet0/0/0]**interface g0/0/1**
[R1-GigabitEthernet0/0/1]**pim sm**
[R1] **pim**
[R1-pim] **static-rp 1.1.1.1**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

PIM-SM Configurations (2)



(Configurations of interface IP addresses and OSPF are not provided here.)

Configurations on R2:

```
[R2]multicast routing-enable
[R2]interface g0/0/0
[R2-GigabitEthernet0/0/0]pim sm
[R2-GigabitEthernet0/0/0]interface g0/0/1
[R2-GigabitEthernet0/0/1]igmp enable
[R2] pim
[R2-pim] static-rp 1.1.1.1
```

Configurations on R3:

```
[R3]multicast routing-enable
[R3]interface g0/0/0
[R3-GigabitEthernet0/0/0]pim sm
[R3-GigabitEthernet0/0/0]interface g0/0/1
[R3-GigabitEthernet0/0/1]igmp enable
[R3] pim
[R3-pim] static-rp 1.1.1.1
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Single) What is the destination IP address of PIM messages?
   A. 224.0.0.2
   B. 224.0.0.1
   C. 224.0.0.5
   D. 224.0.0.13
2. (Multiple) Which of the following mechanisms are used by multicast to prevent duplicate packets?
   A. RPF mechanism
   B. Assert election mechanism
   C. DR election mechanism
   D. IGMP querier election mechanism

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. D
2. A, B, C

Summary

- There are two PIM models:
  - PIM-DM is mainly used on small-scale multicast networks with densely distributed users.
  - PIM-SM is mainly used on large-scale multicast networks with sparsely distributed users. Based on the multicast model, PIM-SM is classified into PIM-SM (ASM) and PIM-SM (SSM). PIM-SM (SSM) mainly serves SSM.
- PIM-DM uses the flooding-prune mechanism to establish an MDT. During MDT generation, PIM-DM uses the Assert and DR election mechanisms to prevent loops. During multicast forwarding, PIM-DM uses the PRF mechanism to prevent loops.
- PIM-SM (ASM) sends group membership information to the RP to establish an RPT. The multicast source sends multicast packets to the RP, which then sends multicast data to group members. In this manner, an MDT (SPT + RPT) is established. To prevent the RPT which is a sub-optimal path from being used to forward multicast traffic, PIM-SM (ASM) initiates an RPT-to-SPT switchover to optimize the MDT.
- PIM-SM (SSM) mainly serves the SSM model. Because this model knows the multicast source address in advance, an MDT can be directly established reversely.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Introduction to IPv6

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- IPv4 is a widely used Internet protocol. In the early stage of the Internet, IPv4 developed rapidly due to its simple design, easy implementation, and good interoperability. However, with the rapid development of the Internet, IPv4's design defects become increasingly obvious. Internet Protocol version 6 (IPv6) was developed to address the defects.
- IPv6 is also called IP next generation (IPng). Designed by the Internet Engineering Task Force (IETF), IPv6 is an upgraded version of IPv4.
- This course describes the basic concepts, address classification, and packet format of IPv6.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- On completion of this course, you will be able to:
  - Describe the development status of IPv6.
  - Explicate the advantages of IPv6 over IPv4.
  - Describe the basic concepts of IPv6.
  - Describe the format of an IPv6 packet header.
  - Describe the IPv6 address format and types.
  - Configure IPv6 addresses and routes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

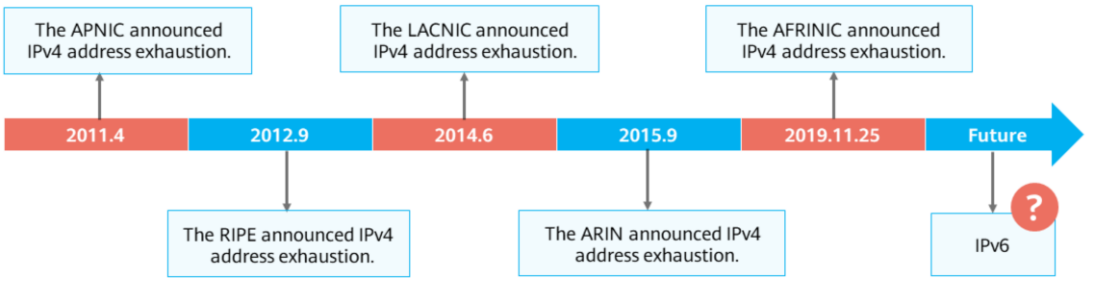
Contents

1. **IPv6 Overview**
2. IPv6 Address Introduction
3. IPv6 Packet Structure
4. Basic IPv6 Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv4 Status

On February 3, 2011, the Internet Assigned Numbers Authority (IANA) announced even allocation of its last 4.68 million IPv4 addresses to five Regional Internet Registries (RIRs) around the world. The IANA thereafter had no available IPv4 address.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- The IANA is responsible for assigning global Internet IP addresses. The IANA assigns some IPv4 addresses to continent-level RIRs, and then each RIR assigns addresses in its regions. The five RIRs are as follows:
  - RIPE: Réseaux IP Européens, which is a European IP address registration center and serves Europe, Middle East, and Central Asia.
  - LACNIC: Latin American and Caribbean Internet Address Registry, which is an Internet address registration center for Latin America and the Caribbean and serves the Central America, South America, and the Caribbean.
  - ARIN: American Registry for Internet Numbers, which is an Internet number registration center in the United States and serves North America and some Caribbean regions.
  - AFRINIC: Africa Network Information Centre, which serves Africa.
  - APNIC: Asia Pacific Network Information Centre, which serves Asia and the Pacific.
- IPv4 has proven to be a very successful protocol. It has survived the development of the Internet from a small number of computers to hundreds of millions of computers. However, this protocol was designed based on then network scale several decades ago. With the expansion of the Internet and the launch of new applications, IPv4 has shown more and more limitations.
- The rapid expansion of the Internet scale was unforeseen at that time. Especially over the past decade, the Internet has experienced explosive growth and has been accessed by numerous households. It has become a necessity in people's daily life. In this case, IPv4 address exhaustion is becoming an urgent issue.
- In the 1990s, the IETF launched technologies such as network address translation (NAT) and classless inter-domain routing (CIDR) to delay IPv4 address exhaustion. However, these transition solutions can only slow down the speed of address exhaustion, but cannot fundamentally solve the issue.

Development Status of IPv6

Significant Increase in the Global IPv6 Deployment Rate

By October 2019, countries or regions with a comprehensive IPv6 deployment rate of over 30% in the global map area accounted for
> 50%

Sharp Increase in the Number of Global IPv6 Users

According to Google website monitoring, by September 2019, users who use IPv6 to access Google websites accounted for
> 30%

Improved IPv6 Support for Mainstream Software and Services

- By October 10, 2019, 14.6% of global websites, 19.2% of the top 1 million websites, and nearly 30% of the top 1000 websites support IPv6 access.
- Major cloud service providers and CDN carriers, such as Cloudflare, Akamai, Microsoft Azure, and Amazon, all support IPv6.
- Windows and MacOS operating systems for mass users support IPv6.

IPv6 Deployment in Global Networks and Domain Name Systems

By October 2019, 1505 of the 1527 top domains worldwide support IPv6,
accounting for 98.6% of the total.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

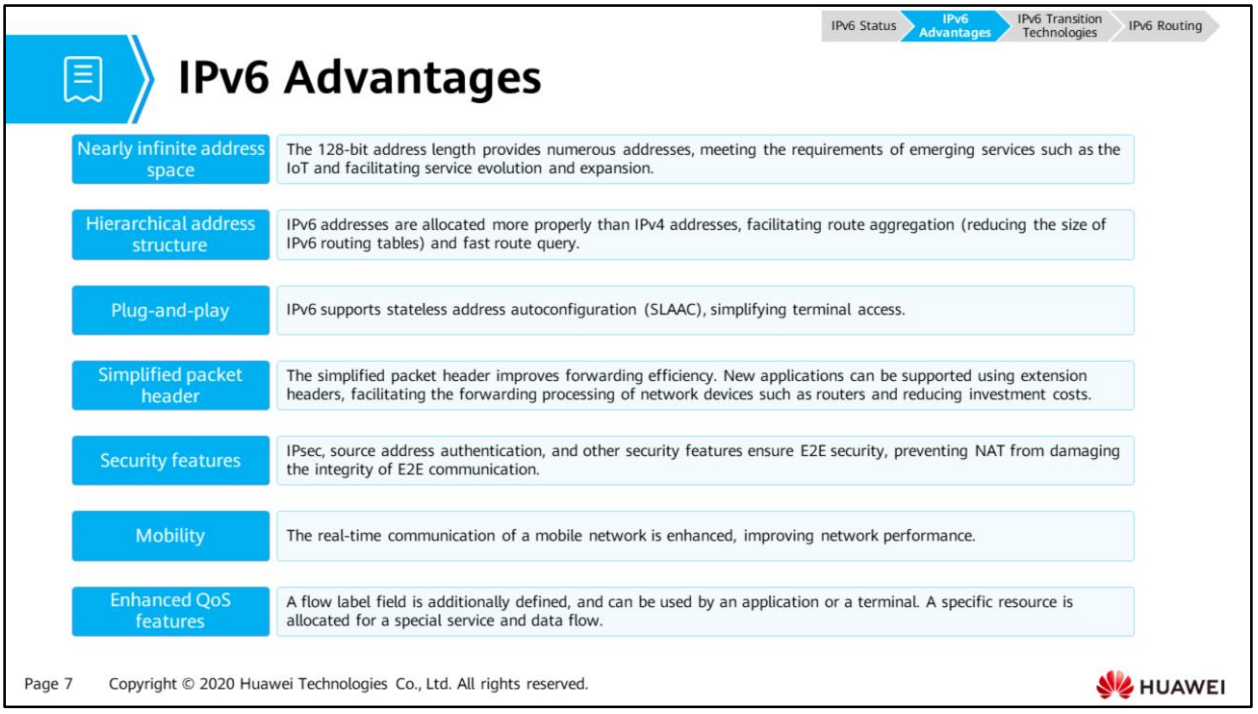
Why IPv6?



IPv4		IPv6
Public address exhaustion		Nearly infinite address space
Improper packet header design		Hierarchical address allocation
Oversized routing table and low table lookup efficiency	VS.	Plug-and-play
Broadcast flooding caused by dependency on ARP		Simplified packet header
...		IPv6 security features
		Integrity of E2E communication
		Mobility
		Enhanced QoS features
		...



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Nearly infinite address space: This is the most obvious advantage over IPv4. An IPv6 address consists of 128 bits. The address space of IPv6 is about 8 x 10^{28} times that of IPv4. It is claimed that IPv6 can allocate a network address to each grain of sand in the world. This makes it possible for a large number of terminals to be online at the same time and unified addressing management, providing strong support for the Internet of Things (IoT).
- Hierarchical address structure: IPv6 addresses are divided into different address segments based on application scenarios thanks to the nearly infinite address space. In addition, the continuity of unicast IPv6 address segments is strictly required, facilitating IPv6 route aggregation and reducing the size of IPv6 address tables.
- Plug-and-play: Any host or terminal must have a specific IP address to obtain network resources and transmit data. Traditionally, IP addresses are assigned manually or automatically using DHCP. In addition to the preceding two methods, IPv6 supports SLAAC.
- E2E network integrity: NAT used widely on IPv4 networks damages the integrity of E2E connections. After IPv6 is used, NAT devices are no longer required, and online behavior management and network monitoring become simple. In addition, applications do not need complex NAT adaptation code.
- Enhanced security: IPsec was initially designed for IPv6. Therefore, IPv6-based protocol packets (such as routing protocol and neighbor discovery packets) can be encrypted in E2E mode, despite the fact that this function is not widely used currently. The security capability of IPv6 data plane packets is similar to that of IPv4+IPsec.

Introduction to IPv6 Transition Technologies (1)

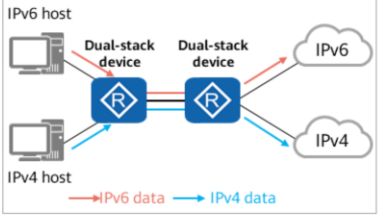
- NAT alleviates IPv4 address insufficiency, and IPv6 is the final solution to this issue. Currently, different regions in the world have different requirements for IPv6 deployment, and IPv4 networks are still dominant. Therefore, IPv6 and IPv4 will coexist in a short period of time.
- The evolution from an IPv4 network to an IPv6 network involves the following technologies:
  - Dual-stack technology: enables both IPv4 and IPv6 stacks on a device.
  - Tunneling technology: encapsulates a type of protocol data into another type of one.
  - Translation technology: translates between IPv6 and IPv4 addresses.
- There is no optimal transition technology solution, and no technical solution can resolve all issues. Generally, multiple technologies are combined into different transition solutions to address different network access scenarios.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Introduction to IPv6 Transition Technologies (2)

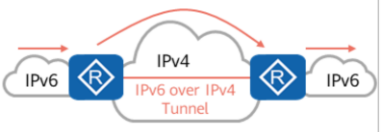
IPv4/IPv6 Dual Stack

- Devices support IPv4/IPv6. IPv4 and IPv6 are independently deployed on a network and coexist within a period of time. This technology has little impacts on existing IPv4 services.
- The evolution solution is simple and easy to understand. The workload of network planning and design is relatively low.
- Much hardware and software (such as network devices, terminals, and operating systems) on live networks support dual stack. Even IPv6 is enabled by default and can be directly used.
- The hardware and software of devices must support dual stack.



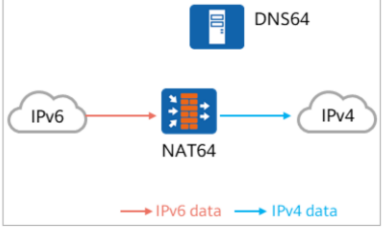
Tunneling Technology

- IPv4 traffic is encapsulated in an IPv6 tunnel, or IPv6 traffic is encapsulated in an IPv4 tunnel.
- This technology implements interconnection between isolated IPv6 islands on an IPv4 network or interconnection between isolated IPv4 islands on an IPv6 network.
- Devices (usually tunnel endpoints) must support dual stack and tunneling technology.
- This technology is applicable to a temporary isolated island interconnection but not long-term and stable services.

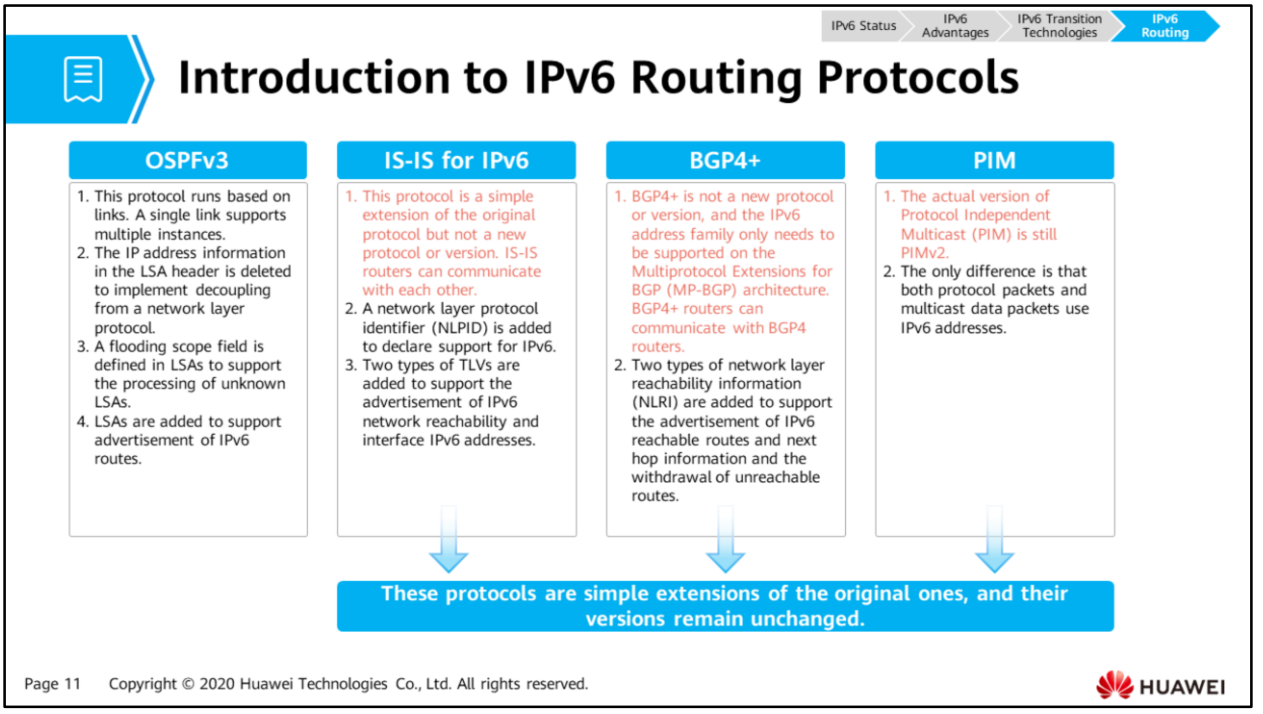


Translation Technology

- IPv4 traffic is translated into IPv6 traffic (mainly IP header modification), or IPv6 traffic is translated into IPv4 traffic.
- This technology implements communication between native IPv4 and IPv6 networks.
- This implementation damages the integrity of E2E connections. The ALG function must be provided for special applications.
- Network management and audit become complex.
- NAT and DNS devices must be deployed on a network.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Similar to IPv4 networks, IPv6 networks also support static routes.

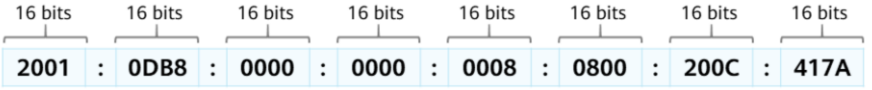
Contents

1. IPv6 Overview
2. **IPv6 Address Introduction**
   - **IPv6 Address Overview**
   - IPv6 Address Types
   - IPv6 Address Planning Example
3. IPv6 Packet Structure
4. Basic IPv6 Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Address

- The length of an IPv6 address is 128 bits. Generally, an IPv6 address is divided into eight segments using colons, and each segment contains 16 bits and is expressed in hexadecimal notation.



The letters in an IPv6 address are case insensitive. For example, **A** is equivalent to **a**.

- Similar to an IPv4 address, an IPv6 address is expressed in the format of *IPv6 address*/*mask length*.

IPv6 address: 2001:0DB8:2345:CD30:1230:4567:89AB:CDEF/64

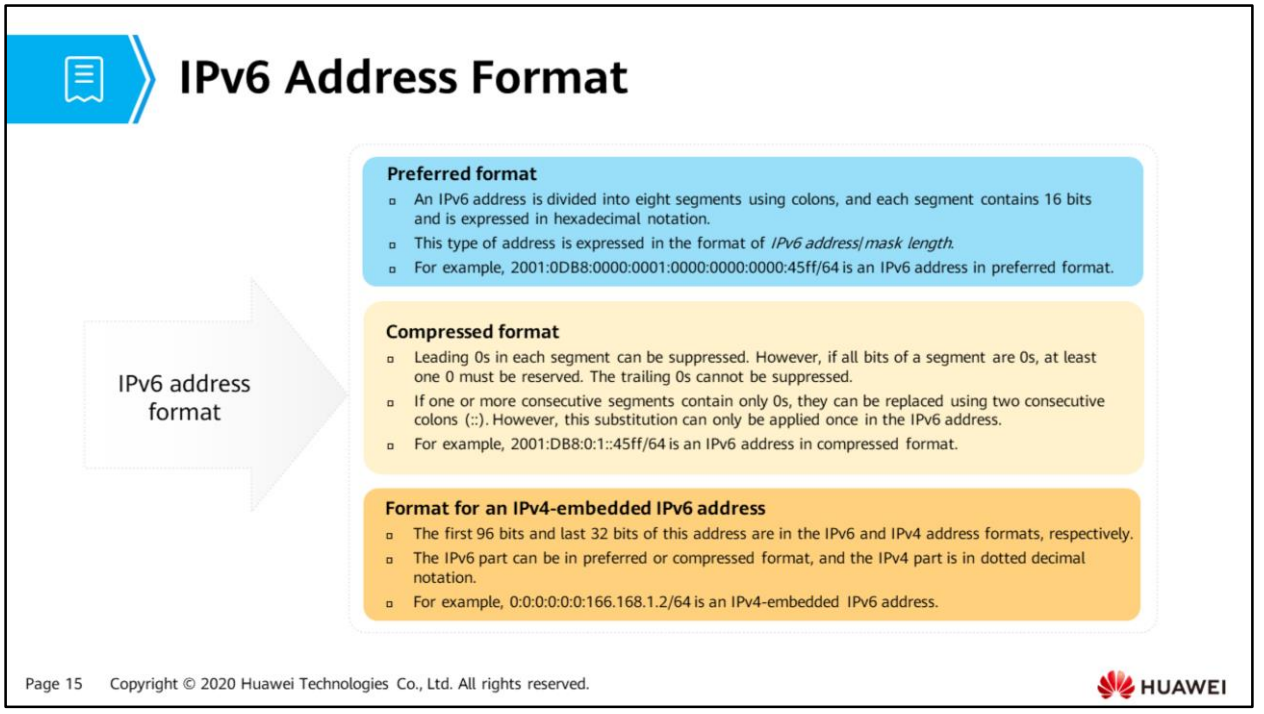
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Address Space

- Why is the length of an IPv6 address 128 bits?
  - The word length processed by the CPU has experienced 4 bits, 8 bits, 16 bits, 32 bits, and 64 bits. When data can be represented by a binary number with an exponent power of 2, the CPU has the highest processing efficiency.
  - Considering data processing efficiency and network expansion in the future, it is very appropriate to set the length of an IPv6 address to 128 bits.

- What is the concept of a 128-bit IPv6 address?
  - There are 2^{32} (4,294,967,296) IPv4 addresses.
  - There are 2^{128} (2^{96} x 2^{32} = 340,282,366,920,938,463,463,374,607,431,768,211,456) IPv6 addresses. Specifically, 67 trillion addresses can be allocated per square meter on Earth's surface
  - It is no exaggeration to say that each grain of sand on Earth can be allocated an IPv6 address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- An IPv4 address consists of four decimal numbers separated by dots and a mask, for example, 192.168.1.1/24. The length of an IPv6 address is 128 bits, and it is suitable for an IPv6 address to inherit the decimal format of an IPv4 address. The IPv6 address format different from the IPv4 address format is defined in RFC 2373.

IPv6 Address Structure

- An IPv6 address is composed of two parts:
  - Network prefix: consists of *n* bits and has the same function as the network ID of an IPv4 address.
  - Interface ID: consists of (128 – *n*) bits and is parallel to the host ID of an IPv4 address.
- Example of an IPv6 unicast address: 2001:0DB8:6101:0001:5ED9:98FF:FECA:A298/64



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Address Prefix

- Due to the limitations of IPv4 address planning and allocation, the IETF classifies IPv6 addresses into different types. Different types of IPv6 addresses are assigned different prefixes and are strictly managed by address allocation organizations.
- Currently, common IPv6 addresses or prefixes are as follows:

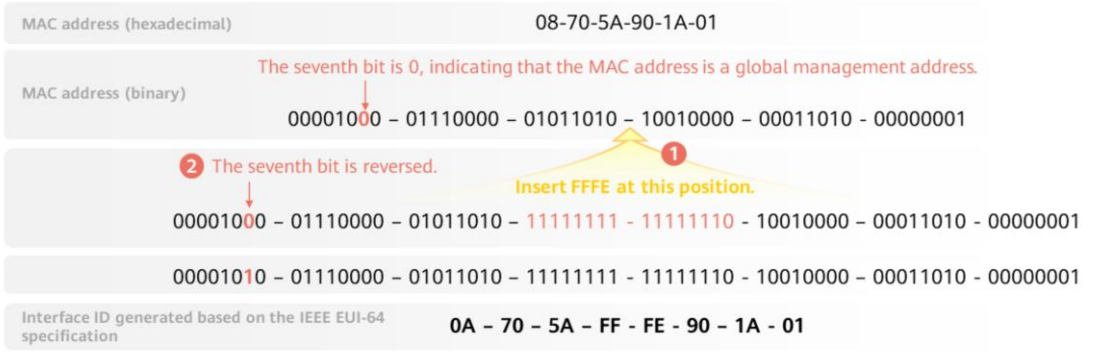
IPv6 Address or Prefix	Description
2001::/16	Used for the IPv6 Internet, similar to a public IPv4 address.
2002::/16	Used for 6to4 tunnels.
FE80::/10	Link-local address prefix, which is used for communication within the local link range.
FF00::/8	Multicast address prefix, which is used for IPv6 multicast.
::/128	Unspecified address, which is similar to 0.0.0.0 in IPv4.
::1/128	Loopback address, which is similar to 127.0.0.1 in IPv4.
...	...

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Latest definition of the IANA for IPv6 prefixes

IPv6 Address's Interface ID

- An interface ID can be manually configured, automatically generated by the system, or generated based on the IEEE EUI-64 specification.
- The method of generating an interface ID based on the IEEE EUI-64 specification is most commonly used. In this method, the MAC address of an interface is converted into an IPv6 address's interface ID.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

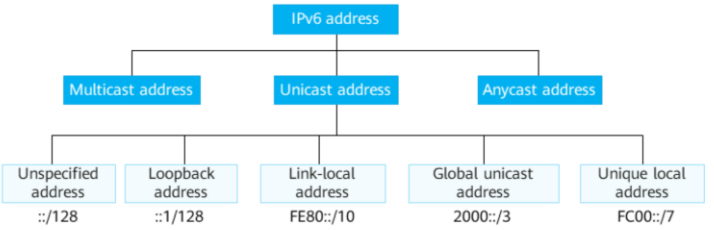
- Currently, the interface ID of an IPv6 address can be generated in the following ways:
  - Generated based on the IEEE EUI-64 specification
    - The typical length of an interface ID is 64 bits. The IEEE EUI-64 specification defines a method of generating an interface ID, that is, transforming a 48-bit MAC address to a 64-bit interface ID.
    - A 48-bit MAC address can be transformed to a 64-bit interface ID by changing the seventh bit 0 to 1 and inserting FFFE in the middle of a MAC address.
    - This method reduces the configuration workload. Only one IPv6 prefix needs to be obtained to form an IPv6 address with the interface ID.
    - The defect of this method is that attackers can deduct IPv6 addresses based on MAC addresses.
  - Randomly generated by a device
    - The device generates an interface ID randomly. Currently, the Windows operating system uses this method.
  - Manually configured
    - An interface ID can be manually specified.

Contents

1. IPv6 Overview
2. **IPv6 Address Introduction**
   - IPv6 Address Overview
   - **IPv6 Address Types**
   - IPv6 Address Planning Example
3. IPv6 Packet Structure
4. Basic IPv6 Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Address Types

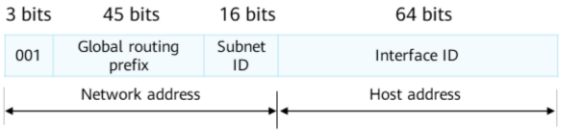


- **Unicast address:** identifies an interface. A packet destined for a unicast address is sent to the interface having that unicast address. In IPv6, an interface may have multiple IPv6 addresses.
- **Multicast address:** identifies multiple interfaces. A packet destined for a multicast address is sent to all the interfaces having that multicast address. Only the interfaces that join a multicast group listen to the packets destined for the corresponding multicast address.
- **Anycast address:** identifies a group of network interfaces (usually on different nodes). A packet sent to an anycast address is routed to the nearest interface having that address, according to the router's routing table.
- **IPv6 does not define any broadcast address.**

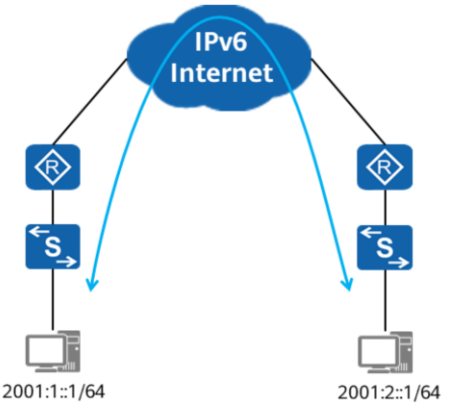
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Common IPv6 Unicast Address — GUA

A global unicast address (GUA) is also called an aggregatable GUA. This type of address is globally unique and is used by hosts that need to access the Internet. It is equivalent to a public IPv4 address.



- Global routing prefix: is assigned by a provider to an organization. Generally, the prefix is at least 48 bits.
- Subnet ID: is defined by an organization based on network requirements.
- Interface ID: identifies a device's interface.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

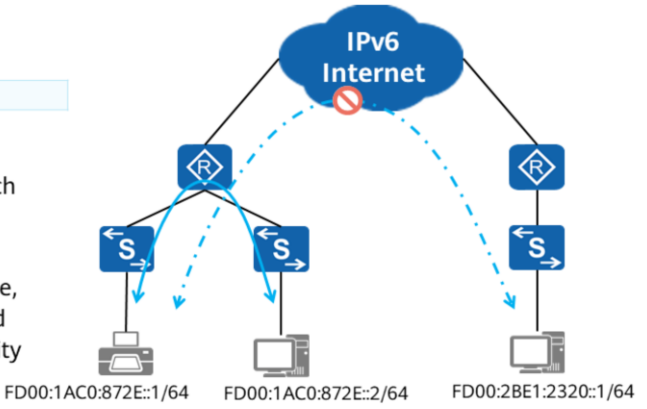
- You can apply for a GUA from a carrier or the local IPv6 address management organization.

Common IPv6 Unicast Address — ULA

A unique local address (ULA) is a private IPv6 address that can be used only on an intranet. This type of address cannot be routed on an IPv6 public network and therefore cannot be used to directly access a public network.

8 bits	40 bits	16 bits	64 bits
1111 1101	Global ID	Subnet ID	Interface ID
Random generation			

- A ULA uses the FC00::/7 address block in which only FD00::/8 is currently used. FC00::/8 is reserved for future expansion.
- Although a ULA is valid only in a limited scope, it also has a globally unique prefix (generated randomly, and therefore the conflict probability is very low).



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

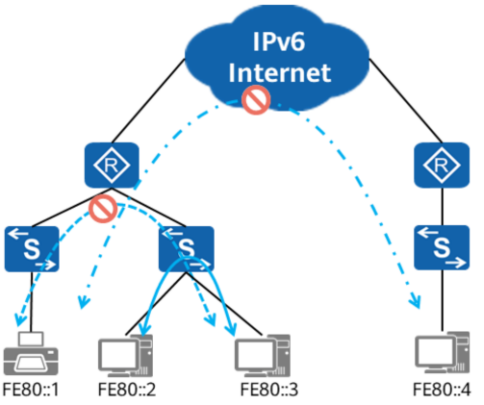
Common IPv6 Unicast Address — LLA

A link-local address (LLA) is another type of IPv6 address with a limited application scope. The valid scope of an LLA is the local link, with the prefix of FE80::/10.

10 bits	54 bits	64 bits
1111 1110 10	0	Interface ID

The value is fixed at 0.

- An LLA is used for communication on a single link, such as during IPv6 SLAAC and IPv6 neighbor discovery.
- Data packets with the source or destination IPv6 address being an LLA are not forwarded out of the originating link. In other words, the valid scope of an LLA is the local link.
- Each IPv6 interface must have an LLA. Huawei devices support automatic generation and manual configuration of LLAs.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Multicast Address

An IPv6 multicast address identifies a multicast group. Packets destined for a multicast address are sent to members in the multicast group. A multicast address is composed of the prefix (FF::/8) as well as the Flags, Scope, and Group ID fields.

8 bits	4 bits	4 bits	80 bits	32 bits
11111111	Flags	Scope	Reserved (must be 0)	Group ID

- **Flags**
  - Indicates a permanent or temporary multicast address.
    - The value 0000 indicates a permanent multicast address.
    - The value 0001 indicates a temporary multicast address.
- **Scope**
  - Indicates the multicast group scope.
- **Group ID**
  - Indicates a multicast group ID.

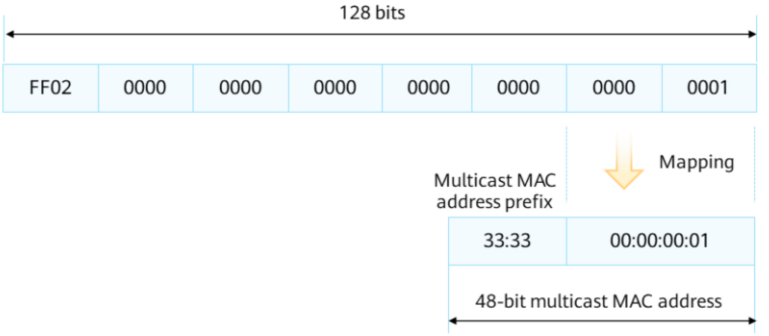
0: reserved
1: interface-local scope, which spans only a single interface on a node and is useful only for loopback transmission of multicast
2: link-local scope (for example, FF02::1)
5: site-local scope
8: organization-local scope
E: global scope
F: reserved

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Unicast Address | IPv6 Multicast Address | IPv6 Anycast Address

IPv6 Multicast MAC Address

- The destination IP address of multicast IPv6 packets is a multicast IPv6 address, and the destination MAC address is a multicast MAC address.
- The first 16 bits of a multicast MAC address are 33:33, which is a MAC address prefix reserved for IPv6 multicast, and the last 32 bits are directly mapped from the last 32 bits of a multicast IPv6 address.

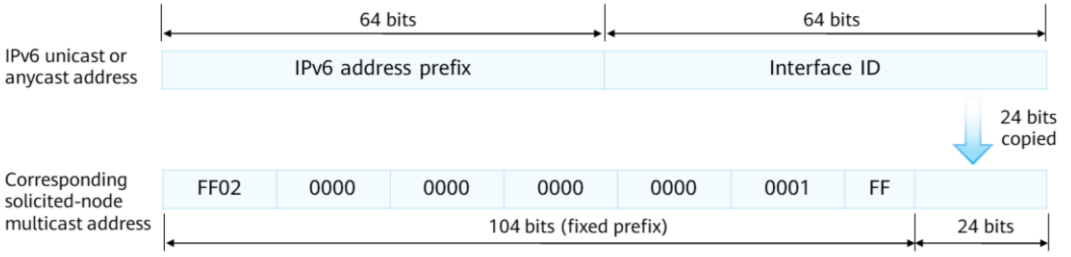


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- When unicast IP packets are transmitted on an Ethernet, they use the MAC addresses of next hops as destination MAC addresses. However, when multicast packets are transmitted, their destination is a group of unspecific members but not a specific receiver. Therefore, they use a multicast MAC address as the destination MAC address.
- IPv4 multicast MAC address
  - As defined by the IANA, the 24 most significant bits of an IPv4 multicast MAC address are 0x01005E, the 25th bit is 0, and the 23 least significant bits are mapped from those of an IPv4 multicast address.
  - The most significant four bits of an IPv4 multicast address are 1110, indicating the multicast identifier. However, only 23 bits of the least significant 28 bits are mapped to the IPv4 multicast MAC address. As a result, 5 bits of the IPv4 multicast address are lost. Therefore, 32 IPv4 multicast addresses are mapped to the same IPv4 multicast MAC address. During Layer 2 processing, a device may need to receive multicast data from multicast groups other than the local IPv4 multicast group. In this case, the redundant multicast data needs to be filtered by the upper layer.
- IPv6 multicast MAC address
  - When an IPv6 multicast packet is sent on an Ethernet link, the corresponding MAC address is 0x3333-A-A-A-A, where A-A-A-A is directly mapped from the last 32 bits of a multicast IPv6 address.

Solicited-Node Multicast Address

If a node has an IPv6 unicast or anycast address, a solicited-node multicast address is generated for the node, and the node joins the corresponding multicast group. This address is used for neighbor discovery and duplicate address detection. A solicited-node multicast address is valid only on the local link.

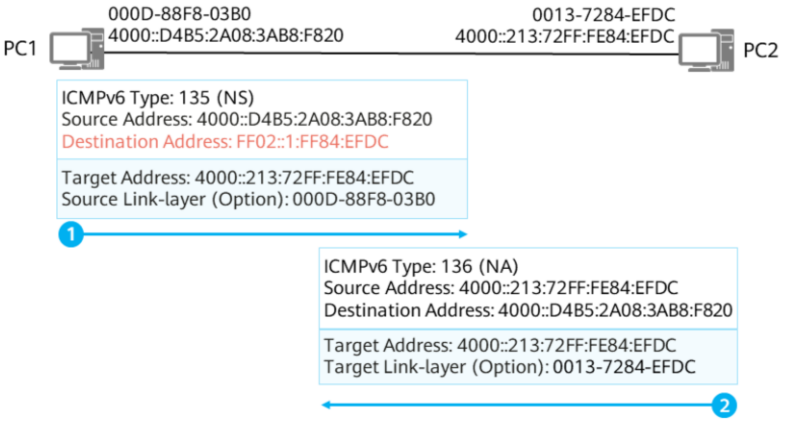


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- An application scenario example of a solicited-node multicast address is as follows: In IPv6, ARP and broadcast addresses are canceled. When a device needs to request the MAC address corresponding to an IPv6 address, the device still needs to send a request packet, which is a multicast packet. The destination IPv6 address of the packet is the solicited-node multicast address corresponding to the target IPv6 unicast address. Because only the target node listens to the solicited-node multicast address, the multicast packet is received only by the target node, without affecting the network performance of other non-target nodes.

Solicited-Node Multicast Address Example

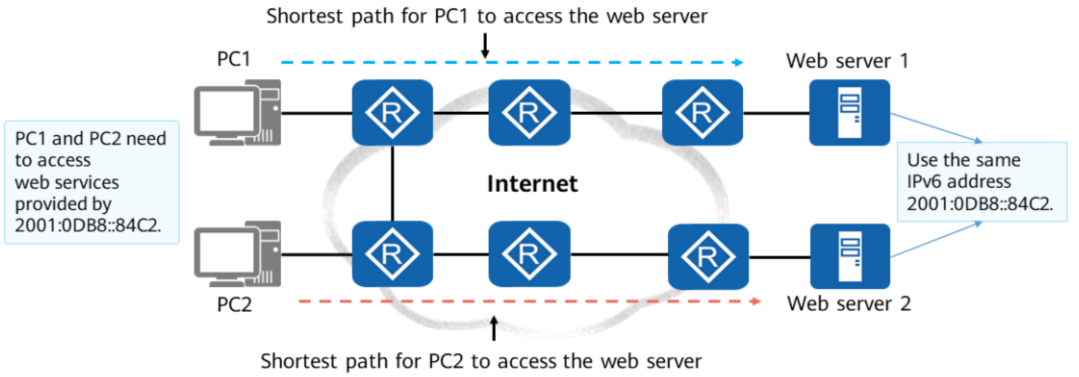
Before sending data to PC2, PC1 needs to obtain the MAC address of PC2. PC1 initiates a process similar to ARP resolution in IPv4. IPv6 uses ICMPv6 NS and NA messages to implement address resolution. The destination IPv6 address of NS messages is the solicited-node multicast address corresponding to the target IPv6 unicast address.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Anycast Address

An anycast address identifies a group of network interfaces, which usually belong to different nodes. An anycast address can be used as the source or destination address of IPv6 packets.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The anycast process involves an anycast packet initiator and one or more responders.
  - An initiator of an anycast packet is usually a host requesting a service (for example, a web service).
  - The format of an anycast address is the same as that of a unicast address. A device, however, can send packets to multiple devices with the same anycast address.
- Anycast addresses have the following advantages:
  - Provide service redundancy. For example, a user can obtain the same service (for example, a web service) from multiple servers that use the same anycast address. These servers are all responders of anycast packets. If no anycast address is used and a server fails, the user needs to obtain the address of another server to establish communication again. If an anycast address is used and a server fails, the user can automatically communicate with another server that uses the same address, implementing service redundancy.
  - Provide better services. For example, a company deploys two servers – one in province A and the other in province B – to provide the same web service. Based on the optimal route selection rule, users in province A preferentially access the server deployed in province A when accessing the web service provided by the company. This increases the access speed, reduces the access delay, and greatly improves user experience.

Comparison Between IPv6 and IPv4 Addresses

	IPv4	IPv6
Address space	2^{32}	2^{128}
Representation mode	Dotted decimal notation	Colon hexadecimal notation
Address type	Unicast, multicast, and broadcast	Unicast, multicast, and anycast
Others	Class A, B, C, D, and E addresses	N/A
	Multicast address 224.0.0.0/4	IPv6 multicast address FF00::/8
	Broadcast address	N/A
	Unspecified address 0.0.0.0/32	Unspecified address ::/128
	Loopback address 127.0.0.0/8	Loopback address ::1/128
	Public IP address	Global unicast address
	Private IP addresses 10.0.0.0/8, 172.16.0.0/12, and 192.168.0.0/16	Unique local address FD00::/8
	APIPA address 169.254.0.0/16	Link-local address FE80::/10

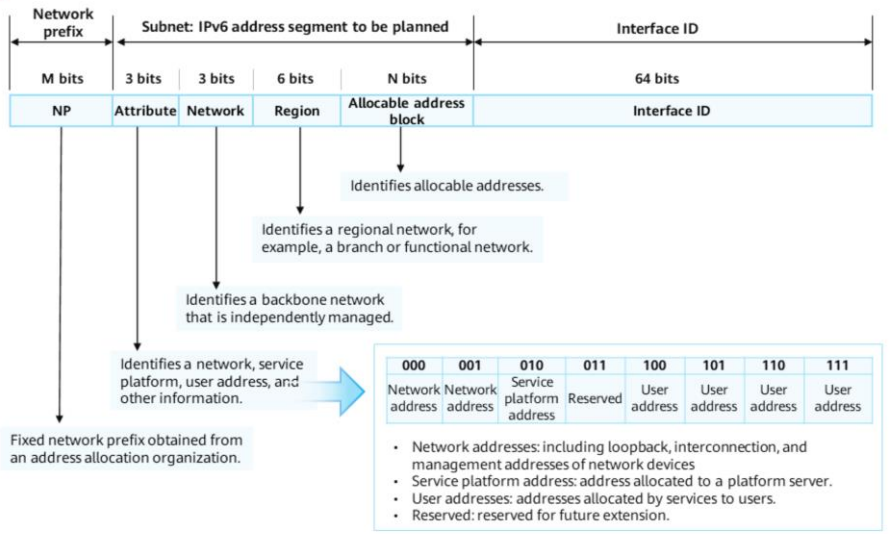
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. IPv6 Overview
2. **IPv6 Address Introduction**
   - IPv6 Address Overview
   - IPv6 Address Types
   - **IPv6 Address Planning Example**
3. IPv6 Packet Structure
4. Basic IPv6 Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Address Planning Example



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

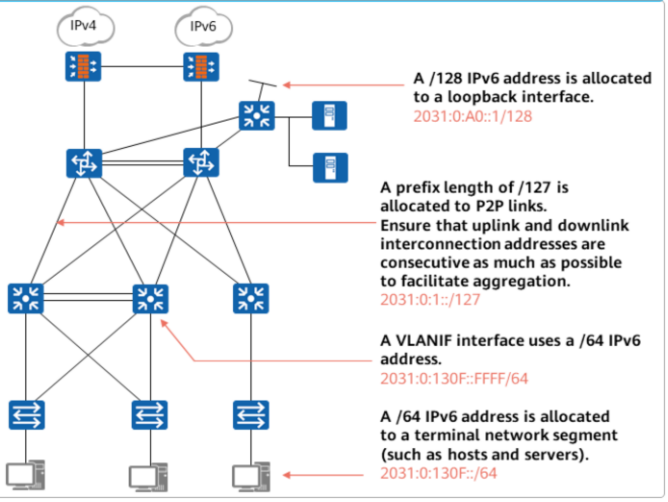
- Address planning and design suggestions:
  - Based on the obtained address prefix, determine the number of functional blocks (for example, 3+3+6+N in the figure) into which the subnet address is divided, and determine the meaning of each functional block and the number of bits occupied by it to avoid address waste.

Suggestions on IPv6 Address Use

Key Points of Address Planning

- Generally, the IPv6 addresses obtained by an ISP from an address allocation organization have a prefix length of at least /32. The prefix length of the IPv6 addresses obtained by an enterprise user is usually /48.
- Although IPv6 addresses have a large space, they must still be properly planned during network planning.
- Ensure that each IPv6 address is unique on the entire network (except for some special applications, such as anycast).
- Ensure the continuity, aggregability, and scalability of IPv6 addresses.
- Service address: Plan appropriate bits in an IPv6 address to carry service, VLAN, or location information, facilitating route planning and QoS deployment.
- User address: Reserve consecutive address segments for users with different service types. Determine user types and distribution areas based on specific bits.
- It is recommended that a prefix length of /64 be allocated to a terminal network segment (such as user hosts and servers).
- It is recommended that a prefix length of /127 be allocated to P2P links.
- It is recommended that a prefix length of /128 be allocated to loopback interfaces.

Address Planning Practice



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

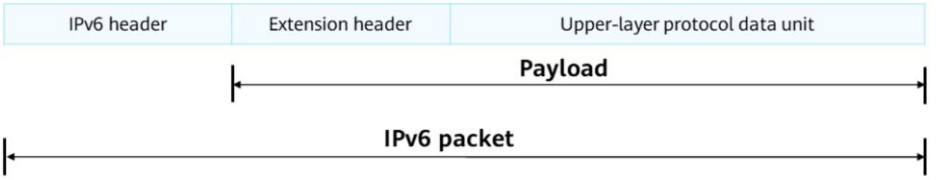
Contents

1. IPv6 Overview
2. IPv6 Address Introduction
3. **IPv6 Packet Structure**
4. Basic IPv6 Configuration

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

IPv6 Packet Composition

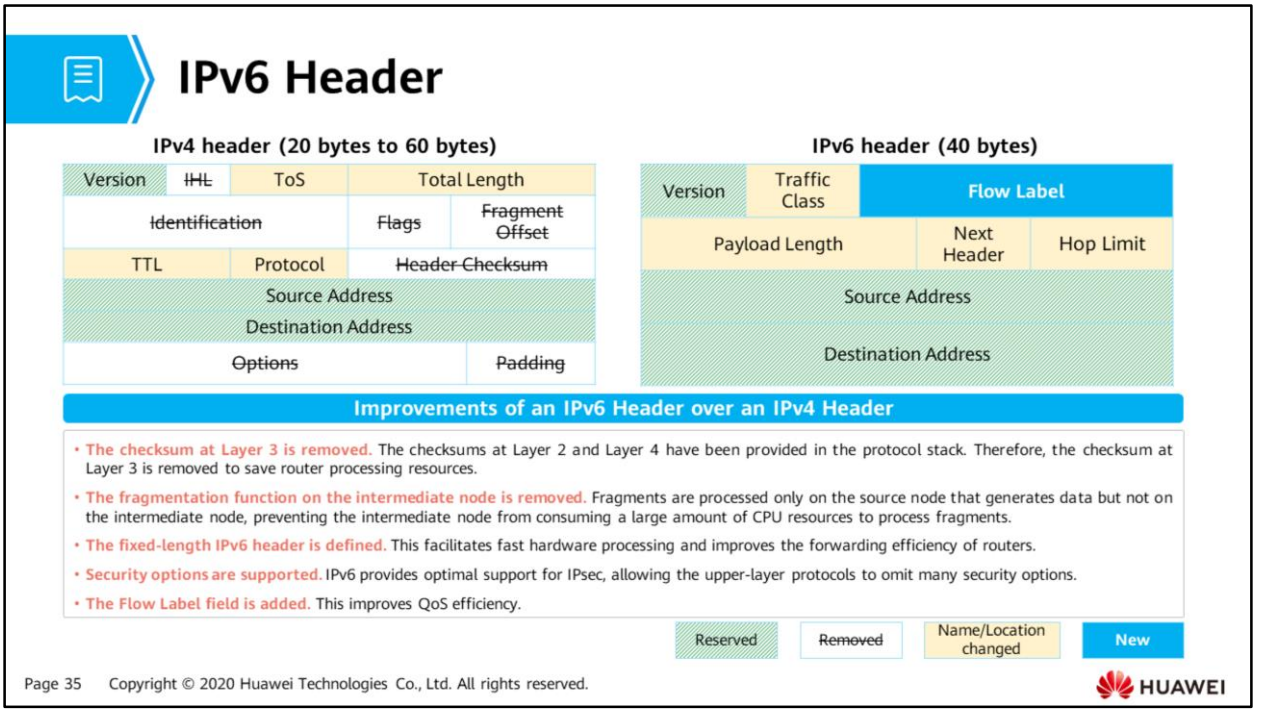


An IPv6 packet is generally composed of the following parts:

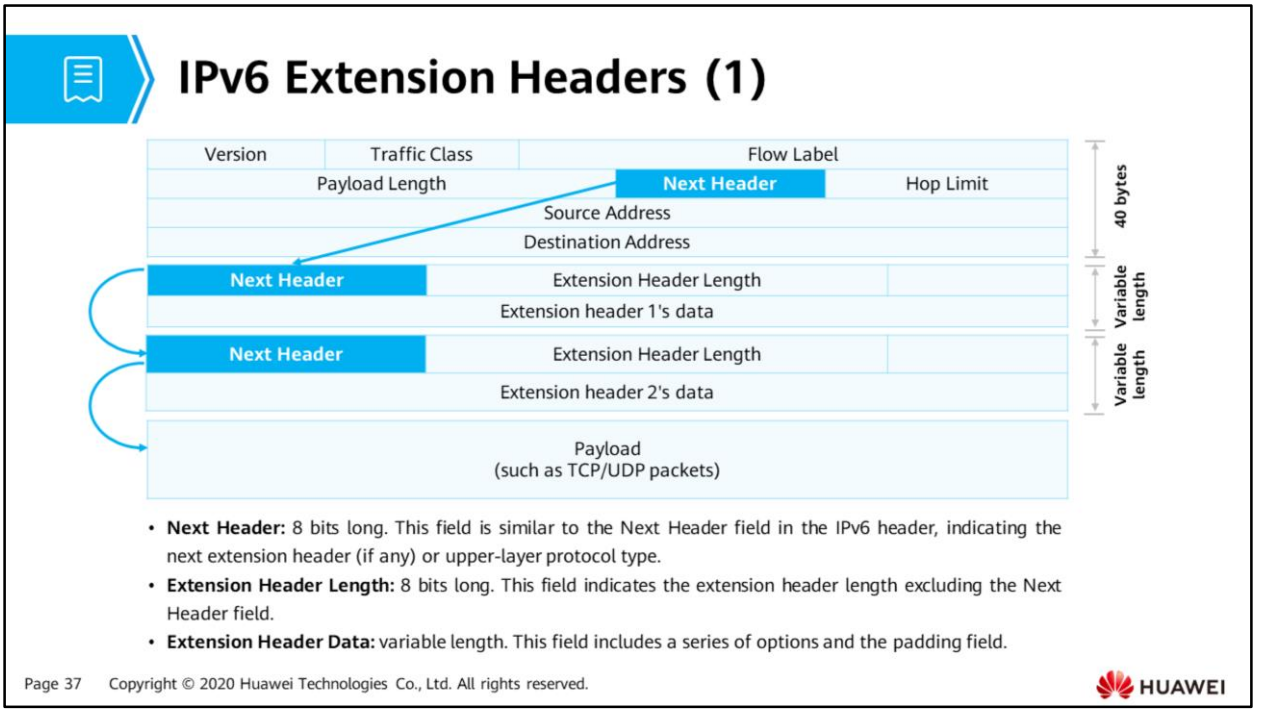
- **IPv6 header:** provides basic packet forwarding information. Routers use the information to forward most packets.
- **Extension headers:** provide extended packet forwarding information, such as segmentation and encryption. This part is optional and does not need to be processed by each router. A sender adds one or more extension headers only when a router or destination node needs to perform special processing.
- **Upper-layer protocol data unit:** is generally composed of an upper-layer protocol header and its payload. This part is similar to the upper-layer protocol data unit in an IPv4 packet.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- As shown in the figure, an IPv6 packet is composed of the following parts:
  - **IPv6 header**
    - Each IPv6 packet must contain a header with a fixed length of 40 bytes.
    - The IPv6 header provides basic packet forwarding information, which is parsed by all routers on a forwarding path.
  - **Extension headers**
    - An IPv6 extension header is an optional header that may follow an IPv6 header. An IPv6 packet can contain no extension header, or it can contain one or more extension headers with different lengths. The IPv6 header and extension headers replace the IPv4 header and its options. The extension headers enhances IPv6 significantly. Unlike the options in an IPv4 header, the maximum length of an extension header is not limited. Therefore, an extension header can contain all the extension data required for IPv6 communication. The extended packet forwarding information provided by an extension header is generally parsed by the destination router but not all routers on a path.
  - **Upper-layer protocol data unit**
    - An upper-layer protocol data unit is composed of the upper-layer protocol header and its payload, which can be an ICMPv6 packet, a TCP packet, or a UDP packet.

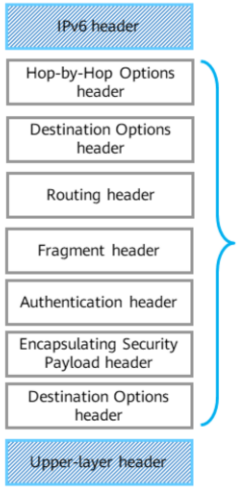


- The IPv6 header is also called a fixed header, which contains eight fields. The total length of the fixed header is 40 bytes. The eight fields are Version, Traffic Class, Flow Label, Payload Length, Next Header, Hop Limit, Source Address, and Destination Address.
- Version
  - This field indicates the version of IP and its value is 6.The length is 4 bits.
- Traffic Class
  - This field indicates the class or priority of an IPv6 packet and its function is similar to that of the ToS field in an IPv4 header. The length is 8 bits.
- Flow Label
  - This field is used by a source to label sequences of packets for which it requests special handling by IPv6 routers. The length is 20 bits. Generally, a flow can be determined based on the source IPv6 address, destination IPv6 address, and flow label.
- Payload Length
  - This field indicates the length of the IPv6 payload. The payload refers to the extension header and upper-layer protocol data unit that follow the IPv6 header. The length is 16 bits. If the payload length exceeds its maximum value of 65535 bytes, the field is set to 0, and the Jumbo Payload option in the Hop-by-Hop Options header is used to express the actual payload length.



- The Options field in an IPv4 header is placed in extension headers of an IPv6 packet. An IPv6 extension header is an optional header that may follow an IPv6 header. Why is an extension header designed in IPv6? Each intermediate router must check whether the options contained in the IPv4 header exist. If the options exist, the intermediate router must process them. This reduces the efficiency for routers to forward IPv4 packets. Therefore, the Options field is placed in extension headers in IPv6 to resolve this issue. In this case, the intermediate router does not need to process each possible option, accelerating packet processing and improving forwarding performance.
- A typical IPv6 packet does not contain any extension header. A sender adds one or more extension headers only when a router or destination node needs to perform special processing. Unlike IPv4, IPv6 has variable-length extension headers, which are not limited to 40 bytes, to facilitate further extension. To improve extension header processing efficiency and transport protocol performance, IPv6 requires that the extension header length be an integral multiple of 8 bytes.

IPv6 Extension Headers (2)



Conventions for extension headers:

- Extension headers must appear in the order on the left.
- Each extension header can appear only once,
- except for the Destination Options header that can appear at most twice (once before the Routing header and once before the upper-layer header). If there is no Routing header, the Destination Options header can appear only once.
- The relationships between the IPv6 header, extension headers, and upper-layer header are as follows:

IPv6 Header Next Header = 6 (TCP)	TCP Segment		
IPv6 Header Next Header = 43 (Routing)	Routing Header Next Header = 6 (TCP)	TCP Segment	
IPv6 Header Next Header = 43 (Routing)	Routing Header Next Header = 44 (Fragment)	Fragment Header Next Header = 6 (TCP)	TCP Segment

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Currently, RFC 2460 defines the following six IPv6 extension headers:
  - Hop-by-Hop Options header: is used to carry multiple options such as the router alarm option that must be examined by every node along a packet's delivery path.
  - Destination Options header: is used to carry multiple options such as the home address option of mobile IPv6 that need to be examined only by a packet's destination node.
  - Routing header: is used by an IPv6 source to list all intermediate nodes to be "visited" on the way to a packet's destination. This function is very similar to IPv4's Loose Source and Record Route option. The destination address in the IPv6 header is not the final destination address of a packet but the first address listed in the Routing header.
  - Fragment header: is used by an IPv6 source to send a packet that is too large to fit in the MTU of the path to its destination. The Fragment header is processed only by the destination node.
  - Authentication header: is used by IPsec and processed only by the destination node.
  - Encapsulating Security Payload header: is used by IPsec and processed only by the destination node.
- The Hop-by-Hop Options and Destination Options headers provide option functions and support extensibility (such as mobility).Options use the TLV mode.

Contents

1. IPv6 Overview
2. IPv6 Address Introduction
3. IPv6 Packet Structure
4. **Basic IPv6 Configuration**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Configuration (1)

1. Enable IPv6 packet forwarding on a device.

```
[<Huawei> system-view
[Huawei] ipv6
```

2. Enable IPv6 on an interface.

```
[Huawei] interface interface-type interface-number
[Huawei-GigabitEthernet1/0/0] ipv6 enable
```

3. Configure global unicast IPv6 addresses.

```
[Huawei-GigabitEthernet1/0/0] ipv6 address { ipv6-address prefix-length | ipv6-address/prefix-length }
```

A maximum of 10 global unicast addresses can be configured on each interface.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration (2)

1. Enable OSPFv3 on a device.

```
[<Huawei> system-view
[Huawei] ospfv3 [ process-id ]
```

OSPFv3 supports multi-process. Multiple OSPFv3 processes running on a router are differentiated by process IDs. You can set an OSPFv3 process ID when enabling OSPFv3. The process ID is only locally valid and does not affect packet exchange with other routers.

2. Configure a router ID for an OSPFv3 process.

```
[Huawei-ospfv3-1] router-id router-id
```

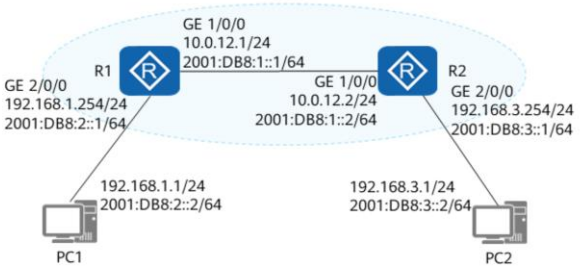
A router ID must be manually configured, and OSPFv3 cannot run properly without a router ID.

3. Enable an OSPFv3 process on an interface and specify an area to which the process belongs.

```
[Huawei] interface interface-type interface-number
[Huawei-GigabitEthernet1/0/0] ospfv3 process-id area area-id
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Configuring a Dual-Stack Network (1)



Configuration requirements:

- Configure interface IP addresses on R1 and R2.
- Configure OSPFv2 and OSPFv3 on R1 and R2 respectively to implement dual-stack communication between PC1 and PC2.

1. Enable IPv6 globally and on related interfaces on R1 and R2. The following uses R1 as an example.

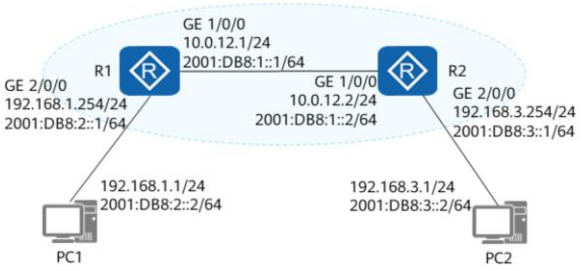
```
[R1]ipv6
[R1]interface GigabitEthernet 1/0/0
[R1-GigabitEthernet1/0/0]ipv6 enable
[R1]interface GigabitEthernet 2/0/0
[R1-GigabitEthernet2/0/0]ipv6 enable
```

2. Configure global unicast IPv4 and IPv6 addresses on the interfaces of the PCs, R1, and R2. The following uses R1 as an example.

```
[R1]interface GigabitEthernet 1/0/0
[R1-GigabitEthernet1/0/0]ip address 10.0.12.1 24
[R1-GigabitEthernet1/0/0]ipv6 address 2001:DB8:1::1 64
[R1]interface GigabitEthernet 2/0/0
[R1-GigabitEthernet2/0/0]ip address 192.168.1.254 24
[R1-GigabitEthernet2/0/0]ipv6 address 2001:DB8:2::1 64
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Configuring a Dual-Stack Network (2)



Configuration requirements:

- Configure interface IP addresses on R1 and R2.
- Configure OSPFv2 and OSPFv3 on R1 and R2 respectively to implement dual-stack communication between PC1 and PC2.

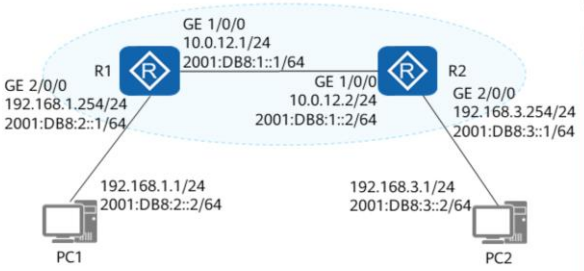
Configure OSPFv2 on R1 and R2 so that PC1 and PC2 can communicate through the IPv4 network.

```
[R1]ospf 1 router-id 10.0.1.1
[R1-ospf-1]area 0
[R1-ospf-1-area-0.0.0.0]network 10.0.12.0 0.0.0.255
[R1-ospf-1-area-0.0.0.0]network 192.168.1.0 0.0.0.255
[R1-ospf-1-area-0.0.0.0]quit
[R1-ospf-1]
```

```
[R2]ospf 1 router-id 10.0.2.2
[R2-ospf-1]area 0
[R2-ospf-1-area-0.0.0.0]network 10.0.12.0 0.0.0.255
[R2-ospf-1-area-0.0.0.0]network 192.168.3.0 0.0.0.255
[R2-ospf-1-area-0.0.0.0]quit
[R2-ospf-1]
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Case: Configuring a Dual-Stack Network (3)



Configuration requirements:

- Configure interface IP addresses on R1 and R2.
- Configure OSPFv2 and OSPFv3 on R1 and R2 respectively to implement dual-stack communication between PC1 and PC2.

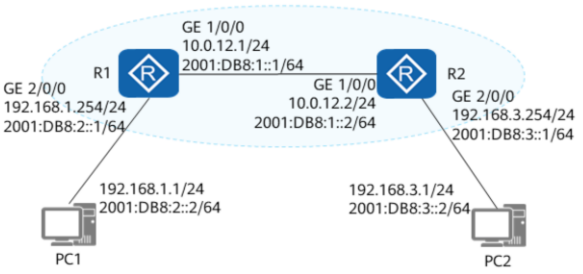
Configure OSPFv3 on R1 and R2 so that PC1 and PC2 can communicate through the IPv6 network.

```
[R1]ospfv3
[R1-ospfv3-1]router-id 10.0.1.1
[R1-ospfv3-1]quit
[R1]interface gigabitethernet1/0/0
[R1-gigabitethernet1/0/0]ospfv3 1 area 0
[R1]interface gigabitethernet2/0/0
[R1-gigabitethernet2/0/0]ospfv3 1 area 0
```

```
[R2]ospfv3
[R2-ospfv3-1]router-id 10.0.2.2
[R2-ospfv3-1]quit
[R2]interface gigabitethernet1/0/0
[R2-gigabitethernet1/0/0]ospfv3 1 area 0
[R2]interface gigabitethernet2/0/0
[R2-gigabitethernet2/0/0]ospfv3 1 area 0
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Verifying the Configuration



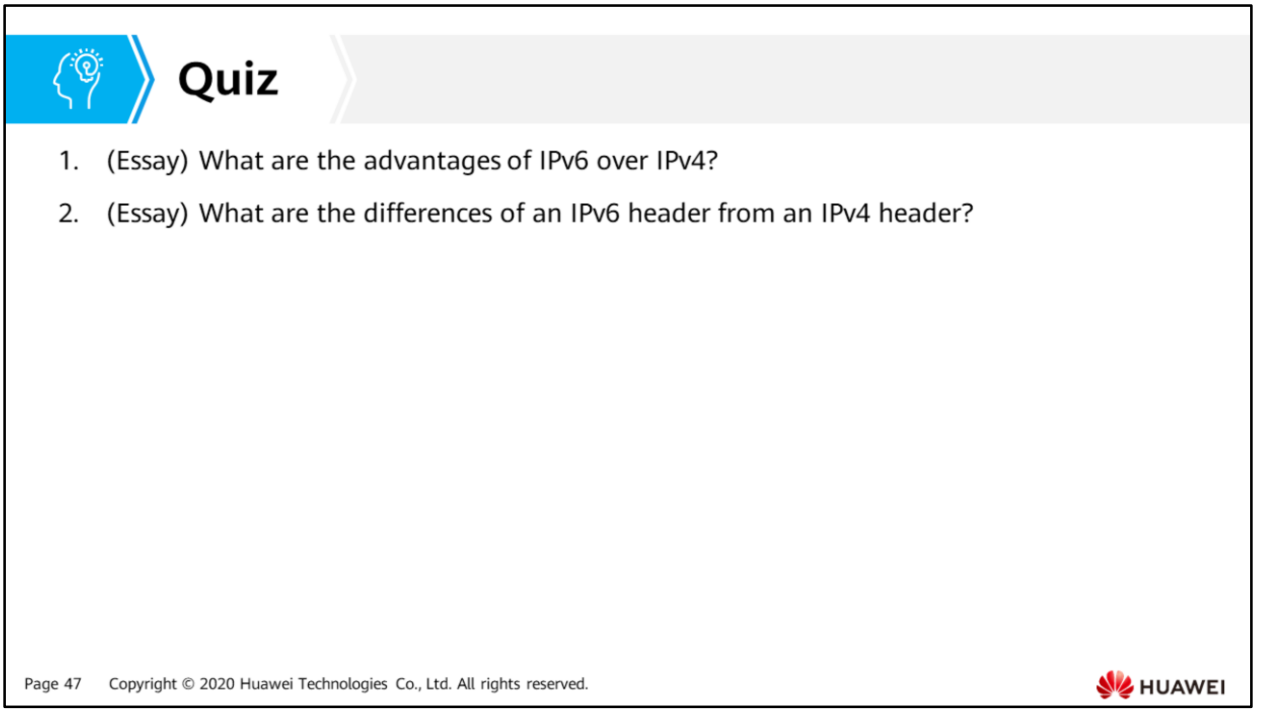
Ping the IPv6 address of PC1 from R1 (pinging the IPv4 address is omitted).

```
<R1>ping ipv6 2001:db88:2::2
  PING 2001:db88:2::2 : 56  data bytes, press CTRL_C to break
    Reply from 2001:DB88:2::2
    bytes=56 Sequence=1  hop limit=255  time = 40 ms
    Reply from 2001:DB88:2::2
    bytes=56 Sequence=2  hop limit=255  time = 10 ms
```

Ping the IPv6 address of PC2 from PC1 (pinging the IPv4 address is omitted).

```
PC1>ping 2001:db88:3::2
Ping 2001:db88:3::2: 32 data bytes, Press Ctrl_C to break
From 2001:db88:3::2: bytes=32 seq=1 hop limit=64 time<1 ms
From 2001:db88:3::2: bytes=32 seq=2 hop limit=64 time=16 ms
From 2001:db88:3::2: bytes=32 seq=3 hop limit=64 time=16 ms
From 2001:db88:3::2: bytes=32 seq=4 hop limit=64 time<1 ms
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI



1. Unlimited address space, hierarchical address structure, plug-and-play, simplified packet header, security features, mobility, and enhanced QoS features.

2. Differences between IPv6 header and IPv4 header

- The packet format of IPv6 header+extension headers is used.
- The checksum at Layer 3 is removed. The checksums at Layer 2 and Layer 4 are sufficiently robust, and therefore the checksum at Layer 3 is removed to save router processing resources.
- The fragmentation function on the intermediate node is removed. Fragments are processed only on the source node that generates data but not on the intermediate router, preventing the intermediate router from consuming a large amount of CPU resources to process fragments.
- The fixed-length IPv6 header is defined to facilitate fast hardware processing and improve the forwarding efficiency of routers.
- Security options are supported. IPv6 provides optimal support for IPsec, allowing the upper-layer protocols to omit many security options.
- The Flow Label field is added to improve QoS efficiency.

Summary

- As a next-generation Internet protocol, IPv6 has more advantages than IPv4 and can meet service development requirements.
- IPv6 has a large address space. It also simplifies packet headers, improving the packet forwarding efficiency of routers. IPv6 addresses are easy to divide and plan, facilitating route aggregation. In addition, IPv6 supports plug-and-play and enhances QoS.
- The dynamic routing protocols commonly used on IPv4 networks can still be used on IPv6 networks through version upgrade (OSPFv3) or protocol extension (IS-IS for IPv6 or BGP4+).

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



ICMPv6 and NDP

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- In IPv4, Internet Control Message Protocol (ICMP) allows hosts or devices to report errors. An ICMP message is encapsulated in an IP packet as the data, and forms a complete IP packet together with an IP header. Common ping and tracert commands are implemented based on ICMP.
- In IPv6, the Internet Control Message Protocol for IPv6 (ICMPv6) provides functions similar to ICMP and extends it. The Neighbor Discovery Protocol (NDP) is implemented based on ICMPv6. As a key IPv6 protocol, NDP provides functions such as prefix discovery, duplicate address detection (DAD), address resolution, and redirection.
- This course describes ICMPv6 and NDP in detail.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- On completion of this course, you will be able to:
  - Describe the functions of ICMPv6.
  - Describe the ICMPv6 message format.
  - Describe ICMPv6 message types.
  - Describe the functions of NDP.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

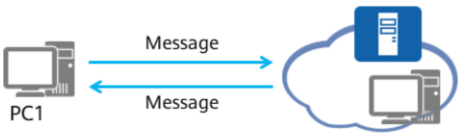
1. **ICMPv6 Introduction**
   - **ICMPv6 Overview**
   - ICMPv6 Message Format
   - ICMPv6 Message Types and Functions
2. NDP Introduction

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

ICMP

ICMP is an auxiliary protocol of IP.

ICMP is used to transmit error and control information between network devices. It plays an important role in collecting network information and diagnosing and rectifying network faults.



Ethernet header	IP header	ICMP message	Ethernet trailer

Type	Code	Checksum
ICMP message content		

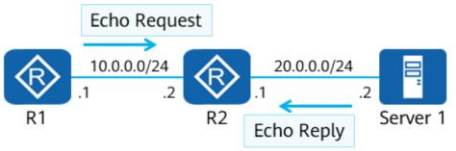
Type	Code	Description
0	0	Echo Reply
3	0	Network unavailable
3	1	Host unreachable
3	2	Protocol unreachable
3	3	Port unreachable
5	0	Redirection
8	0	Echo Request

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- ICMP works at the network layer to ensure the correct forwarding of IP packets, and allows hosts or devices to report errors during packet transmission.
- ICMP message:
  - ICMP messages are encapsulated in IP packets. If the Protocol value in the IP header is 1, the used protocol is ICMP.
  - Field description:
    - The format of an ICMP message depends on the Type and Code fields. The Type field indicates the message type, and the Code field contains specific parameters of the message type.
    - The Checksum field is used to check whether the message is complete.

ICMP Error Check

ICMP Echo messages are used to check the network connectivity between the source and destination. In addition, they provide other information, such as the round-trip time of packets.



Function: ping

Ping is a command used on network devices as well as the Windows, Unix, and Linux platforms. It is a small and practical application based on ICMP.

Ping is often used to check the network reachability to a destination node.

```
[R1]ping 20.0.0.2
  PING 20.0.0.2: 56  data bytes, press CTRL_C to break
    Reply from 20.0.0.2: bytes=56 Sequence=1 ttl=254 time=70 ms
    Reply from 20.0.0.2: bytes=56 Sequence=2 ttl=254 time=30 ms
    Reply from 20.0.0.2: bytes=56 Sequence=3 ttl=254 time=30 ms
    Reply from 20.0.0.2: bytes=56 Sequence=4 ttl=254 time=40 ms
    Reply from 20.0.0.2: bytes=56 Sequence=5 ttl=254 time=30 ms

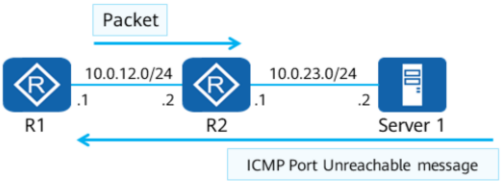
  --- 20.0.0.2 ping statistics ---
    5 packet(s) transmitted
    5 packet(s) received
    0.00% packet loss
    round-trip min/avg/max = 30/40/70 ms
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Ping is a typical application of ICMP. It is a common tool used to check network connectivity and collect information. You can configure various parameters in the **ping** command, such as the length of ICMP messages, number of sent ICMP messages, and timeout period for waiting for a reply. A device constructs and sends ICMP messages based on the configured parameters to perform a ping test.

ICMP Error Report

ICMP defines various error messages for diagnosing network connectivity faults. Based on the error messages, the source can determine the cause of a data transmission failure. For example, when a network device cannot access a target network, it automatically sends an ICMP Destination Unreachable message to the sender.




```
[R1]tracert 10.0.23.2

 traceroute to  10.0.23.2, max hops: 30 ,packet length: 40,press CTRL_C to break

 1 10.0.12.2                80 ms    10 ms    10 ms

 2 10.0.23.2                30 ms    30 ms    20 ms
```

Function: tracert

Tracert traces the packet forwarding path hop by hop based on the TTL value in the packet header. It is an effective method to check the packet loss and delay and to help administrators find routing loops on a network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

ICMPv6 Overview

- ICMPv6 is one of the basic IPv6 protocols.
- If the Next Header field value in the IPv6 header is 58, the used protocol is ICMPv6.
- ICMPv6 messages are used to advertise information or errors.
- ICMPv6 messages are widely used in other protocols, such as NDP and path MTU discovery.
- ICMPv6 controls key IPv6 processes, such as address autoconfiguration, address resolution, address conflict detection, route selection, and error control.

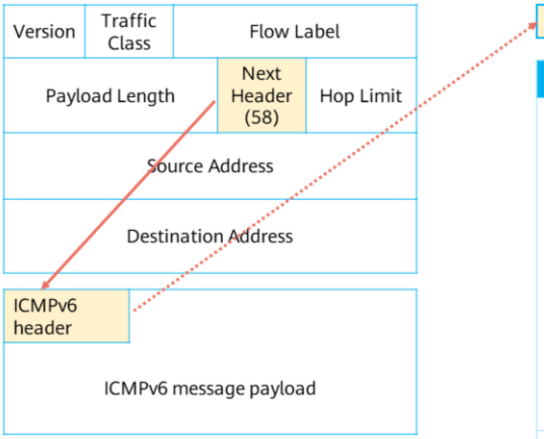
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **ICMPv6 Introduction**
   - ICMPv6 Overview
   - **ICMPv6 Message Format**
   - ICMPv6 Message Types and Functions
2. NDP Introduction

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

ICMPv6 Message Format



Type	Code	Checksum

Message Type	Type	Name	Code
Error messages	1	Destination Unreachable	0: no route to destination
			1: communication with destination administratively prohibited
			2: beyond scope of source address
			3: address unreachable
			4: port unreachable
	2	Packet Too Big	0
	3	Time Exceeded	0: hop limit exceeded in transit
			1: fragment reassembly time exceeded
	4	Parameter Problem	0: erroneous header field encountered
			1: unrecognized Next Header type encountered
			2: unrecognized IPv6 option encountered
Informational messages	128	Echo Request	0
	129	Echo Reply	0

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The payload of an ICMPv6 message is determined by its type.
- **Type:** indicates the message type.
- **Code:** depends on the message type.
- **Checksum:** indicates the ICMPv6 message checksum.

Contents

1. **ICMPv6 Introduction**
   - ICMPv6 Overview
   - ICMPv6 Message Format
   - **ICMPv6 Message Types and Functions**
2. NDP Introduction

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

ICMPv6 Message Types

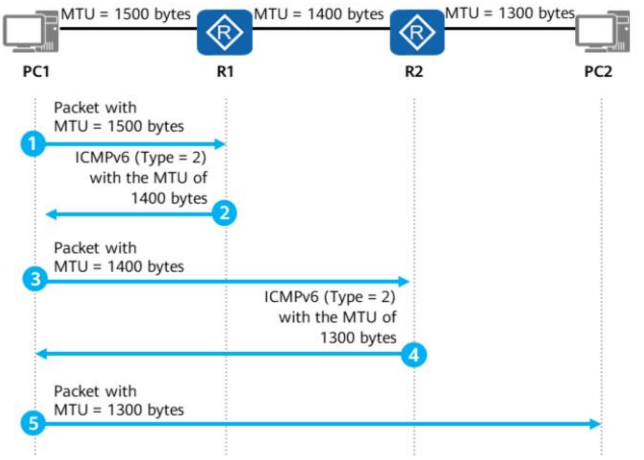
ICMPv6 messages are classified as error or informational messages.

Error Messages	Informational Messages
• The most significant bit of the Type field in error messages is 0. That is, ICMPv6 Type = [0, 127]. • Error messages are used to report errors that occur during IPv6 packet forwarding, such as destination unreachable and timeout.	• The most significant bit of the Type field in informational messages is 1. That is, ICMPv6 Type = [128, 255]. • Informational messages can be used to implement inter-node communication on a link and multicast member management on a subnet.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

ICMPv6 Error Message Application — Path MTU Discovery

- In IPv6, an intermediate forwarding device does not fragment IPv6 packets, and packet fragmentation is performed on the source node.
- Path MTU (PMTU) is the minimum interface MTU on a path.
- Path MTU Discovery (PMTUD) is used to discover the MTU on a path, preventing fragmentation when a packet is forwarded from the source to the destination.
- With PMTUD, the data sender can use the optimal PMTU discovered to communicate with the destination node. This prevents packets from being fragmented by intermediate routers during transmission from the source to the destination.



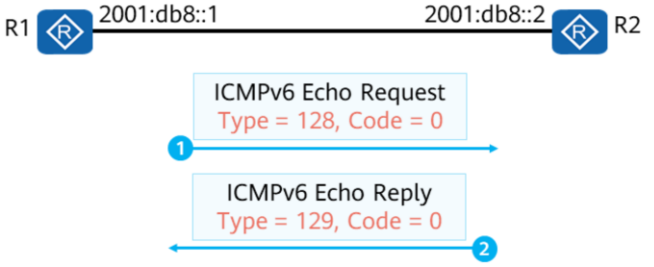
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

1. PC1 sends an IPv6 packet with an MTU of 1500 bytes to PC2.
2. After checking that the packet is too large and the MTU of the outbound interface is 1400 bytes, R1 sends an ICMPv6 (Type = 2) message with the MTU of 1400 bytes to PC1.
3. PC1 sends an IPv6 packet with an MTU of 1400 bytes.
4. After the packet reaches R2, R2 checks that the MTU of the outbound interface is 1300 bytes and sends an ICMPv6 (Type = 2) message with the MTU of 1300 bytes to PC1.
5. PC1 sends an IPv6 packet with an MTU of 1300 bytes.

ICMPv6 Informational Message Application — Ping

Ping is implemented based on ICMPv6 informational messages.

- Echo Request: is sent to a destination node, which responds with an Echo Reply message. The Type and Code field values in an Echo Request message are 128 and 0, respectively.
- Echo Reply: After receiving an Echo Request message, ICMPv6 responds with an Echo Reply message. The Type and Code field values in an Echo Reply message are 129 and 0, respectively.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Other Common ICMPv6 Messages

- Neighbor Discovery (RFC 2461 and RFC 4861)
  - Type = 133: Router Solicitation
  - Type = 134: Router Advertisement
  - Type = 135: Neighbor Solicitation
  - Type = 136: Neighbor Advertisement
  - Type = 137: Redirect
- Multicast Listener Discovery (RFC 2710 and RFC 3810)
  - Type = 130: Multicast Listener Query
  - Type = 131: Multicast Listener Report
  - Type = 132: Multicast Listener Done
  - Type = 143: Version 2 Multicast Listener Report

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. ICMPv6 Introduction
2. **NDP Introduction**
   - **NDP Overview**
   - Router Discovery
   - Address Resolution
   - Neighbor State Tracing
   - DAD
   - Redirection

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

NDP Overview

RFC 2461 defines the IPv6 Neighbor Discovery Protocol (NDP), which is a core component in IPv6. Its functions are as follows:

NDP	Router discovery	Discovers routers on a link and obtains information advertised by them.
	Stateless address autoconfiguration	A terminal automatically generates an IPv6 address based on the address prefix advertised by a router.
	DAD	Detects duplicate addresses and ensures that no address conflict occurs.
	Address resolution	Requests the link-layer address corresponding to a destination network address. This function is similar to ARP in IPv4.
	Neighbor state tracing	Discovers neighbors on a link through NDP and traces the neighbor state.
	Prefix readdressing	A router flexibly sets an advertised address prefix to implement readdressing.
	Redirection	Informs other devices of the optimal next hop to a target network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

NDP Message Types and Functions

NDP uses the following ICMPv6 messages:

- Router Solicitation (RS)
- Router Advertisement (RA)
- Neighbor Solicitation (NS)
- Neighbor Advertisement (NA)

Function \ ICMPv6 Message	RS (Type = 133)	RA (Type = 134)	NS (Type = 135)	NA (Type = 136)	Redirect (Type = 137)
Address resolution			●	●	
Router discovery	●	●			
Prefix readdressing	●	●			
DAD			●	●	
Redirection					●

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Contents

1. ICMPv6 Introduction
2. **NDP Introduction**
   - NDP Overview
   - **Router Discovery**
   - Address Resolution
   - Neighbor State Tracing
   - DAD
   - Redirection

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

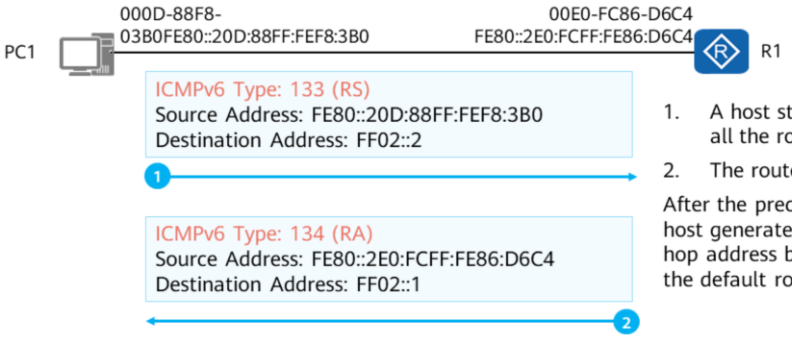
Router Discovery

- Router discovery is a process in which a host discovers a router on a local link and determines its configuration.
- Router discovery can implement the following functions:
  - Router discovery: A host locates neighboring routers and selects a router as the default gateway.
  - Prefix discovery: A host discovers a group of IPv6 prefixes on a local link for address autoconfiguration.
  - Parameter discovery: A host discovers operation parameters, such as the default hop limit of output packets and address configuration mode.
- Used messages:
  - RS message
  - RA message
- The protocol message exchange involves the following situations:
  - A host sends RS messages, and routers respond with RA messages.
  - Routers periodically send RA messages.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Router Discovery Process — Triggered by a Host's Request

When a host starts, it sends RS messages to all the routers on a local link, triggering them to respond with RA messages. After discovering routers on the local link, the host automatically configures a router as the default router. In addition, the host sets up the default routing table and prefix list as well as sets other configuration parameters.



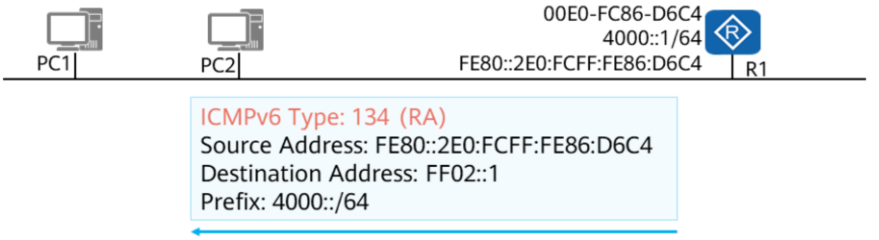
1. A host starts and sends RS messages to all the routers on a local link.
2. The routers respond with RA messages.

After the preceding process is complete, the host generates a default route with the next hop address being the link-local address of the default router.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Router Discovery Process — Triggered by Routers' Periodic RA Message Sending

- Routers periodically send RA messages. The interval for sending RA messages is a random value in a valid range. The default maximum and minimum intervals are 600s and 200s, respectively.
- The addresses of these RA messages must meet the following requirements:
  - Source Address: must be the link-local address of an outbound interface.
  - Destination Address: must be FF02::1.



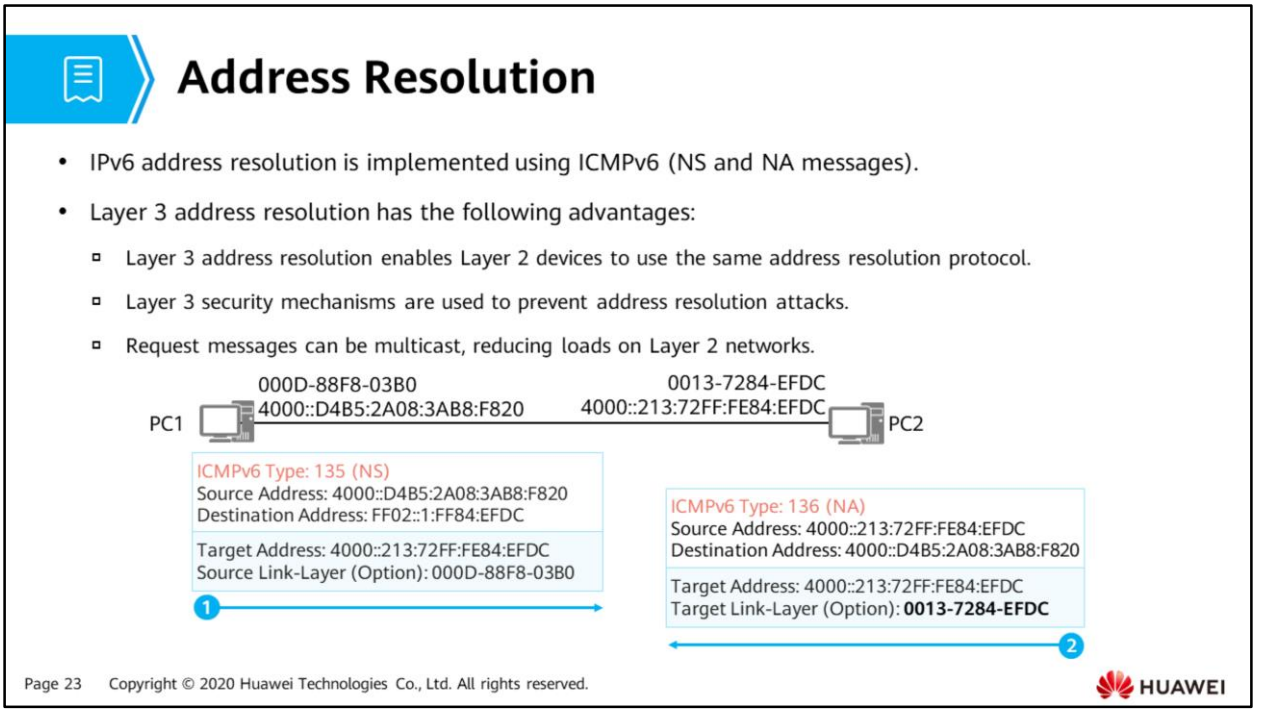
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The **ipv6 nd ra** { **max-interval** *maximum-interval* | **min-interval** *minimum-interval* } command configures an interval at which RA messages are sent.

Contents

1. ICMPv6 Introduction
2. **NDP Introduction**
   - NDP Overview
   - Router Discovery
   - **Address Resolution**
   - Neighbor State Tracing
   - DAD
   - Redirection

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- In this example, if PC1 wants to send a packet to PC2 but does not know PC2's link-layer address, the following protocol exchange process needs to be performed:
  - PC1 sends an NS message with the destination address being PC2's solicited-node multicast address FF02::1:FF84:EFDC. The Option field of the message carries PC1's link-layer address 000D-88F8-03B0 of PC1.
  - After listening to the NS message, PC2 checks that the destination address of the message is FF02::1:FF84:EFDC and determines that it is in the multicast group. Therefore, PC2 processes the message. In addition, PC2 updates its neighbor entries based on the source address and source link-layer address option of the NS message.
  - PC2 responds to the NS message with an NA message in which the Target Link-Layer option carries the link-layer address 0013-7284-EFDC.
  - After receiving the NA message, PC1 obtains PC2's link-layer address and creates a neighbor entry for the target node.
- After the preceding process is complete, PC1 and PC2 obtain each other's link-layer addresses and establish neighbor entries, which are similar to ARP entries in IPv4. PC1 and PC2 can then communicate.

Contents

1. ICMPv6 Introduction
2. **NDP Introduction**
   - NDP Overview
   - Router Discovery
   - Address Resolution
   - **Neighbor State Tracing**
   - DAD
   - Redirection

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

IPv6 Neighbor State Table

An IPv6 neighbor state table caches the mapping between IPv6 and MAC addresses. You can run the **display ipv6 neighbors** command to check an IPv6 neighbor state table.

```
<Huawei>display ipv6 neighbors
------------------------------------------------------------------------------
IPv6 Address     : 2001:DB8::2
Link-layer       : 00e0-fc23-26e3            State : REACH
Interface        : GE0/0/0                   Age   : 0
VLAN             : -                         CEVLAN: -
VPN name         :                           Is Router: TRUE
Secure FLAG      : UN-SECURE

IPv6 Address     : FE80::2E0:FCFF:FE23:26E3
Link-layer       : 00e0-fc23-26e3              State : REACH
Interface        : GE0/0/0                     Age   : 0
VLAN             : -                           CEVLAN: -
VPN name         :                             Is Router: TRUE
Secure FLAG      : UN-SECURE

------------------------------------------------------------------------------
Total: 2       Dynamic: 2       Static: 0
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

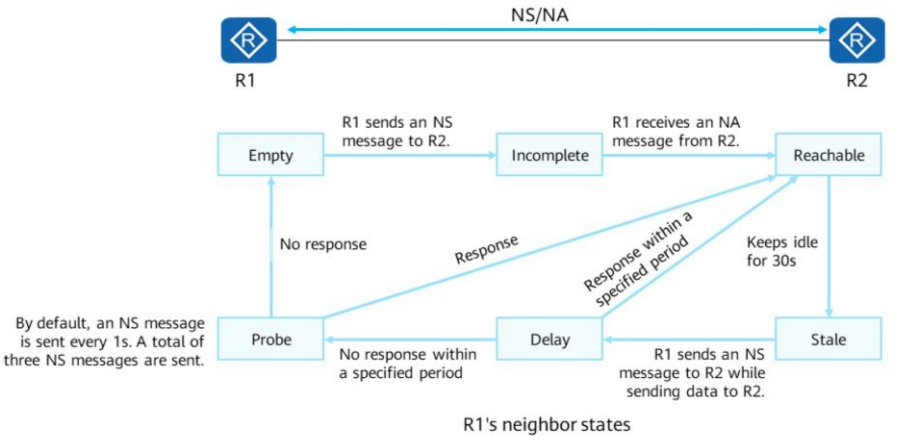
IPv6 Neighbor States

To monitor the state of each neighboring device, IPv6 nodes need to maintain a neighbor table. Five neighbor states are available: Incomplete, Reachable, Stale, Delay, and Probe.

State	Description
Incomplete	A neighbor is unreachable. This state indicates that a device is resolving a neighbor's link-layer address. If the device resolves the address successfully, it sets the neighbor state to Reachable.
Reachable	A neighbor is reachable. This state indicates that a neighbor is reachable within a specified period (30s by default). If the period expires but the neighbor entry is not used, this entry enters the Stale state.
Stale	Whether a neighbor is reachable is unknown. This state indicates that the neighbor entry is not used within a specified period (30s by default). A device does not check a neighbor's reachability unless it sends messages to the neighbor.
Delay	Whether a neighbor is reachable is unknown. A device has sent a message to a neighbor. If the device does not receive any response within a specified period, it sets the neighbor state to Probe.
Probe	Whether a neighbor is reachable is unknown. A device has sent an NS message to a neighbor to check whether the neighbor is reachable. If the device receives an NA message within a specified period, it sets the neighbor state to Reachable. Otherwise, it sets the neighbor state to Incomplete.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Neighbor State Transition



R1's neighbor states

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- R1 sends an NS message and generates a neighbor entry. In this case, the neighbor state is Incomplete.
- If R1 receives an NA message from R2, the neighbor state changes from Incomplete to Reachable. Otherwise, the neighbor state changes from Incomplete to Empty after a specified period.
- After the neighbor reachable time (30s by default) times out, the neighbor state changes from Reachable to Stale, indicating that the neighbor reachable state is unknown.
    - If R1 in the Reachable state receives an unsolicited NA message from R2, and the link-layer address of R2 carried in the message is different from that learned by R1, the neighbor state changes to Stale.
- If R1 in the Stale state sends data to R2, the neighbor state changes from Stale to Delay and R1 sends an NS message.
- After a specified period expires, the neighbor state changes from Delay to Probe. During this period, if R1 receives an NA message, the neighbor state changes from Delay to Reachable.
- R1 in the Probe state sends a specified number of unicast NS messages at a specified interval (1s by default). If R1 receives an NA message, the neighbor state changes from Probe to Reachable. Otherwise, the neighbor state changes to Empty.

Contents

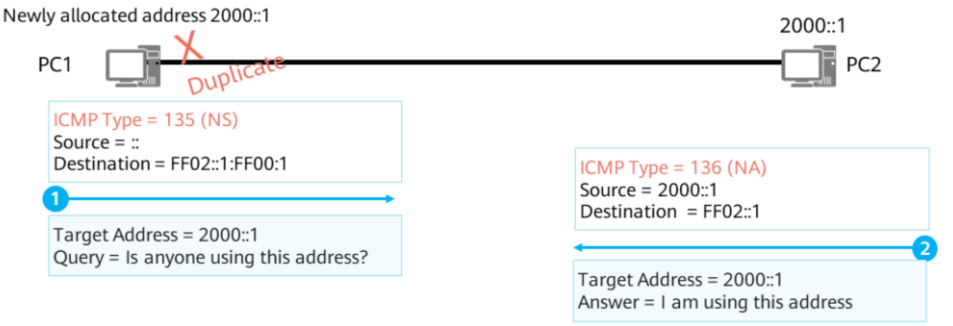
1. ICMPv6 Introduction
2. **NDP Introduction**
   - NDP Overview
   - Router Discovery
   - Address Resolution
   - Neighbor State Tracing
   - **DAD**
   - Redirection

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

DAD (1)

- DAD enables a node to detect whether an IPv6 address is being used by another node before using this address for an interface. This ensures that no two duplicate unicast addresses exist on a network.
- Before any unicast IPv6 address (including a link-local address) is used for an interface, DAD must be performed.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- DAD is a process in which a node checks whether an address to be used is being used by another node. Before configuring a unicast IPv6 address for an interface automatically, a node must ensure that this address is unique on a local link and is not used by another node. A node sends an NS message onto a local link by default. If a node does not receive an NA message within a specified period, it considers that the temporary unicast address is unique on a local link and can be allocated to an interface. Otherwise, it considers that this address is duplicate and cannot be used.

DAD (2)

- An address is called a tentative address before it passes DAD. An interface cannot use a tentative address for unicast communication.
- If DAD is performed for two nodes with the same address configured, the receiver does not use this address when receiving an NS message.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Special scenario: Two hosts are assigned the same IP address. Assume that both PC1 and PC2 want to use the address 2000::1. If PC1 sends an NS message first, PC2 does not send any NS message (or NA message) after receiving the NS message. Instead, PC2 stops using the address 2000::1 and waits for a new address to be generated in other modes. If both PC1 and PC2 receive an NS message, they do not use the address 2000::1.

Contents

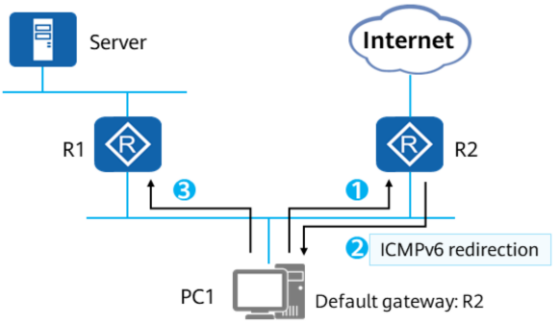
1. ICMPv6 Introduction
2. **NDP Introduction**
   - NDP Overview
   - Router Discovery
   - Address Resolution
   - Neighbor State Tracing
   - DAD
   - **Redirection**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Redirection

When a sender sends packets to a gateway, the gateway sends a Redirect message to instruct the sender to send packets to another optimal gateway.



1. If PC1 wants to send a packet to the server, it sends the packet to the default gateway R2.
2. After receiving the packet, R2 checks the packet's information and finds that the packet should be forwarded to the optimal gateway R1 that is on the same network segment as PC1. R2 then sends a Redirect message to PC1 to instruct PC1 to send the packet to R1.
3. After receiving the Redirect message, PC1 sends the packet to R1, which then forwards the packet to the server.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



Quiz

1. (Multiple) What are ICMPv6 message types?
   A. Error messages
   B. Informational messages
   C. Other messages
   D. Parameter messages
2. (Multiple) Which of the following messages are used to implement IPv6 address resolution?
   A. RS
   B. RA
   C. NS
   D. NA

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. A, B
2. C, D

Summary

- ICMPv6 is a basic IPv6 protocol that uses error or informational messages to report errors or information generated during packet processing. Error messages are used to report errors that occur during IPv6 packet forwarding, such as destination unreachable and timeout; informational messages are used to implement various functions, such as router discovery, DAD, and multicast member management.
- NDP is implemented using five ICMPv6 informational messages. RS and RA messages are used to implement router discovery, gateway discovery, and address autoconfiguration; NS and NA messages are used to resolve link-layer addresses of neighbors and detect duplicate addresses; Redirect messages are used to implement redirection.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



IPv6 Address Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- IPv6 is the core protocol of the next-generation Internet. It resolves many defects of IPv4, such as address scarcity, oversized routing table, and insufficient support for mobile devices.
- IPv6 supports address autoconfiguration for network nodes and implements plug-and-play, significantly simplifying network management.
- This course describes the fundamentals and implementation of IPv6 address autoconfiguration.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- On completion of this course, you will be able to:
  - Describe the fundamentals of IPv6 stateless address autoconfiguration.
  - Describe the fundamentals of DHCPv6 address autoconfiguration.
  - Perform IPv6 address autoconfiguration.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

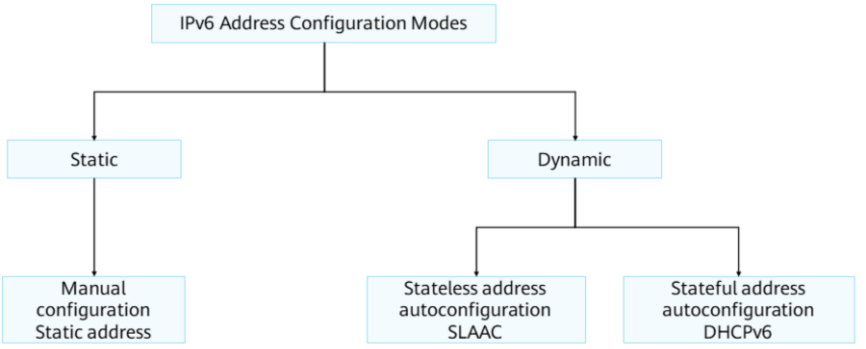
Contents

1. **IPv6 Address Configuration Modes**
2. IPv6 Stateless Address Autoconfiguration
3. DHCPv6
4. Implementation of IPv6 Address Autoconfiguration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Address Configuration Modes

IPv6 addresses can be configured statically or dynamically. Dynamic address configuration can be classified as stateless address autoconfiguration (SLAAC) or stateful address autoconfiguration.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Classification of IPv6 Address Autoconfiguration

Stateless address autoconfiguration

- IPv6 stateless address autoconfiguration does not require an IPv6 address allocation server to store and manage the status information of each node.
- Stateless address autoconfiguration is implemented based on the Neighbor Discovery Protocol (NDP).

Stateful address autoconfiguration

- IPv6 stateful address autoconfiguration requires an IPv6 address allocation server to store and manage the status information of each node.
- Stateful address autoconfiguration is implemented based on the Dynamic Host Configuration Protocol for IPv6 (DHCPv6).

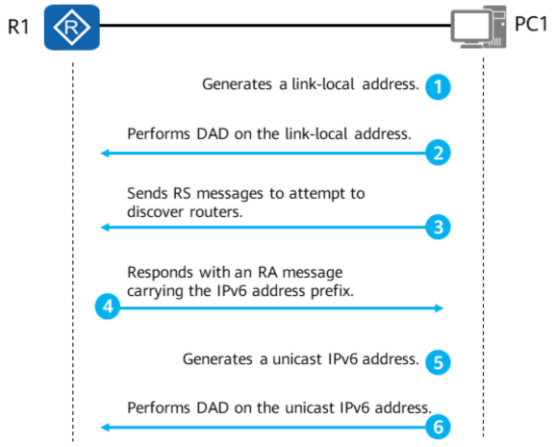
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. IPv6 Address Configuration Modes
2. **IPv6 Stateless Address Autoconfiguration**
3. DHCPv6
4. Implementation of IPv6 Address Autoconfiguration

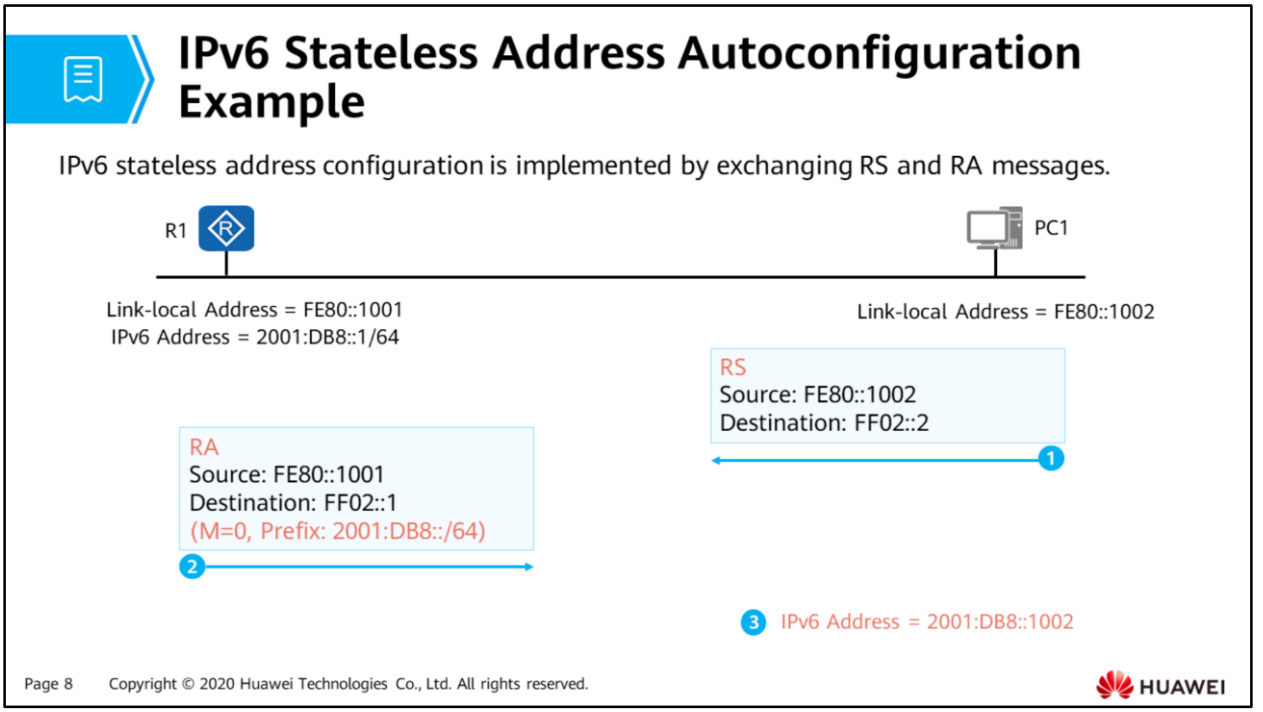
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Stateless Address Autoconfiguration Process



1. PC1 automatically generates a link-local address based on a local interface ID.
2. PC1 performs duplicate address detection (DAD) on the link-local address. If the address does not conflict with another, it can be used. In this case, PC1 has the IPv6 connection capability.
3. PC1 sends RS messages to attempt to discover IPv6 routers on the link.
4. R1 sends an RA message, which carries the IPv6 address prefix that can be used for stateless address autoconfiguration. (The router can also send RA messages even if it does not receive RS messages.)
5. PC1 parses the RA message to obtain the IPv6 address prefix, and generates a unicast IPv6 address based on this prefix and local interface ID.
6. PC1 performs DAD on the unicast IPv6 address. If no conflict is detected, PC1 uses this address.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- The process of NDP-based IPv6 stateless address autoconfiguration is as follows (DAD is omitted):
    1. PC1 generates the link-local address FE80::1002 and sends RS messages to all routers on the local link.
    2. R1 sends an RA message carrying a prefix for stateless address autoconfiguration. In this example, the prefix is 2001:DB8::/64.
    3. After receiving the RA message, PC1 generates the IPv6 address 2001:DB8::1002 based on the prefix and interface ID.

Flags Field in an RA Message

Internet Control Message Protocol V6

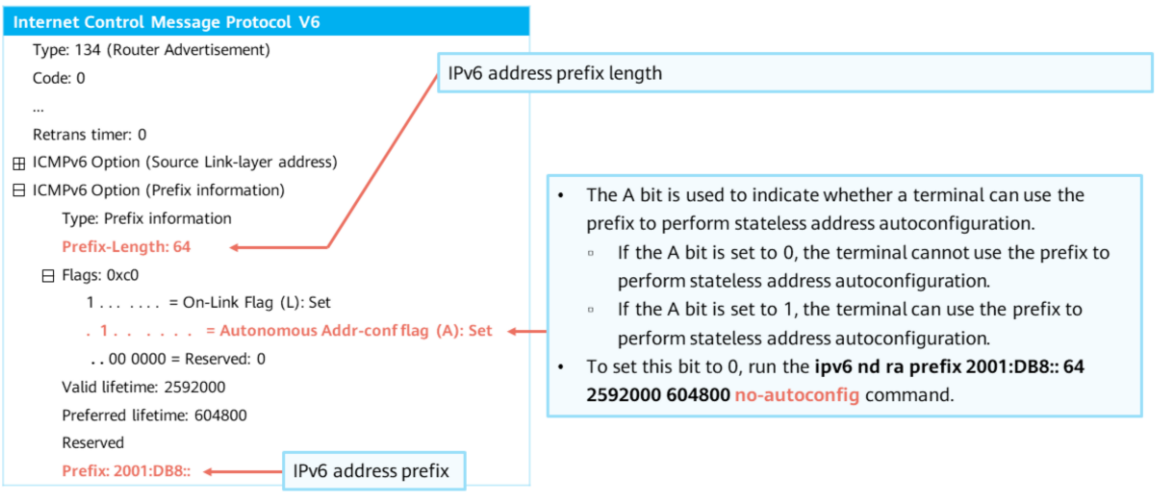
Type: 134 (Router Advertisement)
Code: 0
Checksum: 0x4a68 [Correct]
Cur Hop Limit: 64
⊟ Flags: 0x00
0 . . . . . = Not managed
. 0 . . . . . = Not Other
. . 0 . . . . = Not Home Agent
. . . 0 . . . . = Router Preference: Medium
. . . . . 0 . . = Not Proxied
Router Lifetime: 1800
Reachable time: 0
Retrans timer: 0
⊞ ICMPv6 Option (Source Link-layer address)
⊞ ICMPv6 Option (MTU)
⊞ ICMPv6 Option (Prefix information)

- M bit: managed address configuration flag
  - By default, the M bit is 0, and a terminal that receives an RA message uses the IPv6 address prefix contained in the message to construct a unicast address. That is, stateless address autoconfiguration is used.
  - If the M bit is set to 1 in an RA message, a terminal that receives the message uses DHCPv6 to obtain an address and other information. When the M bit is 1, the terminal can ignore the O bit. That is, the terminal uses DHCPv6 to obtain the address and other information (such as DNS).
- To set the M bit to 1, run the **ipv6 nd autoconfig managed-address-flag** command.

- O bit: other stateful configuration flag
  - By default, the O bit is 0, and a terminal that receives an RA message performs stateless address autoconfiguration. That is, a routing device advertises other configuration information except an IPv6 address to the terminal through the RA message.
  - If the O bit is set to 1 in an RA message, a terminal that receives the message uses DHCPv6 to obtain other parameters except an address.
- To set the O bit to 1, run the **ipv6 nd autoconfig other-flag** command.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Optional Information in an RA Message: Address Prefix



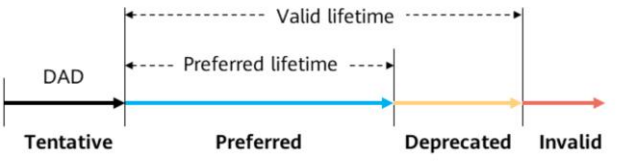
- The A bit is used to indicate whether a terminal can use the prefix to perform stateless address autoconfiguration.
  - If the A bit is set to 0, the terminal cannot use the prefix to perform stateless address autoconfiguration.
  - If the A bit is set to 1, the terminal can use the prefix to perform stateless address autoconfiguration.
- To set this bit to 0, run the **ipv6 nd ra prefix 2001:DB8:: 64 2592000 604800** no-autoconfig command.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Optional Information in an RA Message: Lifetime

```
Internet Control Message Protocol V6
    Type: 134 (Router Advertisement)
    Code: 0
    ...
    Retrans timer: 0
⊞ ICMPv6 Option (Source Link-layer address)
⊟ ICMPv6 Option (Prefix information)
        Type: Prefix information
        Prefix-Length: 64
    ⊟ Flags: 0xc0
            1... .... = On-Link Flag (L): Set
            .1.. .... = Autonomous Addr-conf flag (A): Set
            ..00 0000 = Reserved: 0
        Valid lifetime: 2592000
        Preferred lifetime: 604800
        Reserved
        Prefix: 2001:DB8::
```

- After a terminal obtains a prefix and generating a unicast IPv6 address, the address enters the Tentative state. After passing DAD, the address enters the Preferred state and remains in this state within the preferred lifetime.
- In the Preferred state, the terminal can send and receive packets normally. After the preferred lifetime expires, the address enters the Deprecated state and remains in this state within the valid lifetime.
- In the Deprecated state, the address is still valid and can be used for existing connections, but cannot be used to establish new connections.
- After the valid lifetime expires, the address enters the Invalid state, indicating that the address cannot be used.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Contents

1. IPv6 Address Configuration Modes
2. IPv6 Stateless Address Autoconfiguration
3. **DHCPv6**
4. Implementation of IPv6 Address Autoconfiguration

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

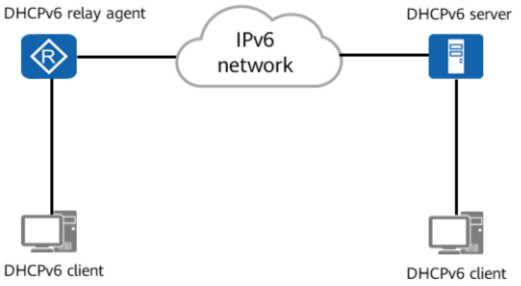
HUAWEI

DHCPv6 Overview

- DHCPv6 is designed for IPv6 addressing. It can be used to allocate IPv6 prefixes, IPv6 addresses, and other network configuration parameters to clients. In addition, DHCPv6 can record the information allocated to clients, facilitating network management.
- DHCPv6 address allocation can be implemented in any of the following modes:
  - DHCPv6 stateful autoconfiguration: A DHCPv6 server automatically configures IPv6 addresses/prefixes and other network configuration parameters, such as DNS, NIS, and SNTP server addresses.
  - DHCPv6 stateless autoconfiguration: Host IPv6 addresses are automatically generated through RA messages. A DHCPv6 server allocates other configuration parameters such as the DNS server address except for IPv6 addresses.
  - DHCPv6 prefix delegation (PD) autoconfiguration: A downstream device requests IPv6 address prefixes from an upstream device, and the upstream device allocates appropriate prefixes to the downstream device. In this way, you do not need to configure IPv6 address prefixes for user-side links on the downstream device. The downstream device divides the obtained prefix (which is typically less than 64 bits long) into subnet segments with 64-bit prefixes and sends an RA message on the link that IPv6 hosts directly connect to. This enables IPv6 hosts to automatically configure IPv6 addresses, implementing hierarchical address deployment.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

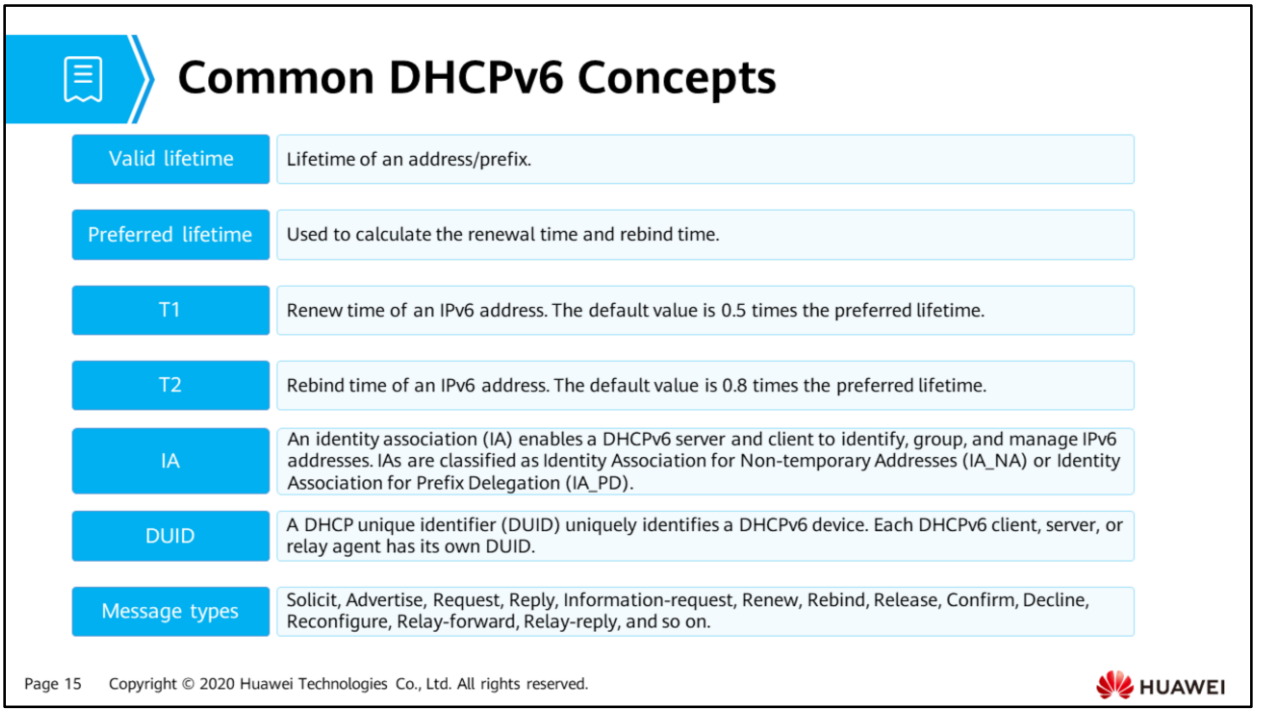
DHCPv6 Network Composition



The DHCPv6 architecture involves the following roles:

- DHCPv6 client: exchanges DHCPv6 messages with a DHCPv6 server to obtain an IPv6 address/prefix and other network configuration parameters.
- DHCPv6 server: processes requests for address allocation, lease renewal, and release from a DHCPv6 client or a DHCPv6 relay agent, and allocates IPv6 addresses/prefixes and other network configuration parameters to the client.
- DHCPv6 relay agent: relays DHCPv6 messages between a DHCPv6 client and server to help the client configure its address.

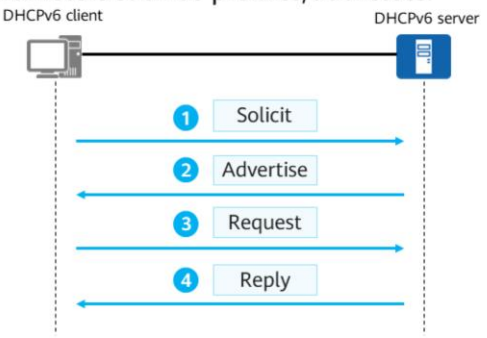
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Valid lifetime: lifetime of an address/prefix. After an address/prefix expires, all the users who use it are logged out. The valid lifetime must not be shorter than 3 hours or the preferred lifetime.
- Preferred lifetime: used to calculate the renew time and rebind time. The preferred lifetime must not be shorter than 2 hours.

DHCPv6 Stateful Autoconfiguration — Four-Message Exchange

In the four-message exchange, a DHCPv6 client and server exchange four messages to obtain parameters such as prefixes/addresses.

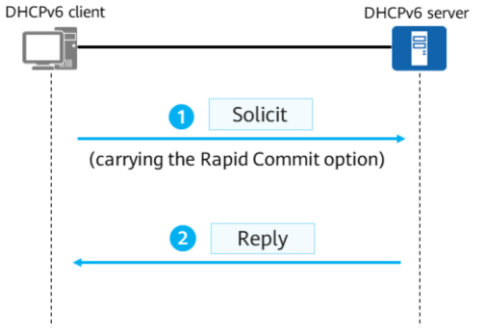


1. The DHCPv6 client sends a Solicit message to request the DHCPv6 server to allocate an IPv6 address/prefix and network configuration parameters.
2. The DHCPv6 server sends an Advertise message to the DHCPv6 client, notifying the client of the IPv6 addresses/prefixes and network configuration parameters that can be allocated.
3. If the DHCPv6 client receives Advertise messages from multiple DHCPv6 servers, it selects a server based on the sequence in which the messages are received and the server priorities. The client then sends a Request message to the server, requesting an IPv6 address/prefix and network configuration parameters.
4. The DHCPv6 server responds with a Reply message, which contains the allocated IPv6 address/prefix and network configuration parameters.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCPv6 Stateful Autoconfiguration — Two-Message Exchange

A DHCPv6 client sends a Solicit message that carries the Rapid Commit option, indicating that the client expects the server to quickly allocate an address/prefix and network configuration parameters.



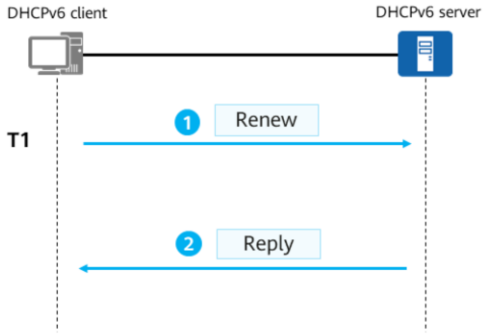
1. The DHCPv6 client sends a Solicit message carrying the Rapid Commit option.
2. After receiving the Solicit message, the DHCPv6 server checks whether it supports fast address allocation. If so, the server returns a Reply message carrying the allocated IPv6 address/prefix and other network configuration parameters. If the DHCPv6 server does not support fast address allocation, the four-message exchange is used.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The two-message exchange improves the efficiency of DHCPv6 address allocation, but it is applicable when only one DHCPv6 server exists on a network. On a network with multiple DHCPv6 servers, these servers can allocate IPv6 addresses/prefixes and other configuration parameters to a DHCPv6 client. However, a client can use only the IPv6 address/prefix and configuration parameters allocated by a DHCPv6 server.

Address/Prefix Lease Renewal (1)

An IPv6 address/prefix allocated by a DHCPv6 server has a valid lifetime. After the valid lifetime of an address/prefix expires, a DHCPv6 client cannot use the address/prefix. Therefore, if the DHCPv6 client wants to continue to use the address/prefix before the valid lifetime expires, the lease of the address/prefix needs to be renewed.

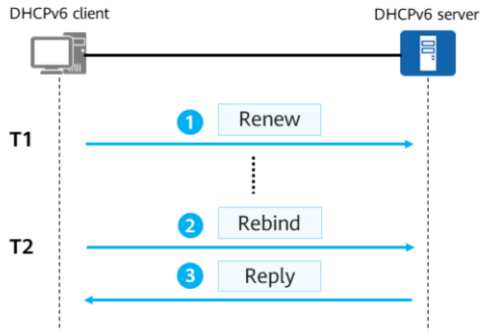


1. The DHCPv6 client sends a Renew message at T1 (half the preferred lifetime by default) to request address/prefix lease renewal.
2. If the DHCPv6 client can continue to use the address/prefix, the DHCPv6 server responds with a Reply message, indicating that the address/prefix lease is renewed successfully. Otherwise, the DHCPv6 server responds with a Reply message, indicating that the address/prefix lease fails to be renewed.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Address/Prefix Lease Renewal (2)

If no DHCPv6 server responds to a Renew message sent by a DHCPv6 client at T1, the client multicasts a Rebind message to all DHCPv6 servers at T2 (0.8 times the preferred lifetime by default) to request lease renewal.

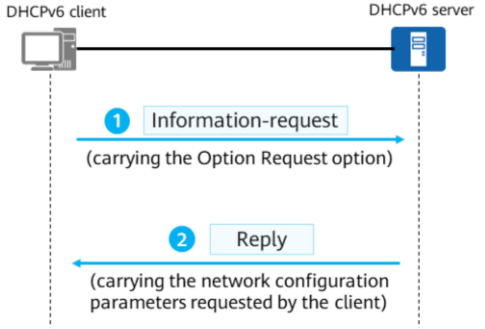


1. The DHCPv6 client sends a Renew message to request lease renewal at T1, but it does not receive a reply message from DHCPv6 servers.
2. The DHCPv6 client multicasts a Rebind message to all DHCPv6 servers at T2 to request lease renewal.
3. If the DHCPv6 client can continue to use the address/prefix, the DHCPv6 server responds with a Reply message, indicating that the address/prefix lease is renewed successfully. Otherwise, the DHCPv6 server responds with a Reply message, indicating that the address/prefix lease fails to be renewed.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCPv6 Stateless Autoconfiguration

DHCPv6 stateless autoconfiguration enables a DHCPv6 server to allocate network configuration parameters except IPv6 addresses/prefixes to clients that already have the addresses/prefixes.



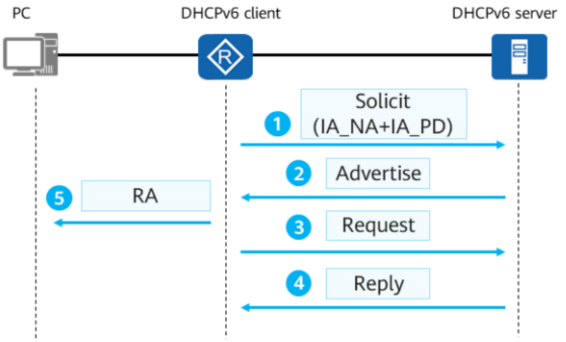
1. The DHCPv6 client multicasts an Information-request message with the Option Request option to the DHCPv6 server. This option specifies the configuration parameters that the client needs to obtain from the server.
2. After receiving the message, the server unicasts a Reply message containing the allocated network configuration parameters to the client. The client checks the allocated network configuration parameters in the Reply message. If these parameters are the same as those in the Information-request message, the client performs network configuration based on the allocated network configuration parameters. Otherwise, the client ignores the allocated parameters.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- After a host generates a link-local address and detects no address conflict, it initiates a router discovery process. Specifically, the host sends an RS message, and the router replies with an RA message. If the M bit is 0 and the O bit is 1 in the RA message, the host obtains other configuration parameters except addresses/prefixes, such as DNS, SIP, and SNTP server configuration parameters, through DHCPv6 stateless autoconfiguration.

DHCPv6 PD Autoconfiguration

On a hierarchical network, IPv6 addresses are generally configured manually, which has poor extensibility and is not beneficial for uniform IPv6 address planning and management. DHCPv6 PD can resolve this issue.



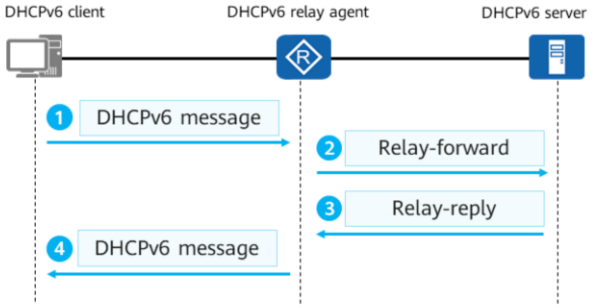
1. The DHCPv6 client sends a Solicit message, requesting an IA_NA address and IA_PD prefix from the DHCPv6 server.
2. The DHCPv6 server sends an Advertise message to notify the client of the allocated IPv6 address/prefix.
3. If multiple servers send Advertise messages to the client, the client selects the Advertise message from the server with the highest priority based on parameters such as the server priorities in the Advertise messages (if the server priorities are the same, the client selects the Advertise message that carries the required configuration parameters) and sends a Request message to the server, requesting the server to allocate an IPv6 address/prefix.
4. The DHCPv6 server responds with a Reply message, which carries the allocated IPv6 address/prefix.
5. After receiving the PD prefix, the DHCPv6 client exchanges RS/RA messages with the PC and delivers the obtained PD prefix to the PC through an RA message.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- DHCPv6 PD applies to a scenario where a router (for example, the DHCPv6 client in this example) needs to allocate prefixes to its connected IPv6 hosts to implement automatic address configuration for the hosts. In this way, the hierarchical layout of the entire IPv6 network is implemented.
- In Step 1, the DHCPv6 client requests the DHCPv6 server to allocate an IA_NA address and an IA_PD prefix, which are the address allocated to the client's WAN interface and the prefix allocated to the client's LAN side, respectively.

Working Process of a DHCPv6 Relay Agent

If a DHCPv6 server and client are on different network segments, a DHCPv6 relay agent is required to obtain IPv6 addresses/prefixes and other network configuration parameters.

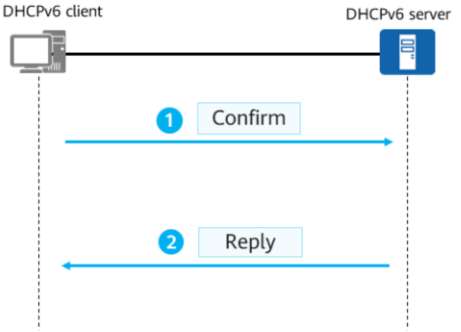


1. The DHCPv6 client sends a Request message to all DHCPv6 servers and the multicast address FF02::1:2 of the DHCPv6 relay agent.
2. After receiving the message, the DHCPv6 relay agent encapsulates it in the Relay Message option of a Relay-forward message and sends the Relay-forward message to the DHCPv6 server or next-hop relay agent.
3. The DHCPv6 server obtains the client's request from the Relay-forward message and selects an IPv6 address and other configuration parameters. The server then sends a Relay-reply message to the DHCPv6 relay agent.
4. The DHCPv6 relay agent obtains the server's reply message from the Relay-reply message and forwards the reply message to the DHCPv6 client.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCPv6 Address Confirmation

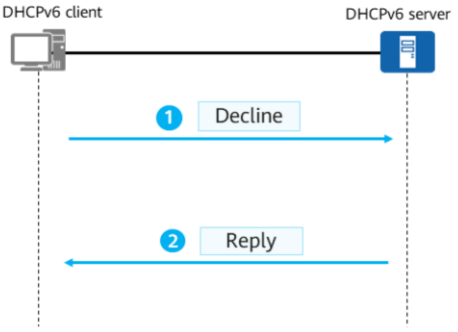
In some scenarios, for example, when a DHCPv6 client is powered off, goes offline, or roams, it sends a Confirm message to confirm whether its IPv6 address is available. If the client confirms that the address is valid, the server responds. If the server does not respond, the client needs to restart the address application process.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCPv6 DAD

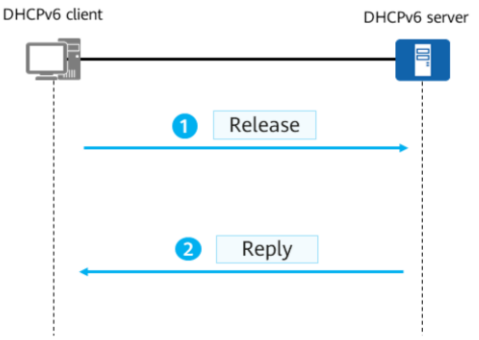
A DHCPv6 client performs DAD on an allocated address before using the address. If DAD detects an address conflict, the client sends a Decline message to notify the server that it does not use the address.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCPv6 Address Release

When a DHCPv6 client does not need to use an address, it sends a Release message to a server, requesting address release.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCPv6 Message Summary

Message Type	Function
Solicit	A DHCPv6 client sends a Solicit message to request a DHCPv6 server to allocate an IPv6 address/prefix and network configuration parameters.
Advertise	A DHCPv6 server sends an Advertise message to a DHCPv6 client, notifying the client of the IPv6 addresses/prefixes and network configuration parameters that can be allocated.
Request	If a DHCPv6 client receives Advertise messages from multiple DHCPv6 servers, it selects a server based on the sequence in which the messages are received and the server priorities. The client then sends a Request message to the server, requesting an IPv6 address/prefix and network configuration parameters.
Reply	A DHCPv6 server responds with a Reply message, which contains the allocated IPv6 address/prefix and network configuration parameters.
Information-request	A DHCPv6 client multicasts an Information-request message with the Option Request option to DHCPv6 servers. This option specifies the configuration parameters that the client needs to obtain from a DHCPv6 server.
Renew	When the address/prefix lease reaches T1, a DHCPv6 client unicasts a Renew message to the DHCPv6 server that allocates the address/prefix to renew the address/prefix lease.
Rebind	If a DHCPv6 client sends a Renew message to renew the lease at T1 but does not receive any reply message from the corresponding DHCPv6 server, the client multicasts a Rebind message to all DHCPv6 servers at T2 to request lease renewal.
Confirm	In some scenarios, for example, when a DHCPv6 client is powered off, goes offline, or roams, it sends a Confirm message to confirm whether its IPv6 address is available.
Decline	When a DHCPv6 client detects an address conflict, it sends a Decline message to a DHCPv6 server.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Address Autoconfiguration Comparison

Address Configuration	Based on DHCPv6	Based on NDP
Address management	Stateful. A server stores information about address or prefix allocation and release.	Stateless. A server does not store information about user address allocation.
Deployment value	128-bit addresses and prefixes with different lengths can be allocated, resulting in high extensibility.	Only 64-bit prefixes are supported, resulting in poor extensibility.
Implementation	Complex configuration.	Simple configuration.
Security	DHCPv6 works at the application layer and provides high security.	Low security.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. IPv6 Address Configuration Modes
2. IPv6 Stateless Address Autoconfiguration
3. DHCPv6
4. **Implementation of IPv6 Address Autoconfiguration**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Basic IPv6 Configurations (1)

1. Enable a device to forward IPv6 unicast packets.

   [Huawei] **ipv6**

2. Enable the interface to forward IPv6 unicast packets.

   [Huawei-GigabitEthernet0/0/0] **ipv6 enable**

3. Manually configure a link-local address for the interface.

   [Huawei-GigabitEthernet0/0/0] **ipv6 address** *ipv6-address* **link-local**

4. Automatically configure a link-local address for the interface.

   [Huawei-GigabitEthernet0/0/0] **ipv6 address auto link-local**

5. Manually configure a global unicast address for the interface.

   [Huawei-GigabitEthernet0/0/0] **ipv6 address** { *ipv6-address prefix-length* | *ipv6-address*/*prefix-length* }

6. Automatically configure a global unicast address for the interface.

   [Huawei-GigabitEthernet0/0/0] **ipv6 address auto** { **global** | **dhcp** }

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic IPv6 Configurations (2)

7. Display IPv6 information on the interface.

```
[Huawei] display ipv6 interface [ interface-type interface-number | brief ]
```

8. Display neighbor entry information.

```
[Huawei] display ipv6 neighbors
```

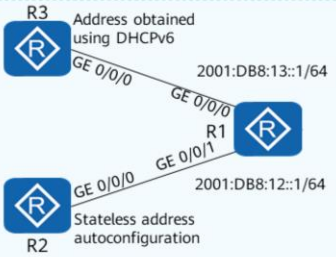
9. Enable the interface to send RA messages.

```
[Huawei-GigabitEthernet0/0/0] undo ipv6 nd ra halt
```

By default, Huawei router interfaces do not send ICMPv6 RA messages. In such a case, hosts on the network cannot periodically receive information about updated IPv6 address prefixes. To enable an interface to periodically send RA messages that carry both the IPv6 address prefix and stateful address autoconfiguration flag to a host, run the **undo ipv6 nd ra halt** command.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Address Configuration Examples — Static Configuration



Configuration requirements:

- Manually configure IPv6 addresses for R1's GE 0/0/0 and GE 0/0/1.
- Configure R2's GE 0/0/0 to obtain an IPv6 address through stateless address autoconfiguration.
- Configure R3's GE 0/0/0 to obtain an IPv6 address through DHCPv6.

1. Enable IPv6 on R1, R2, and R3, and configure interfaces to automatically generate link-local addresses (use R1 as an example).

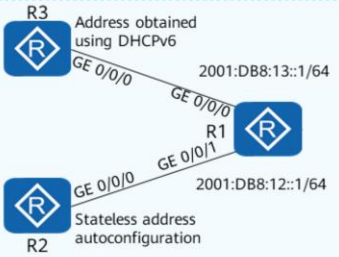
```
[R1]ipv6
[R1]interface GigabitEthernet 0/0/0
[R1-GigabitEthernet0/0/0]ipv6 enable
[R1-GigabitEthernet0/0/0]ipv6 address auto link-local
```

2. Configure global unicast IPv6 address for R1's interfaces.

```
[R1]interface GigabitEthernet 0/0/0
[R1-GigabitEthernet0/0/0]ipv6 address 2001:DB8:13::1 64
[R1-GigabitEthernet0/0/0]interface GigabitEthernet 0/0/1
[R1-GigabitEthernet0/0/1]ipv6 address  2001:DB8:12::1 64
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

IPv6 Address Configuration Examples — Stateless Address Autoconfiguration



Configuration requirements:

- Manually configure IPv6 addresses for R1's GE 0/0/0 and GE 0/0/1.
- Configure R2's GE 0/0/0 to obtain an IPv6 address through stateless address autoconfiguration.
- Configure R3's GE 0/0/0 to obtain an IPv6 address through DHCPv6.

1. Enable RA message advertisement on R1.

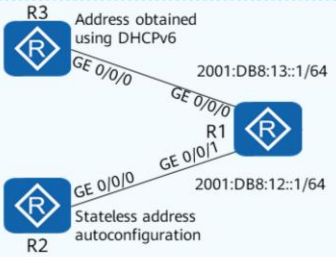
```
[R1]interface GigabitEthernet 0/0/1
[R1-GigabitEthernet0/0/1]undo ipv6 nd ra halt
```

2. Configure R2's GE 0/0/0 to automatically obtain an IPv6 address.

```
[R2]interface GigabitEthernet 0/0/0
[R2-GigabitEthernet0/0/0]ipv6 address auto global
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Address Configuration Examples — DHCPv6



Configuration requirements:

- Manually configure IPv6 addresses for R1's GE 0/0/0 and GE 0/0/1.
- Configure R2's GE 0/0/0 to obtain an IPv6 address through stateless address autoconfiguration.
- Configure R3's GE 0/0/0 to obtain an IPv6 address through DHCPv6.

1. Configure the DHCPv6 server function on R1.

```
[R1]dhcp enable
[R1]dhcpv6 pool pool1
[R1-dhcpv6-pool-pool1]address prefix 2001:DB8:13::/64
[R1-dhcpv6-pool-pool1]excluded-address 2001:DB8:13::1
[R1]interface GigabitEthernet 0/0/0
[R1-GigabitEthernet0/0/0]dhcpv6 server pool1
```

2. Configure R3's GE 0/0/0 to obtain an IPv6 address through DHCPv6.

```
[R3]dhcp enable
[R3]interface GigabitEthernet 0/0/0
[R3-GigabitEthernet0/0/0]ipv6 address auto dhcp
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IPv6 Address Autoconfiguration — Verification

Verify the IPv6 address obtained by R2 through stateless address autoconfiguration.

```
[R2]display ipv6 interface GigabitEthernet 0/0/0
GigabitEthernet0/0/0 current state : UP
IPv6 protocol current state : UP
IPv6 is enabled, link-local address is FE80::2E0:FCFF:FEAD:675F
  Global unicast address(es):
   2001:DB8:12::2E0:FCFF:FEAD:675F,
    subnet is 2001:DB8:12::/64 [SLAAC]
  Joined group address(es):
    FF02::1:FFAD:675F
    FF02::2
    FF02::1
  MTU is 1500 bytes
  ND DAD is enabled, number of DAD attempts: 1
  ND reachable time is 30000 milliseconds
  ND retransmit interval is 1000 milliseconds
  Hosts use stateless autoconfig for addresses
```

Run the **display dhcpv6 client** command on R3 to check address allocation information.

```
[R3]display dhcpv6 client
GigabitEthernet0/0/0 is in stateful DHCPv6 client mode.
State is BOUND.
Preferred server DUID   : 0003000100E0FC401312
  Reachable via address : FE80::2E0:FCFF:FE40:1312
IA NA IA ID 0x00000031 T1 43200 T2 69120
     Obtained           : 2020-06-17 09:49:58
      Renews            : 2020-06-17 21:49:58
      Rebinds           : 2020-06-18 05:01:58
      Address           : 2001:DB8:13::3
    Lifetime valid 172800 seconds, preferred 86400 seconds
    Expires at 2020-06-19 09:49:58(172159 seconds left)
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Single) Which of the following statements about NDP-based stateless address autoconfiguration is true?
   A. NDP-based stateless address autoconfiguration can be used to allocate 128-bit IPv6 addresses to hosts.
   B. NDP-based stateless address autoconfiguration can be used to allocate IPv6 prefixes or addresses to hosts.
   C. NDP-based stateless address autoconfiguration can be used to allocate only IPv6 prefixes to hosts.
   D. NDP-based stateless address autoconfiguration can be used to allocate DNS information to hosts.
2. (Multiple) Which of the following processes use Reply messages in DHCPv6?
   A. DHCPv6 information obtaining
   B. DHCPv6 prefix allocation's two-message exchange
   C. DHCPv6 address allocation's four-message exchange
   D. DHCPv6 address release
   E. DHCPv6 address lease renewal

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. D
2. A, B, C, D, and E

Summary

- IPv6 addresses can be configured statically or dynamically. Usually, the interconnection IPv6 addresses of network devices and IPv6 addresses of loopback interfaces are statically configured, whereas terminals are dynamically configured.
- This course mainly describes NDP-based stateless address autoconfiguration and DHCPv6-based stateful address autoconfiguration. Stateless address autoconfiguration can be used to allocate only 64-bit prefixes, resulting in poor extensibility as well as address management and control. Stateful address autoconfiguration provides more functions, such as PD prefix allocation.
- The M and O bits in an RA message determine an IPv6 address autoconfiguration mode used by a client. Determine an appropriate address autoconfiguration mode based on an actual scenario.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Huawei Firewall Technology

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- The word "firewall" is first used in the construction field. The function of a firewall is to isolate the fire, preventing the fire from spreading from one area to another. In the communications field, a firewall is usually deployed to logically isolate networks for some purposes. The firewall defends against various attacks on the network to ensure transmission of communication packets.
- This course describes what is a firewall in the communications field, why a firewall is required, and the working mechanism and configurations of the firewall.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this source, you will be able to:
  - Describe the definition, type, and development history of firewalls.
  - Distinguish firewalls from routers and switches.
  - Describe basic firewall concepts, such as the security zone, security policy, session table, and server map.
  - Perform basic firewall configurations.

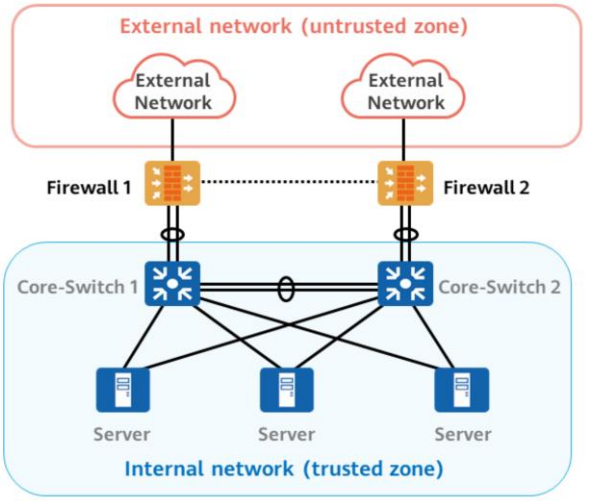
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Introduction to Firewalls**
2. Basic Firewall Concepts
3. Basic Firewall Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Why Do We Need Firewalls?



- Security is required everywhere. Routers and switches construct an interconnected network, which brings convenience and security risks.
- For example, enterprises have the following security requirements at a network border:
  - External network security isolation
  - Internal network security control
  - Content filtering
  - Intrusion prevention
  - Antivirus

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

What Is a Firewall?

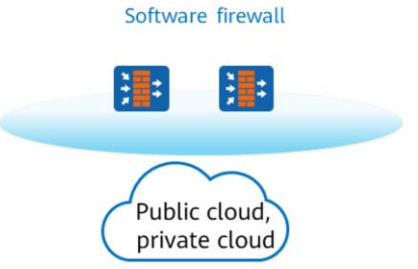
- In the communications field, a firewall is a security device. It is used to protect a network area against attacks and intrusions from another network area. It is usually deployed at a network border, such as the enterprise Internet egress, enterprise internal service border, and data center border.
- Firewalls are classified into modular firewalls, fixed firewalls, and software firewalls. Firewalls can be flexibly deployed locally and on the cloud.



Modular firewall



Fixed firewall

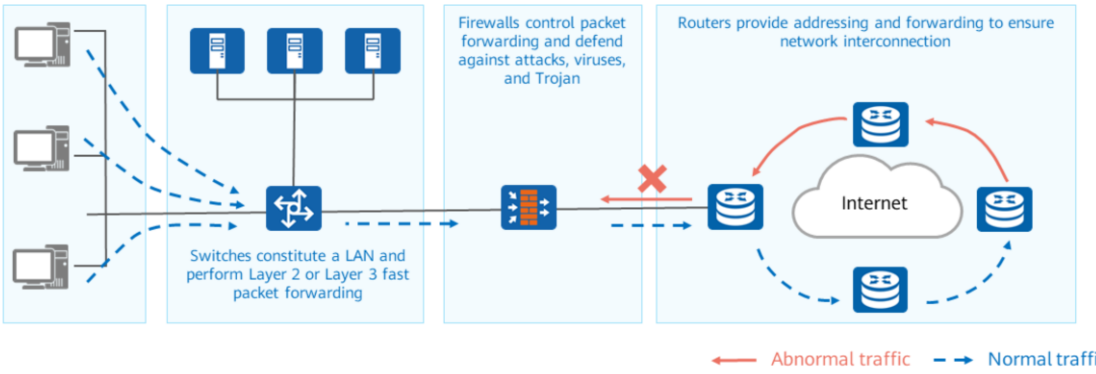


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Firewalls have other models, such as desktop firewalls (a type of fixed firewalls). Desktop firewalls apply to small enterprises, industry branches, and chain business organizations. Huawei fixed firewalls support both the traditional and cloud management modes. In cloud management mode, the cloud manages secure access of branches in a unified manner, and supports plug-and-play devices, automatic service configuration, visualized O&M, and network big data analysis.
- This course focuses on modular and fixed physical firewalls, and does not describe desktop firewalls and software firewalls.

Comparison Between Firewalls and Switches and Routers

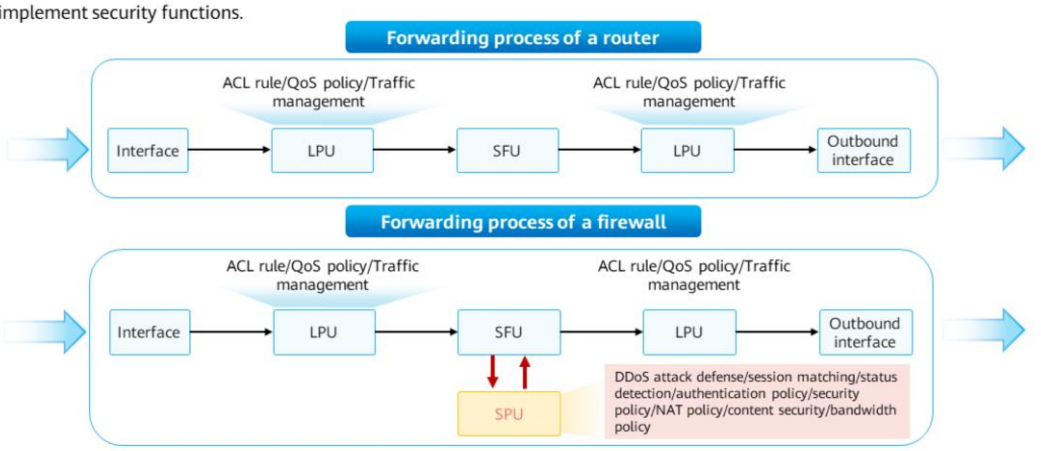
- For example, on a campus network, switches are used to connect terminals and summarize internal routes to build a LAN for internal communication.
- Routers are used for route distribution, addressing, and forwarding to construct an external connection network.
- Firewalls are used to implement traffic control and security protection and to distinguish and isolate different security zones.



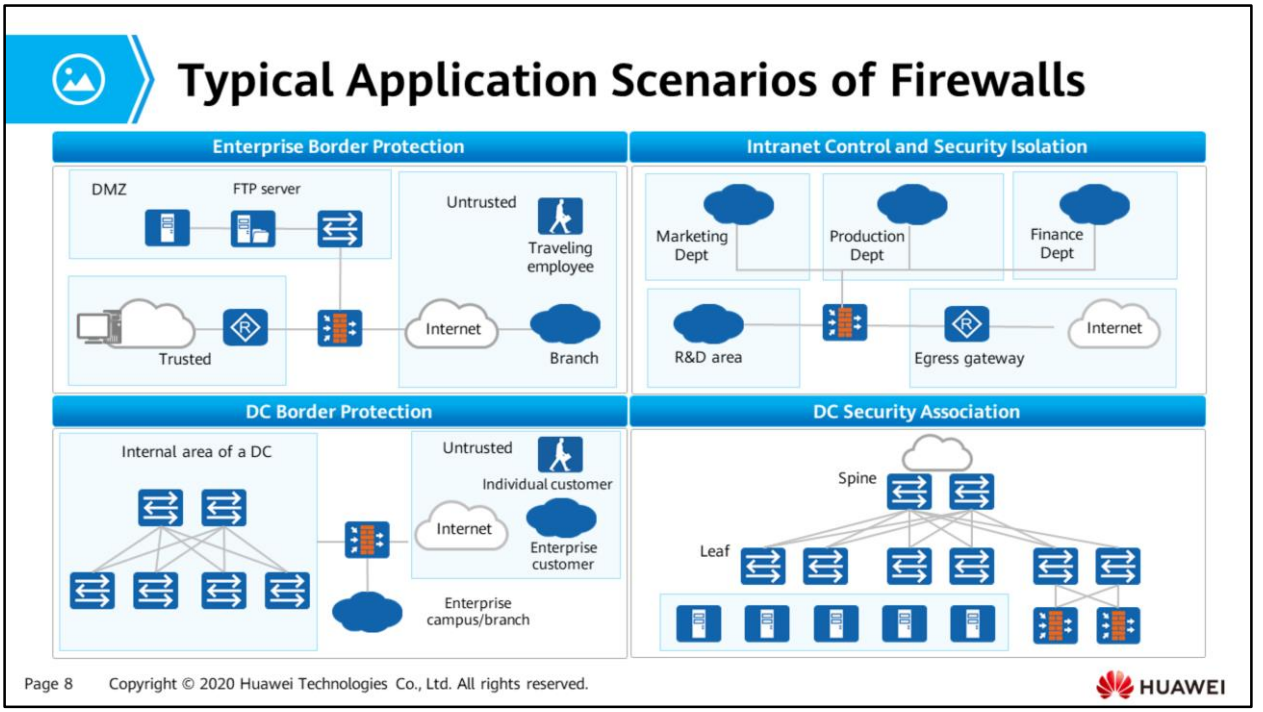
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Comparison Between Forwarding Processes of a Firewall and a Router

- The forwarding process of a firewall is more complex than that of a router. A modular firewall is used as an example. In addition to interfaces, line processing units (LPUs), and switch fabric units (SFUs), a modular firewall has service processing units (SPUs) to implement security functions.



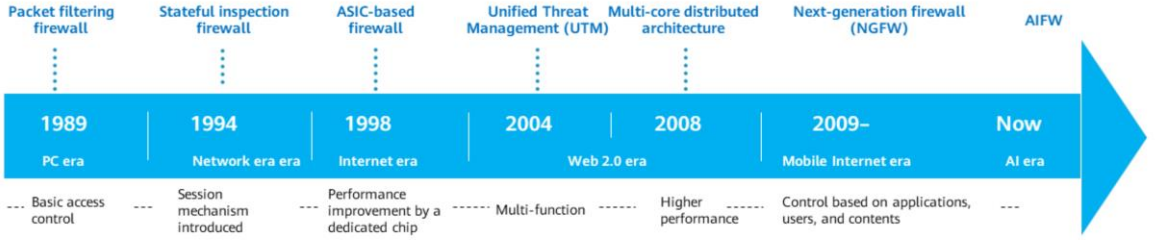
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI



- A Demilitarized Zone (DMZ) is originally a military term, referring to a partially controlled area between a military control area and a public area. A DMZ configured on a firewall is logically and physically separated from internal and external networks. In an enterprise, it is usually used to accommodate servers.
- Data center networks often use the spine-leaf architecture. Spine nodes forward traffic at a high speed, and leaf nodes connect to servers, firewalls, or other devices. Spine and leaf nodes are fully meshed at Layer 3.

Firewall Development History

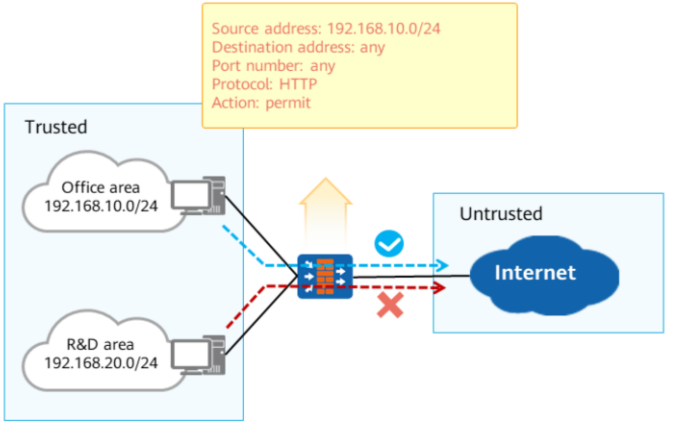
- Throughout the firewall development history, the firewall has gone through the process from low-level to high-level and from simple functions to complex functions. Development of network technologies and emergence of new requirements promote the development of firewalls.
- Firewalls develop from the packet filtering firewall, status detection firewall, UTM, and NGFW to AIFW, with the following characteristics:
  - More refined access control
  - More powerful protection capability
  - Higher performance



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Packet Filtering Firewall

- A packet filtering firewall checks each packet based on the 5-tuple and forwards or discards packets based on the configured security policies.
- The packet filtering firewall uses access control lists (ACLs) to filter data packets.



- In this example, firewall security policies are configured to allow only IP addresses in the office area to access the Internet.
- Problems of the packet filtering firewall:
  - Per-packet detection has low performance.
  - ACL rules cannot meet dynamic requirements.
  - Application-layer data is not checked.
  - The packet filtering firewall does not provide packet association analysis, so spoofing may easily occur.

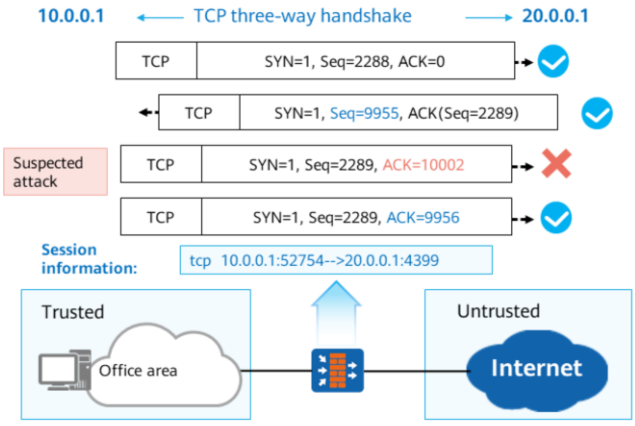
Question: How to solve problems of the packet filtering firewall?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A packet filtering firewall filters packets based on information such as the source/destination IP address, source/destination port number, IP identifier, and packet transmission direction in the packets.
- The packet filtering firewall is simple in design, easy to implement, and cost-effective.
- The disadvantages of the packet filtering firewall are as follows:
  - With the increase of ACL complexity and length, filtering performance decreases exponentially.
  - Static ACL rules cannot meet dynamic security requirements.
  - The packet filtering firewall does not check the session status or analyze data, which makes it easy for attackers to escape. For example, an attacker sets the IP address of the host to an IP address permitted by a packet filtering firewall. In this way, packets from this host can easily pass through the packet filtering firewall.

Stateful Inspection Firewall

- Stateful inspection firewalls are introduced based on packet filtering technology. It considers correlation between packets and detects the connection status but not a single packet.



- In this example, the stateful inspection firewall detects that a TCP connection is established between hosts at 10.0.0.1 and 20.0.0.1 and generates session information. The third handshake packet does not match any session entry and is discarded.

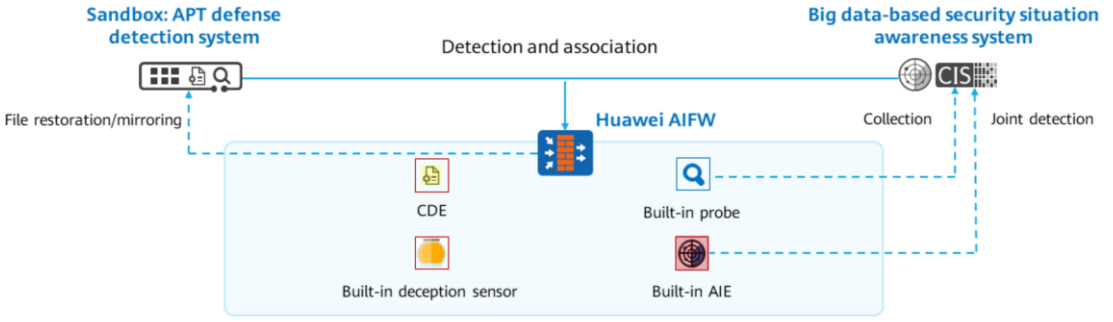
The NGFW is also a stateful inspection firewall. The NGFW greatly improves content security and processing performance.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The stateful inspection firewall detects the first data packet of a connection to determine the status of the connection. Subsequent data packets are forwarded or blocked based on the status of the connection.

AIFW

- The AIFW is a next-generation firewall that integrates AI technology. It further improves the security protection capability and performance of the firewall by using AI algorithms or AI chips.
- Huawei AIFW has built-in core detection engine (CDE) — malicious file detection engine, deception sensor, APT inspection engine (AIE), and probe. It can interwork with the sandbox and Huawei Cybersecurity Intelligence System (CIS) — big data analytics platform for detection, building an intelligent defense system.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Huawei HiSecEngine USG6000E series is the first AIFW launched in the industry. There is no unified standard for AIFWs. For example, firewalls are trained using a large amount of data and algorithms so that they can proactively identify threats. The built-in AI chip of firewalls helps improve application identification and forwarding performance.
- Advanced Persistent Threats (APTs) persistently attack specific targets using advanced attack methods.
- The sandbox is a security device used to detect viruses. It builds a virtual environment for suspected viruses and detects viruses by observing their subsequent behaviors. The sandbox is an important device for APT detection. Huawei FireHunter is a sandbox.
- The CIS can effectively collect network traffic, and network and security logs of various devices. Based on real-time and offline analysis of big data and machine learning technology, expert reputation, and intelligence, the CIS can effectively detect potential and advanced threats on a network, implement security situation awareness of the entire network, and effectively complete the closed-loop handling of threats with the help of Huawei HiSec solution.
- Huawei-developed CDE uses the PE Class 2.0 AI algorithm to restore all files and perform in-depth detection on file content. (Flow detection is the mainstream in the industry. This technology is fast, but it restores only the file header and does not check the file content.
- Huawei's unique AIE APT detection engine uses AI algorithms to continuously defend

against the latest threats.

- For more information about AIFWs, see https://e.huawei.com/en/products/enterprise-networking/security.

Section Summary

- This section describes the basic concepts, application scenarios, and development history of firewalls.
  - Firewalls are security devices and fall into fixed, modular, desktop, and software firewalls.
  - Firewalls can be used for but not limited to enterprise border protection, intranet control and security isolation, DC border protection, and DC security association.
  - The firewall technology is developing from the early packet filtering firewall to the current AIFW.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

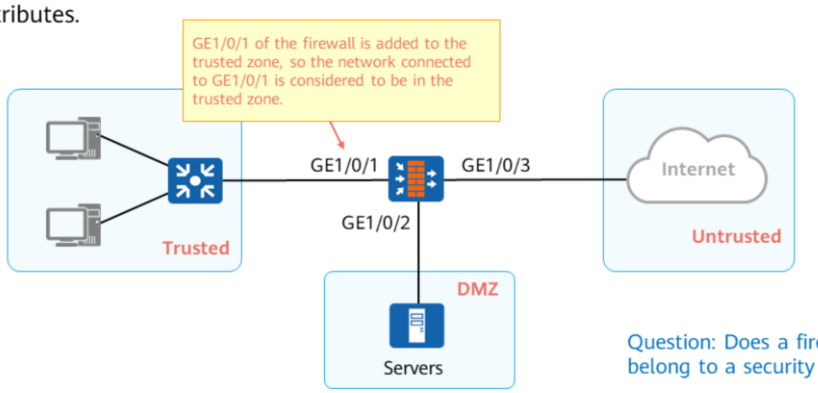
1. Introduction to Firewalls
2. **Basic Firewall Concepts**
3. Basic Firewall Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Basic Firewall Concepts: Security Zone

- A security zone is an important concept of a firewall. Most security policies are implemented based on security zones.
- A security zone is a set of networks connected through interfaces. Users in a zone have the same security attributes.



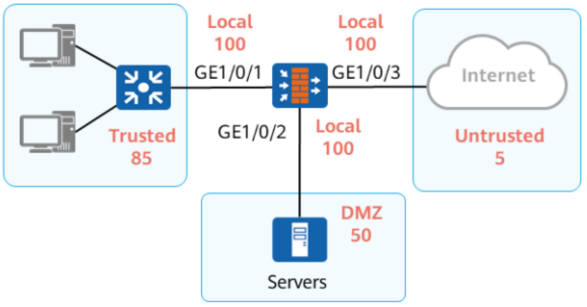
Question: Does a firewall interface belong to a security zone?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Security Zone | Security Policy | Session Table | Server Map

Default Security Zones

- Four security zones have been created on Huawei firewalls: untrusted, DMZ, trusted, and local zones. Security zones have the following characteristics:
  - The default security zone cannot be deleted, and the security priority cannot be changed.
  - Each security zone must be configured with a security priority. A larger value indicates a higher priority of the security zone.
  - You can create customized zones as required.



Security Zone Name	Default Security Priority
Untrusted zone	5 (low security level)
DMZ	50 (medium security level)
Trusted zone	85 (high security level)
Local zone	A local zone is a device itself, including interfaces on the device. The local zone has the highest security level, and its priority is 100.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

The default security zones are as follows:

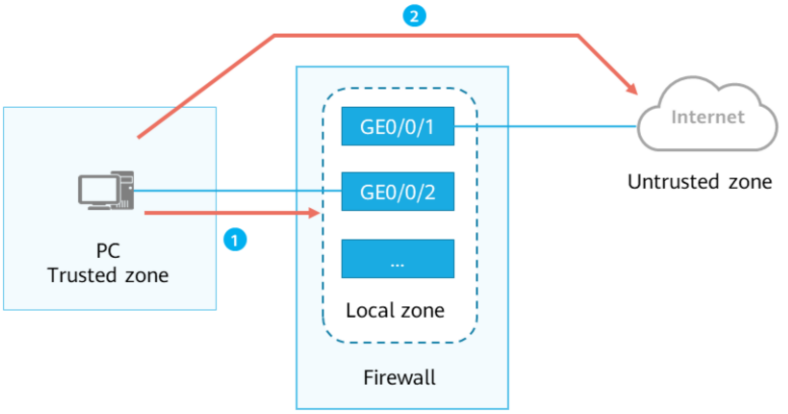
- Untrusted zone: defines an insecure network, such as the Internet.
- DMZ: defines the zone where internal network servers reside. Internal network servers are frequently accessed by external network devices but cannot proactively access the external network, which causes huge security risks. These servers are deployed in a DMZ with a lower level than a trusted zone but a higher level than an untrusted zone.
  - A DMZ is originally a military term, referring to a partially controlled area between a military control area and a public area. A DMZ configured on a firewall is logically and physically separated from internal and external networks.
  - Devices that provide network services for external users are deployed in the DMZ. These devices such as web servers and FTP servers provide services for extranet devices. If the servers are placed on an internal network, their security vulnerabilities may be used by external malicious users to attack the internal network. If the servers are deployed on the external network, security cannot be ensured.
- Trusted zone: defines the zone where internal network terminals reside.
- A local zone is a device itself, including interfaces on the device. All packets constructed on and proactively sent from the device are regarded as packet sent from the local zone; the packets to be responded and processed by the device (including the packets to be detected or directly forwarded) are regarded as packets received in the local zone. Configurations of the local zone cannot be changed, for example, interfaces

cannot be added to the local zone.

- A security policy for exchanging packets between the local zone and the security zone of a peer can be configured for applications where devices need to receive and send packets by themselves.

Interzone Example

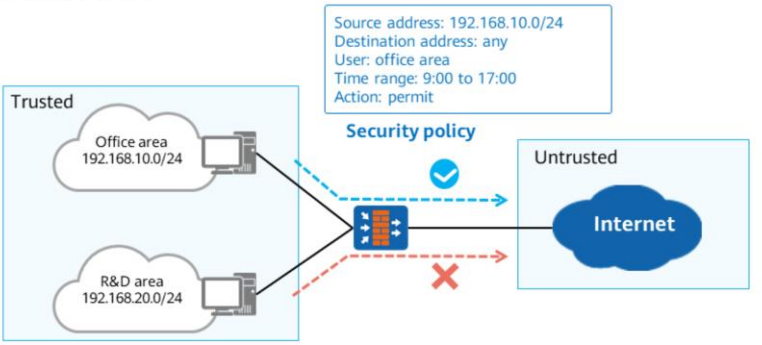
- The source and destination addresses of the traffic determine the zones that can access each other. For 1, traffic from the PC to the firewall interface is transmitted from the trusted zone to the local zone. For 2, the traffic from the PC to Internet is transmitted from the trusted zone to the untrusted zone.



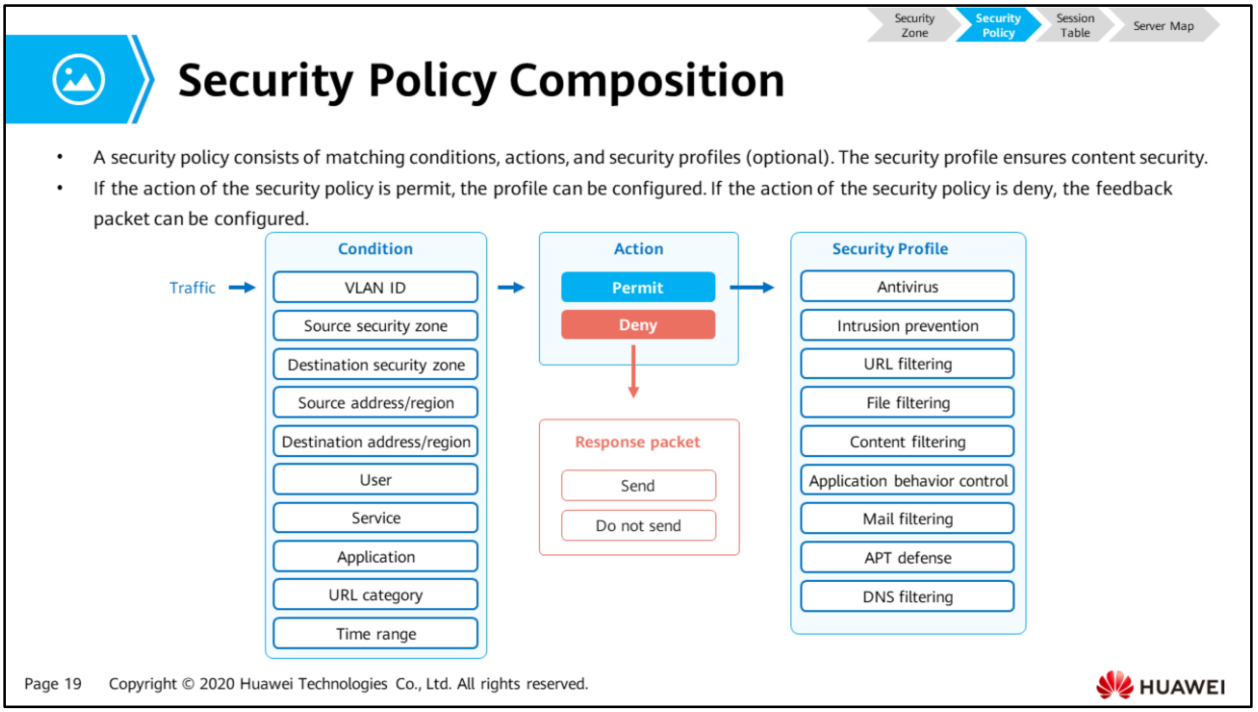
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Basic Firewall Concepts: Security Policy

- A security policy is used by a firewall to control traffic forwarding and perform integrated detection over the traffic content.
- When receiving traffic, the firewall identifies traffic attributes (5-tuple, user, and time range) and matches them with the conditions of the security policy. If the traffic matches the conditions, the corresponding action is performed on the traffic.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



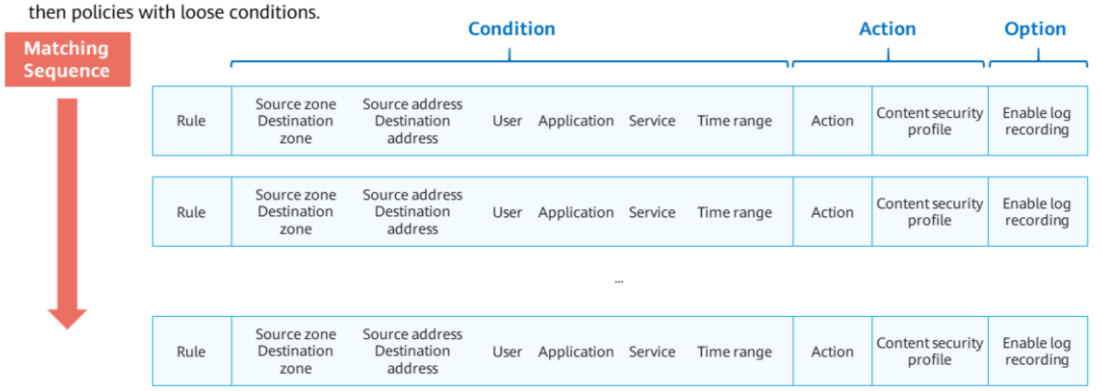
- Actions:
  - Permit: If the action is **permit**, a firewall processes the traffic as follows:
    - If content security detection is not configured, the firewall allows the traffic to pass through.
    - If content security detection is configured, the firewall determines whether to permit the traffic based on the content security detection result. Content security detection includes antivirus and intrusion prevention, which are implemented by referencing security profiles in security policies. If one security profile blocks the traffic, the firewall blocks the traffic. If all security profiles permit the traffic, the firewall allows the traffic to pass through.
  - Deny: The firewall does not allow the traffic that matches a security policy to pass through.
    - If the action is **deny**, the firewall discards the packet and can send a corresponding feedback packet based on the packet type. After the client/server receives the blocking packets from the firewall, it can rapidly terminate sessions and users can detect that the requests have been blocked.
      - Reset client: The firewall sends a TCP reset packet to the TCP client.
      - Reset server: The firewall sends a TCP reset packet to the TCP server.
      - ICMP unreachable: The firewall sends an ICMP unreachable packet to

the client.

- For details, see "Security Policy" in https://support.huawei.com/hedex/hdx.do?docid=EDOC1100092598&lang=en.

Security Policy Matching Process

- When multiple security policies are configured, they are matched according to their configuration sequence. That is, the security policies are matched one by one from the top of the security policy list. If a packet matches a security policy, it will no longer match the following security policies.
- Therefore, the configuration sequence of security policies is important. You need to configure policies with precise conditions and then policies with loose conditions.

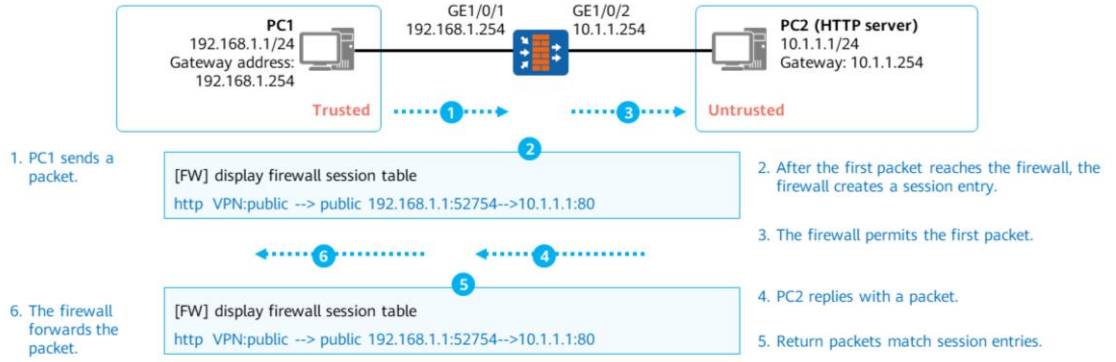


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The system has a default security policy named **default**. The default security policy is located at the bottom of the policy list and has the lowest priority. All matching conditions of the default security policy are **any** and the default action is **deny**. If all the configured policies are not matched, the default security policy is used.

Basic Firewall Concepts: Session Table

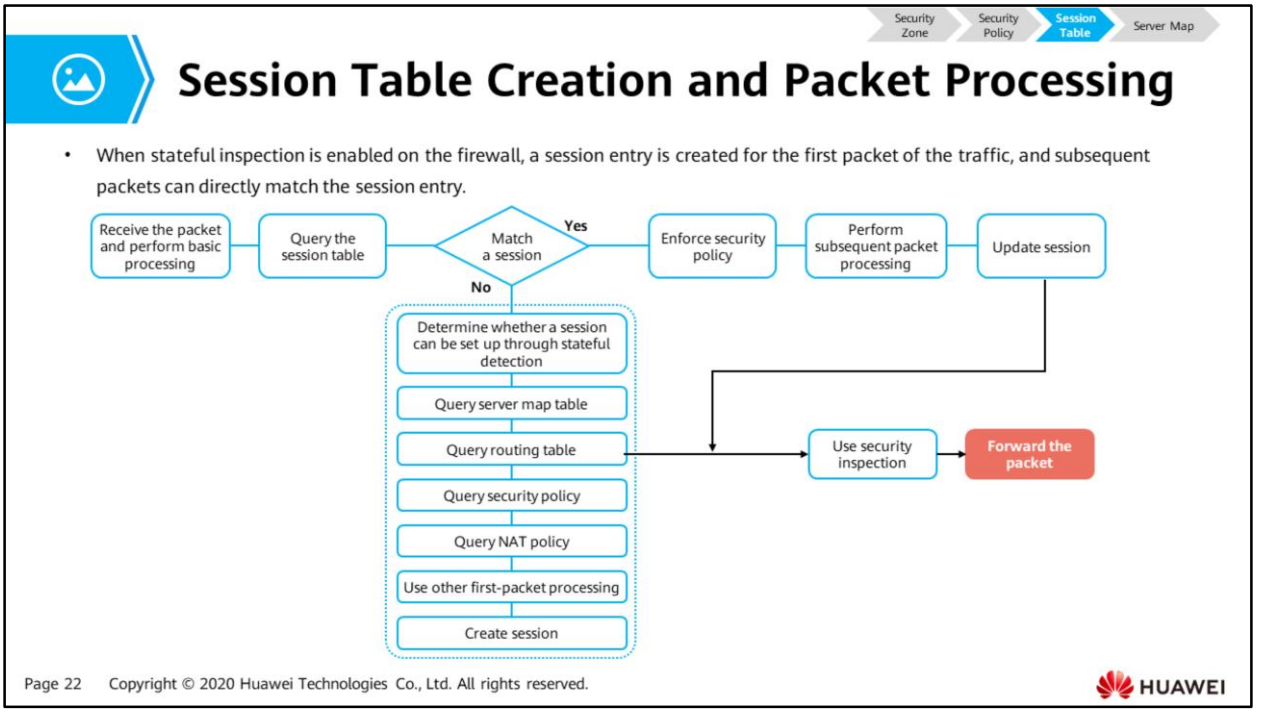
- A session table is used to record the connection status of protocols such as TCP, UDP, and ICMP. It is an important basis for a firewall to forward packets.
- The firewall uses a stateful inspection mechanism which determines the status of a connection based on the first packet or the first several packets. Subsequent packets are then processed based on the connection status. The stateful inspection mechanism improves the detection and forwarding efficiency. The session table is used for maintaining the connection status. To forward TCP, UDP, and ICMP packets, the firewall must look up the session table for the connection status and then process the packets accordingly.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- In this example, PC1 initiates an HTTP connection to PC2, so the firewall marks the HTTP protocol and connection information in the session table and identifies that the traffic is forwarded based on the public routing table (VPN:public in the figure).

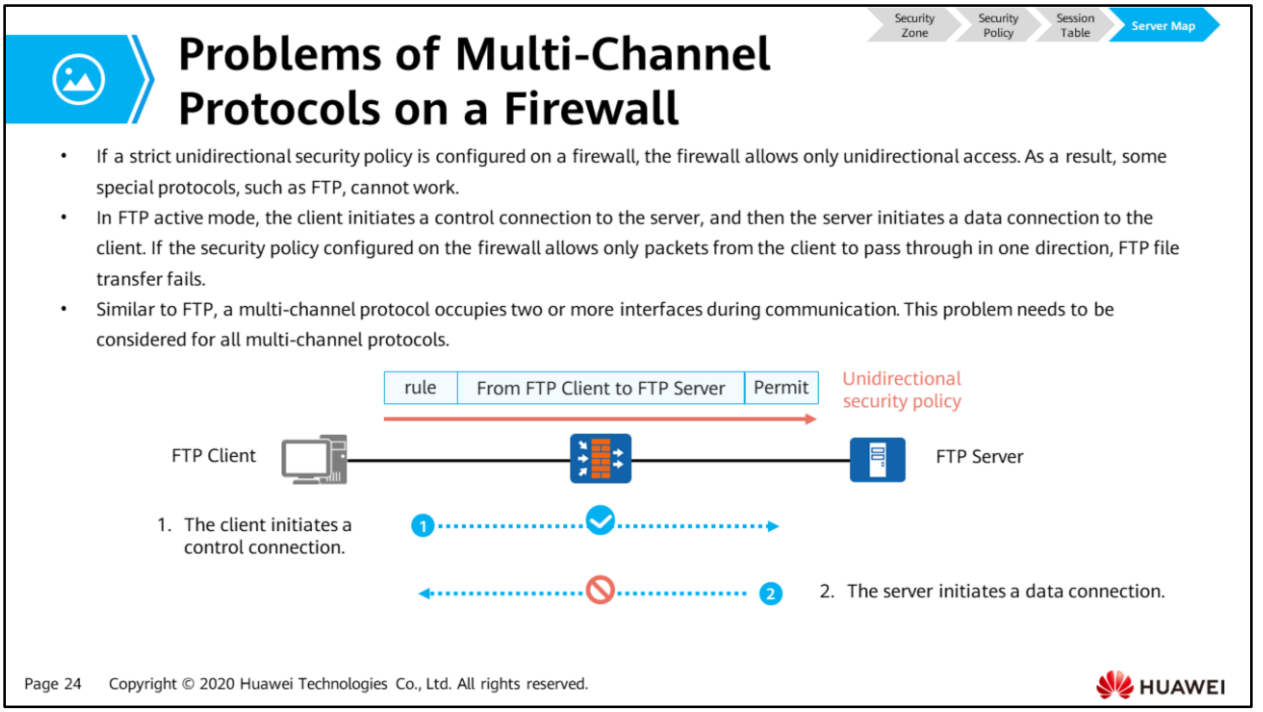


- The flowchart shows the basic processing sequence of each module of a Huawei firewall. In practice, packet processing may be different from the preceding flowchart (if there is no corresponding configuration) and depends on specific product implementation.
- For details, see "Packet Forwarding Process" in the product documentation of the specified firewall model.

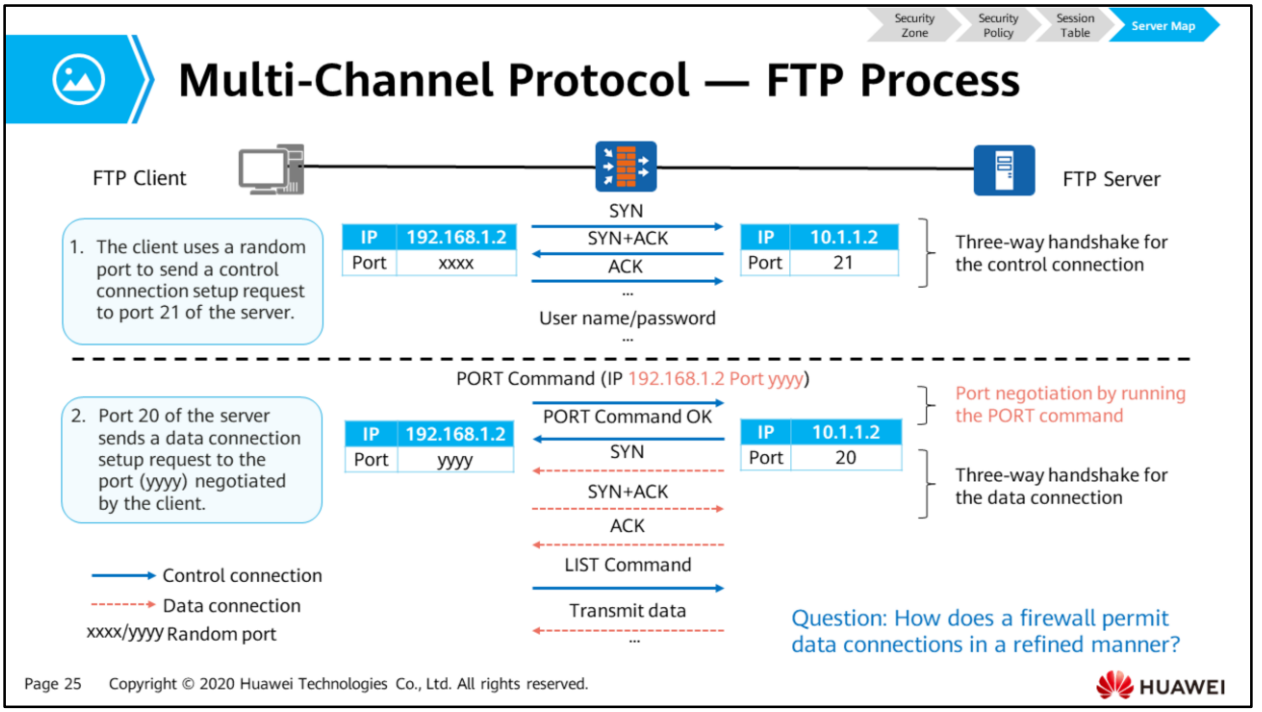
Aging Time and Persistent Connection of the Session Table

- Firewalls provide the session aging mechanism for protocols. A session that is not matched by any packet will be deleted from the session table. This mechanism prevents firewall resources from being consumed by a large number of useless and outdated session entries.
- For some special services, however, the interval between two consecutive packets of a session can be very long. For example:
  - When a user downloads large files through FTP, the interval between control packets along the control channel can be very long.
  - A user may query the data on a database server now and then, and the interval between query operations may be far greater than the aging time of the TCP session.
- If a session entry is deleted, the service is interrupted. The persistent connection mechanism can set an overlong aging time for some connections.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Single-channel protocol: uses only one port during communication. For example, WWW uses only port 80.
- Multi-channel protocol: uses two or more ports for communication.
- FTP is a typical multi-channel protocol. Two connections are set up between the FTP client and server: control and data connections. A control connection is used to transmit FTP instructions and parameters, including information required for establishing a data connection. A data connection is used to obtain server directories and transfer data. The port number used for the data connection is negotiated during the control connection. FTP works in either active (PORT) or passive (PASV) mode, determined by the mode of initiating a data connection. In active mode, port 20 of the FTP server initiates a data connection to the FTP client. In passive mode, the FTP server accepts the data connection initiated by the FTP client. The mode can be set on the FTP client. Here, the active mode is used as an example.
- When multi-channel protocols exist, a firewall can be configured with security policies that define loose conditions to solve the problem of protocol unavailability. However, this brings security risks.



- Most multimedia application protocols (such as H.323 and SIP), FTP, and NetMeeting use prescribed ports to initialize a control connection and then dynamically negotiate a port for data transmission. The port selection is unpredictable. Some applications may even use multiple ports at one time. Packet filtering firewalls can use ACLs to match applications of single-channel protocols to protect internal networks against attacks. However, ACLs can block only applications using fixed ports, and cannot match multi-channel protocol applications that use random ports, bringing security risks.



ASPF and Server Map

- To solve problems of multi-channel protocols, a firewall needs to identify addresses and ports negotiated at the application layer. The Application Specific Packet Filter (ASPF) function must be enabled.
- ASPF is also called status-based packet filtering. It can automatically detect application-layer information of certain packets and define corresponding permit rules according to application-layer information. That is, a server map is generated.
- The server map also records the connection status similar to that in the session table. The server map is equivalent to a simplified session table and is generated before actual traffic arrives. When the traffic reaches the firewall, the firewall generates a session table based on the server map and forwards the traffic.
- ASPF is enabled to solve the multi-channel protocol problem. It is a method of generating a server map.

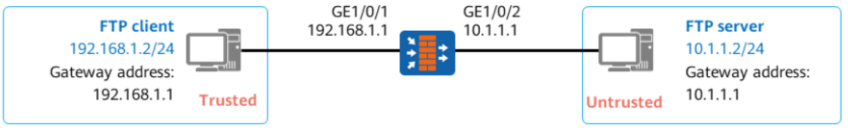
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- When ASPF, NAT server, or source NAT (SNAT) in No-PAT mode is configured on a firewall, the firewall generates corresponding server map entries.

Example of ASPF and Server Map

- After ASPF is configured on a firewall, the firewall checks the negotiated data connection port information in the FTP control connection and generates a server map entry. The server map entry contains information about the data channel negotiated in the FTP control channel. Then the firewall creates a session table for the packets that match server map entries.



1. Create a server map on the firewall.

```
[FW] display firewall server-map
Type: ASPF,  10.1.1.2 -> 192.168.1.2:2097,  Zone:---
Protocol: tcp(Appro: ftp-data),  Left-Time:00:00:10
Vpn: public --> public
```

2. Create a session table on the firewall.

```
[FW] display firewall session table
ftp  VPN: public --> public   192.168.1.2:2095 +-> 10.1.1.2:21
ftp-data  VPN: public --> public  10.1.1.2:20 --> 192.168.1.2:2097
```

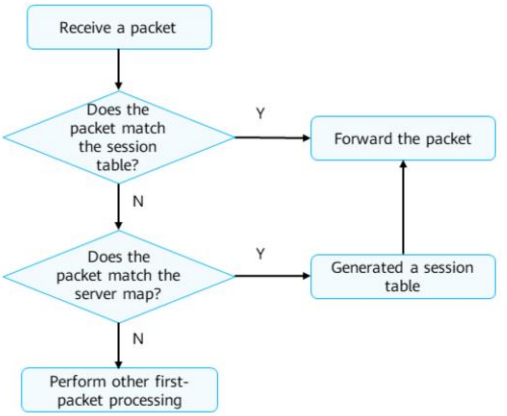
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The relationship between a server map and a session table:
  - A server map records key information about application-layer data. If a packet matches the server map, the security policy is invalid for the packet.
  - A session table represents the connection status of two communication parties.
  - The server map does not represent the current connection status. It predicts subsequent packets based on the analysis of an existing connection.
  - When receiving a packet, a firewall first checks whether the packet matches the session table.
  - If not, the firewall checks whether the packet matches the server map.
  - The security policy is invalid for the packet matching the server map.
  - Then the firewall creates a session table for the packet matching the server map.

Server Map and Simplified Packet Forwarding Process

- When a firewall receives a packet that does not match a session table, the firewall starts the first-packet processing to check whether the packet matches the server map. If so, the firewall generates a session table to forward packets. If not, the firewall performs other packet processing.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Section Summary

- This section describes basic concepts and features of firewalls, including:
  - Security zone
  - Security policy
  - Session table
  - Server map

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Introduction to Firewalls
2. Basic Firewall Concepts
3. **Basic Firewall Configurations**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Basic Firewall Configurations — Interface

1. Create an interface or enter the interface view.

   [Huawei] **interface** *interface-type interface-number*

   The **interface** command on a firewall is used to create an interface or enter the view of a specified interface, which is similar to that on switches and routers.

2. Configure the protocol allowed by the interface.

   [Huawei-GigabitEthernet0/0/1] **service-manage { http | https | ping | ssh | snmp | netconf | telnet | all } { permit | deny }**

   The **service-manage** command allows or blocks access to a firewall through HTTP, HTTPS, ping, SSH, SNMP, NETCONF, or Telnet.
   By default, an interface has the access control and management function enabled. HTTP, HTTPS, and ping permissions are enabled on the management interface only. All permissions on non-management interfaces are disabled.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Firewall Configurations — Security Zone

1. Create a security zone and enter the security zone view.

   [Huawei] **firewall zone name** *zone-name* [**id** *id*]

   firewall zone name creates a security zone and displays the security zone view. *id* specifies the ID of a security zone. The value ranges from 4 to 99 and increases in ascending order.
   firewall zone displays the security zone view. The default four security zones of a firewall cannot be deleted.
2. Set the priority of the security zone.

   [Huawei-zone-*name*] **set priority** *security-priority*

   The priority value ranges from 1 to 100. It is unique. A larger value indicates a higher priority. Default security zones cannot be deleted, and their priorities cannot be reconfigured or deleted.

3. Add an interface to the security zone.

   [Huawei] **add interface** *interface-type { interface-number | interface-number.subinterface-number }*

   A security zone needs to be associated with a specific interface of a firewall. That is, the interface needs to be added to the security zone. The interface can be a physical or logical one.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Firewall Configurations — Security Policy (1)

1. Enter the security policy view.

   [Huawei] **security-policy**

   You can create, copy, move, or rename a security policy rule in the security policy view.

2. Configure a security policy rule in the security policy view and enter the security policy rule view.

   [Huawei-policy-security] **rule name** *rule-name*

3. Configure the source security zone of the security policy rule in the security policy rule view.

   [Huawei-policy-security-rule-name] **source-zone** *{ zone-name* &<1-6> | **any** *}*

   The security zone must exist in the system. You can add or delete a maximum of six security zones in a security policy rule at a time.

4. Configure the destination security zone of the security policy rule in the security policy rule view.

   [Huawei-policy-security-rule-name] **destination-zone** *{ zone-name* &<1-6> | **any** *}*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Firewall Configurations — Security Policy (2)

5. Configure the source IP address of the security policy rule.

   [Huawei-policy-security-rule-name] **source-address** *ipv4-address { ipv4-mask-length* | **mask** *mask-address}*

   In the command, *mask-address* is a wildcard mask.

6. Configure the destination IP address of the security policy rule.

   [Huawei-policy-security-rule-name] **destination-address** *ipv4-address { ipv4-mask-length* | **mask** *mask-address}*

   In the command, *mask-address* is a wildcard mask.

7. Configure a service.

   [Huawei] **service** *{ service-name &<1-6>* | **any** *}*

   The **service** command configures a service. For example, the **service protocol** command references a TCP/UDP/SCTP port or an IP-layer protocol in a security policy.

8. Configure an action in the security policy rule.

   [Huawei] **action** { **permit** | **deny** }

   The default action of a firewall is **deny**.

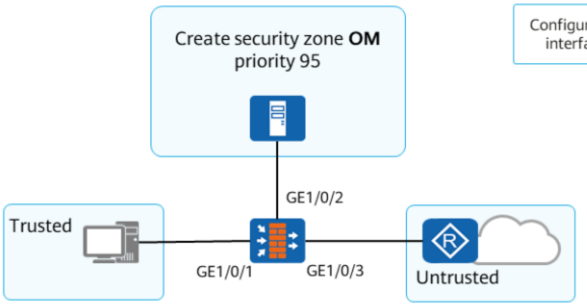
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The source and destination IP addresses specified in the security policy rule view can have many optional parameters, such as the IP address group, region, and region group. This course does not describe these optional parameters. For more information, see the product documentation.

Firewall Configuration Examples

Description:

- The firewall isolates the network into three security zones: trusted, untrusted, and **OM**. The priority of the **OM** zone is 95. The requirements are as follows:
  - GE1/0/1 of the firewall can respond to ping requests.
  - ICMP traffic in the **OM** zone can reach the untrusted zone.





The configuration process consists of four steps:

- Configure a firewall interface.
- Configure a security zone.
- Configure a security policy.
- Verify the configuration.

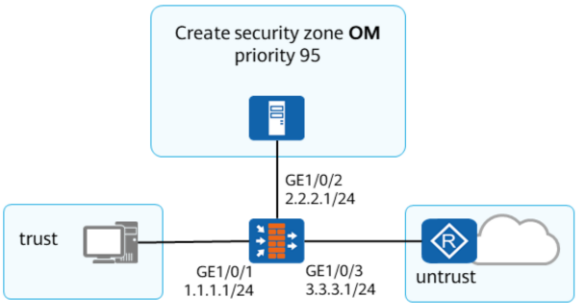
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example — Interface

Configure an interface → Configure a security zone → Configure a security policy → Verify the configuration

Task list:

- Configure an IP address for a firewall interface according to the planning.
- Configure GE1/0/1 to allow the ping service.



Configure an IP address for the interface and allow the ping service on GE1/0/1.

```
[FW] interface GigabitEthernet 1/0/1
[FW-GigabitEthernet1/0/1] ip address 1.1.1.1 24
[FW-GigabitEthernet1/0/1] service-manage ping permit
[FW-GigabitEthernet1/0/1] interface GigabitEthernet 1/0/2
[FW-GigabitEthernet1/0/2] ip address 2.2.2.1 24
[FW-GigabitEthernet1/0/2] interface GigabitEthernet 1/0/3
[FW-GigabitEthernet1/0/3] ip address 3.3.3.1 24
```

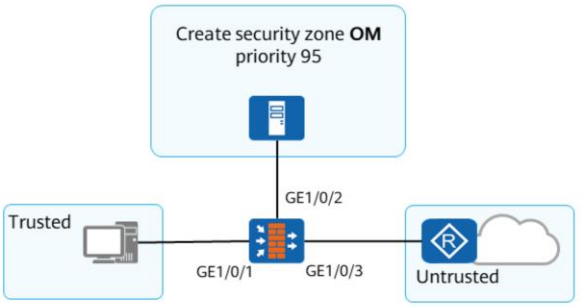

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example — Security Zone

Configure an interface → **Configure a security zone** → Configure a security policy → Verify the configuration

Task list:

- Create security zone **OM** and set its priority to 95.
- Add interfaces to the planned security zones.




```
# Create a security zone.
[FW] firewall zone name OM
[FW-zone-OM] set priority 95
[FW-zone-OM] quit

# Add interfaces to corresponding security zones.
[FW] firewall zone trust
[FW-zone-trust] add interface GigabitEthernet 1/0/1
[FW] firewall zone OM
[FW-zone-OM] add interface GigabitEthernet 1/0/2
[FW] firewall zone untrust
[FW-zone-untrust] add interface GigabitEthernet 1/0/3
```

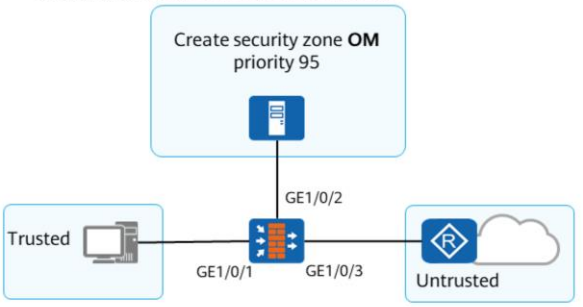

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example — Security Policy Configuration

Configure an interface → Configure a security zone → **Configure a security policy** → Verify the configuration

Task list:

- Create security policy **R1**.
- Configure a security policy rule, including the source and destination zones, service type, and action.



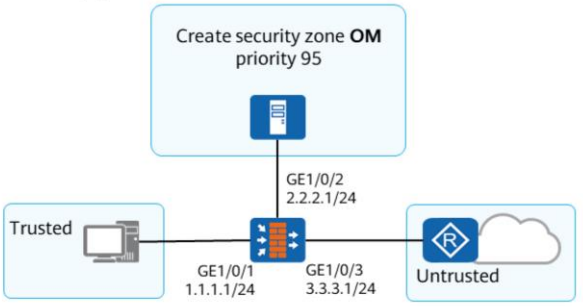

```
# Create a security policy.
[FW-policy-security] rule name R1
[FW-policy-security-rule-R1] source-zone OM
[FW-policy-security-rule-R1] destination-zone untrust
[FW-policy-security-rule-R1] service icmp
[FW-policy-security-rule-R1] action permit
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example — Verification (1)

Configure an interface → Configure a security zone → Configure a security policy → **Verify the configuration**

1. The PC in the trusted zone initiates a ping test to GE1/0/1 of the firewall.



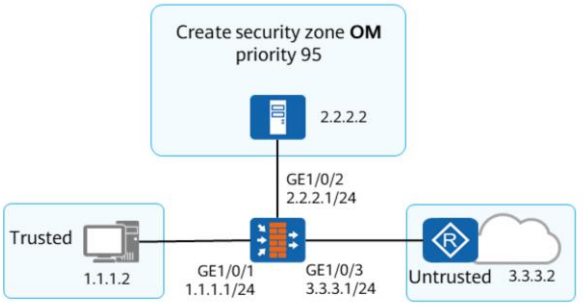

```
# Check the session table on the firewall.
[FW]display firewall session table
2020-03-11 10:31:21.010
 Current Total Sessions : 4
 icmp  VPN: public --> public  1.1.1.2:14265 --> 1.1.1.1:2048
 icmp  VPN: public --> public  1.1.1.2:15289 --> 1.1.1.1:2048
 icmp  VPN: public --> public  1.1.1.2:14777 --> 1.1.1.1:2048
 icmp  VPN: public --> public  1.1.1.2:15033 --> 1.1.1.1:2048
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example — Verification (2)

Configure an interface → Configure a security zone → Configure a security policy → **Verify the configuration**

2. The device at 2.2.2.2 in security zone **OM** initiates a ping request to the device at 3.3.3.2 in the untrusted zone.




```
# Check the session table on the firewall.
[FW]display firewall session table
2020-03-11 10:30:15.150
 Current Total Sessions : 4
  icmp  VPN: public --> public  2.2.2.2:63928 --> 3.3.3.2:2048
  icmp  VPN: public --> public  2.2.2.2:63672 --> 3.3.3.2:2048
  icmp  VPN: public --> public  2.2.2.2:63416 --> 3.3.3.2:2048
  icmp  VPN: public --> public  2.2.2.2:62904 --> 3.3.3.2:2048
```

Question: ICMP does not have a port number. What is the port number in the firewall session table?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- ICMP does not have a port. However, the firewall generates a port number when generating the session table corresponding to ICMP traffic to meet status detection requirements.

Section Summary

- This section describes basic configuration commands of the firewall. After learning basic firewall configurations, you can use commands to perform the following configurations:
  - Configure firewall interfaces.
  - Assign firewall security zones.
  - Configure security policies.
  - Check firewall session information.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Single) By default, a firewall has ( ) security zones.

   A. 1 B. 2 C. 3 D. 4

2. (TorF) The firewall must generate a server map before generating a session table.
3. (Multiple) Which of the following statements about Huawei AIFWs are correct?

   A. Built-in CDE - malicious file detection engine

   B. Built-in probe

   C. Built-in APT detection engine (AIE)

   D. Built-in deception sensor

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- 1. D
- 2. F
- 3. ABCD

Summary

- Firewalls are one of the most important security devices in the communications field. This course helps you learn basic concepts, development history, and basic configurations of firewalls.
- The security zone, security policy, session table, and server map are basic and important concepts of firewalls and are the prerequisite for further learning the firewall technology.
- Firewalls provide more security functions, such as content security filtering, antivirus, and intrusion prevention. For more information about firewall technologies, see Huawei Certification — Security.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

More Information

- Huawei HiSecEngine USG6000E series AIFW documents.
- https://support.huawei.com/enterprise/en/security/usg6600e-pid-23176231

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Recommendations

- For more training materials and videos, visit https://e.huawei.com/en/talent/#/home.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Security Features of Network Devices

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- IP networks of large scales usually have numerous network devices and various communication protocols, adding complexity to network management. It is difficult to balance security and service flexibility, as well as security and management & maintenance convenience. When striking a balance between these, administrators with varying degrees of technical and management expertise perform differently. Administrators may ignore security protection capabilities in the pursuit of service availability, resulting in the failure to configure necessary security measures. The devices, therefore, cannot effectively defend against attacks as they are designed to.
- This course describes common security hardening policies of network devices and provides examples for common security configurations.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Learn common device security hardening policies.
  - Configure STelnet.
  - Configure local attack defense.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

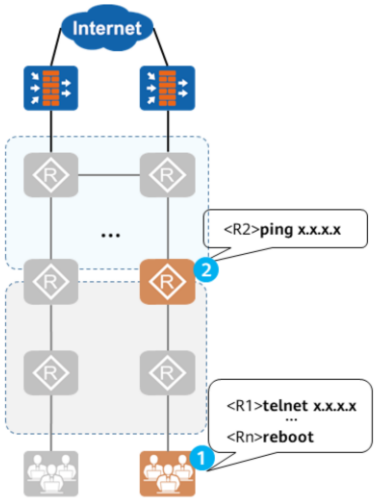
Contents

1. **Common Security Hardening Policies**
2. Examples for Security Hardening Policy Deployment

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Why Is Security Vital to Network Devices?



- Network security is a systematic project. Everything on the network may become a target of attacks, and network devices are no exception.
- Common attacks on network devices include the following:
  1. Malicious users log in to network devices to perform unauthorized operations (for example, restarting the devices), leading to malfunctioning networks.
  2. A large number of control packets (for example, ICMP packets) are forged and sent to the CPU of a target device, leading to a surge in the CPU usage.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Common Security Hardening Policies

Common security hardening policies on network devices include the following:

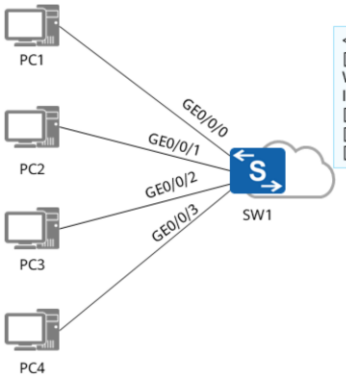
- Disabling unused services and ports
- Discarding insecure access channels
- Access control based on trusted paths
- Local attack defense

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Disabling Unused Services and Ports

Based on service requirements analysis and the minimum authorization principle, disable unused services and ports.

- Disable unused physical ports, which then cannot be used for communication even when network cables are connected.
- Disable unused protocol ports, for example, common Telnet, FTP, and HTTP ports, which then cannot be accessed from external systems.




```
<SW1> system-view
[SW1] undo ftp server
Warning: The operation will stop the FTP server. Do you want to continue? [Y/N]:y
Info: Succeeded in closing the FTP server.
[SW1]port-group protgroup1
[SW1-port-group-protgroup1]group-member GigabitEthernet 0/0/4 to GigabitEthernet0/0/48
[SW1-port-group-protgroup1]shutdown
```

Disable the FTP function and unused ports on SW1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Discarding Insecure Access Channels

The access requirements of services must be preferentially fulfilled based on service requirements analysis. When an access requirement involves multiple access channel services, secure channels must be selected and insecure ones must be discarded.

Access Requirement	Insecure Channel	Secure Channel
Remote login	Telnet	SSHv2
File transfer	FTP and TFTP	SFTP
NE management	SNMPv1 and SNMPv2	SNMPv3
NMS login	HTTP	HTTPS

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- When you log in to a device through the CLI, web UI, or NMS, you are advised to use the corresponding SSH, HTTPS, or SNMPv3 channel.
- SFTP is recommended for data transmission between devices and between devices and terminals.

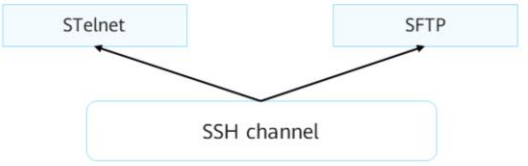
Secure Data Access Channel

- To ensure device security, select secure access channels as long as conditions permit.
- Common scenarios and protocols for device data transmission security:
  - Remote login:
    - Telnet: transmits data in clear text using TCP.
    - STelnet: provides secure information protection and powerful authentication functions based on SSH.
  - File operations:
    - FTP: supports file transfer and file directory operations, provides authorization and authentication functions, and transmits data in clear text.
    - TFTP: supports only file transfer and transmits data in clear text. It does not support authorization or authentication.
    - SFTP: supports file transfer and file directory operations; strictly encrypts data and protects data integrity.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SSH Overview

- Secure Shell (SSH) provides secure remote login, file transfer, and TCP/IP tunnels on insecure networks. It encrypts not only the password during login, but also the executed command data after login.
- After an authorized user logs in from the client and the user name and password are verified, the client attempts to establish a session with the server. Each session is an independent logical channel and can be used by different upper-layer applications.
- STelnet and SFTP each use a logical channel to encrypt data through SSH, implementing secure data transmission.



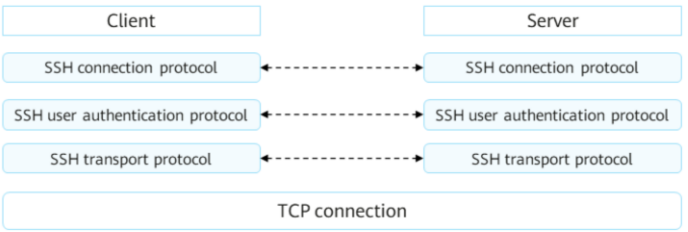
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- SSH is developed by the IETF. The latest version is V2.0. Earlier versions 1.3 and 1.5 have security risks and are gradually obsolete.
- SSH supports two-way authentication between the server and client, and provides security services such as confidentiality and integrity protection.

SSH Protocol Structure

The SSH protocol framework has three pillar protocols: transport protocol, user authentication protocol, and connection protocol.

- Transport protocol: provides version negotiation, encryption algorithm negotiation, key exchange, server authentication, and information integrity verification.
- User authentication protocol: authenticates clients for the server.
- Connection protocol: multiplexes an encrypted information tunnel into multiple logical channels, which are provided for upper-layer application protocols (STelnet and SFTP). Various upper-layer application protocols can be relatively independent of the SSH basic system. They also rely on this basic system to use the SSH security mechanism through the connection protocol.

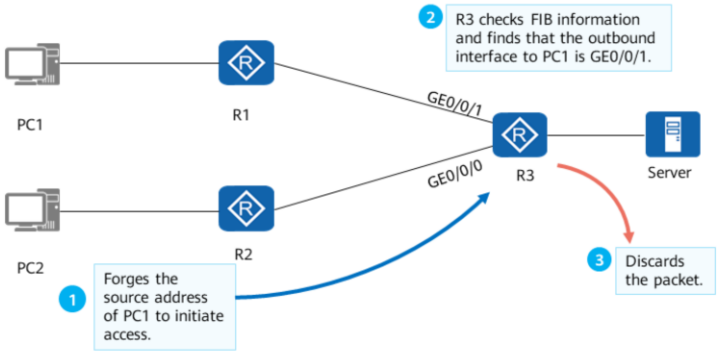


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- SSH uses the following types of algorithms:
  - MAC algorithms for data integrity protection, such as HMAC-MD5 and HMAC-MD5-96
  - Data encryption algorithms, such as 3DES-CBC, AES128-CBC, and DES-CBC
  - Key exchange algorithm used to generate session keys, such as diffle-hellman-group-exchange-sha1
  - Host public key algorithm used for digital signature and authentication, such as RSA and DSA

Access Control Based on Trusted Paths

- Configure access control policies based on trusted paths on network devices to improve network security.
- Deploy URPF to determine whether the source address of a packet is valid. If the path of the packet is inconsistent with the path learned by URPF, the packet is discarded. In this way, URPF helps to prevent network attacks with spoofed IP source addresses.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The openness of IP networks determines that anyone can access or attack the target host as long as routes are reachable.
- For a host, the path of the packets sent to it from a client is fixed, especially at the edge of a network.
- Unicast Reverse Path Forwarding (URPF) can be classified into strict URPF and loose URPF, and the mode in which matching the default route is allowed can be configured. During the URPF check, the device checks whether source IP addresses of packets are valid based on the routing table.
    - In strict mode, if a packet matches a specific route and the inbound interface of the packet is the same as the outbound interface of the route, the packet is allowed to pass. Otherwise, the packet is discarded.
    - In loose mode, if a packet matches a specific route, the packet is allowed to pass. Otherwise, the packet is discarded. In this mode, the interface is not checked. By default, the device does not match packets with the default route. You can configure the device to match packets with the default route.
    - Matching the default route must work with strict URPF. When a packet matches a specific route or the default route and the inbound interface of the packet is the same as the outbound interface of the matched route, the packet is allowed to pass. Otherwise, the packet is discarded. Matching the default route cannot be configured with loose URPF because attack defense cannot be achieved in this way. Loose URPF and strict URPF are mutually exclusive.

Local Attack Defense

- In addition to numerous normal service packets, CPUs of devices on a network may also receive large numbers of attack packets. If a CPU is busy processing attack packets for an extended period, other services, or even the system itself, will experience interruption. Similarly, if a large number of normal packets are sent to the CPU, the CPU usage will surge and device performance will deteriorate, adversely affecting services.
- To ensure that the CPU can properly process and respond to normal services, the device provides the local attack defense function, which has been specifically designed for packets sent to the CPU and is primarily used to protect the device from attacks and ensure consistency of existing services when an attack occurs.
- Local attack defense includes CPU attack defense and attack source tracing.
  - CPU attack defense can rate-limit packets destined for the CPU so that only a limited number of packets are sent to the CPU within a certain period of time. This ensures that the CPU can properly process services.
  - Attack source tracing defends against Denial of Service (DoS) attacks. A device enabled with attack source tracing analyzes packets sent to the CPU, collects statistics on the packets, and allows a packet rate threshold to be set for the packets. Packets sent at a threshold-crossing rate are considered as attack packets. The device finds the source user address or source interface of the attacker by analyzing the attack packets and generates logs or alarms to alert a network administrator. The network administrator then takes measures to protect the device against the attack or configure the device to discard packets sent by the attack source.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

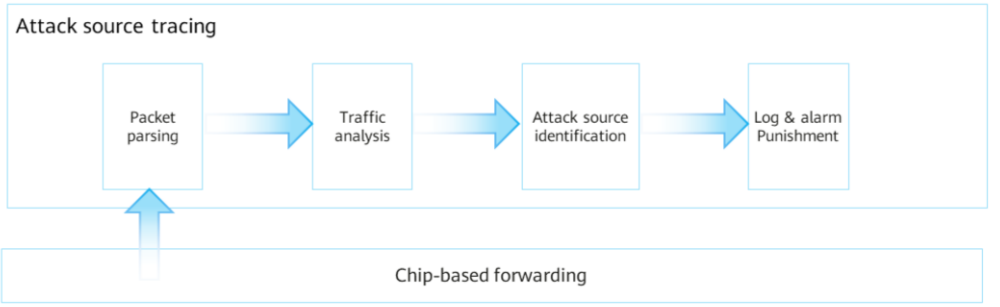
CPU Attack Defense

- Network devices employ four-level security mechanisms to protect their security:
  - Level 1: filter invalid packets sent to the CPU by using blacklists.
  - Level 2: Use Control Plane Committed Access Rate (CPCAR) to rate-limit the packets sent to the CPU based on the protocol type, preventing excess packets of a protocol from being sent to the CPU.
  - Level 3: schedule packets sent to the CPU based on the protocol priority to ensure that packets with higher protocol priorities are preferentially processed.
  - Level 4: uniformly rate-limit all packets sent to the CPU and randomly discard the excess packets to ensure CPU security.
- Regarding rate limiting on packets sent to the CPU in active link protection, when a device detects the establishment of an SSH, Telnet, HTTP, FTP, or BGP session, it enables active link protection for the session. Subsequent packets that match the session characteristics will be sent to the CPU at a high rate, ensuring the reliability and stability of services related to the session.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Working Mechanism of Attack Source Tracing

- Attack source tracing involves four steps: parsing packets, analyzing traffic, identifying an attack source, and sending logs or alarms to a network administrator or taking punishment measures.



- The device locates the attack source, and the network administrator rate-limits the packets sent from the attack source by configuring ACLs or blacklists to protect the CPU.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Common Security Hardening Policies
2. Examples for Security Hardening Policy Deployment
   - STelnet Configuration
   - Local Attack Defense Configuration

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Basic SSH Configuration (1)

1. Enable the SSH server function.

```
[Huawei] stelnet server enable
```

2. Configure the authentication mode for an SSH user.

```
[Huawei] ssh user user-name authentication-type { password | rsa | password-rsa | all }
```

If RSA authentication is used, configure the public key of the SSH client on the SSH server. When the SSH client connects to the SSH server, the SSH client passes the authentication if its private key matches the configured public key.

3. Configure the public key generated by the SSH client.

```
[Huawei] rsa peer-public-key key-name [ encoding-type { der | openssh | pem } ]
[Huawei-rsa-public-key] public-key-code begin
```

The public key must be a hexadecimal character string in the public key encoding format, and generated by the SSH client.

4. Exit the public key editing view, edit the public key view, and return to the system view.

```
[Huawei-rsa-key-code] public-key-code end
[Huawei-rsa-public-key] peer-public-key end
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic SSH Configuration (2)

1. Assign an RSA public key to the SSH user.

   [Huawei] **ssh user** *user-name* **assign** { **rsa-key** | **ecc-key** } *key-name*

2. Generate the local RSA host and server key pairs.

   [Huawei] **rsa local-key-pair create**

   If an RSA key pair exists, the system prompts you to confirm whether to replace the original key pair. After you run this command, the system prompts you to enter the number of bits of the host key. The difference between the bits in the server and host key pairs must be 128 bits or more. The host key pair length and server key pair length both range from 512 to 2048, in bits, and their default values are both 2048.

3. Enable first authentication on the SSH client.

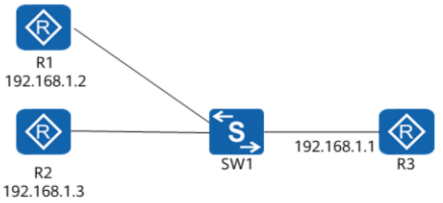
   [Huawei] **ssh client first-time enable**

   When an SSH client accesses an SSH server for the first time and the public host key of the SSH server is not configured on the SSH client, enable the first authentication function. The SSH client then can access the SSH server and save the public host key on the SSH client. When the SSH client accesses the SSH server next time, the saved public host key is used to authenticate the SSH server.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SSH Configuration Example

- Configure STelnet for secure remote login to R3.
- Configure two login users client001 and client002 on R3. R1 uses client001 to log in to R3 in password authentication mode, and R2 uses client002 to log in to R3 in RSA authentication mode. Configure a security policy to ensure that only users from R1 and R2 can log in to R3.

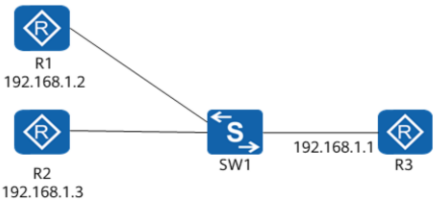


Procedure:

1. Configure the SSH server (R3) to generate a local key pair for secure data exchange between the server and clients (R1 and R2).
2. Configure SSH users client001 and client002 on the SSH server.
3. Enable the STelnet server function on the SSH server.
4. Set the service type of SSH users client001 and client002 to STelnet on the SSH server.
5. Customize an SSH service port number on the SSH server. Then attackers cannot connect to the server through the standard SSH service port, ensuring security.
6. Log in to the SSH server as users client001 and client002 through STelnet.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SSH Configuration Commands (1)

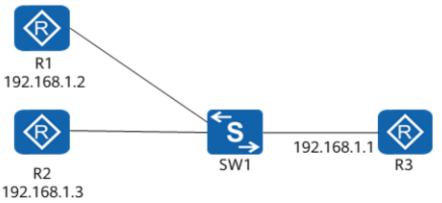


1. Configure the SSH server (R3) to generate a local key pair for secure data exchange between the server and clients (R1 and R2).
2. Configure SSH users client001 and client002 on the SSH server.
3. Enable the STelnet server function on the SSH server.
4. Set the service type of SSH users client001 and client002 to STelnet on the SSH server.
5. Customize an SSH service port number on the SSH server. Then attackers cannot connect to the server through the standard SSH service port, ensuring security.
6. Log in to the SSH server as users client001 and client002 through STelnet.

```
[R3] rsa local-key-pair create
The key name will be: Host
RSA keys defined for Host already exist. Confirm to replace them? (y/n):y
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is less than 2048,
       It will introduce potential security risks.
Input the bits in the modulus[default = 2048]:2048
Generating keys... ..........++++++++ ..........++++++++
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SSH Configuration Commands (2)



1. Configure the SSH server (R3) to generate a local key pair for secure data exchange between the server and clients (R1 and R2).
2. Configure SSH users client001 and client002 on the SSH server.
3. Enable the STelnet server function on the SSH server.
4. Set the service type of SSH users client001 and client002 to STelnet on the SSH server.
5. Customize an SSH service port number on the SSH server. Then attackers cannot connect to the server through the standard SSH service port, ensuring security.
6. Log in to the SSH server as users client001 and client002 through STelnet.

```
[R3] user-interface vty 0 4
[R3-ui-vty0-4] authentication-mode aaa
[R3-ui-vty0-4] protocol inbound ssh
[R3-ui-vty0-4] quit
[R3] aaa
[R3-aaa] local-user client001 password irreversible-cipher Huawei@123
[R3-aaa] local-user client001 privilege level 3
[R3-aaa] local-user client002 password irreversible-cipher Huawei@123
[R3-aaa] local-user client002 privilege level 3
[R3-aaa] quit
[R3] ssh user client001 authentication-type password
[R3] ssh user client002 authentication-type rsa
```

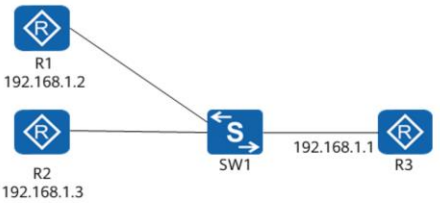
```
[R3] stelnet server enable
```

```
[R3] aaa
[R3-aaa] local-user client001 service-type ssh
[R3-aaa] local-user client002 service-type ssh
[R3-aaa] quit
```

```
[R3] ssh server port 1025
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SSH Configuration Commands (3)



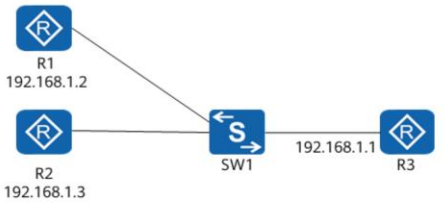
1. Configure the SSH server (R3) to generate a local key pair for secure data exchange between the server and clients (R1 and R2).
2. Configure SSH users client001 and client002 on the SSH server.
3. Enable the STelnet server function on the SSH server.
4. Set the service type of SSH users client001 and client002 to STelnet on the SSH server.
5. Customize an SSH service port number on the SSH server. Then attackers cannot connect to the server through the standard SSH service port, ensuring security.
6. Log in to the SSH server as users client001 and client002 through STelnet.

```
# Generate a local key pair on R2.
[R2] rsa local-key-pair create
The key name will be: Host
RSA keys defined for Host already exist.
Confirm to replace them? (y/n):y
The range of public key size is (512 ~ 2048).
NOTES: If the key modulus is less than 2048,
        It will introduce potential security risks.
Input the bits in the modulus[default = 2048]:2048
Generating keys... ..........................................................
+++ ....+++ ..........................++++++++ ..............++++++++
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The configurations in this slide and the following two slides enable user client002 to log in to R3.

SSH Configuration Commands (4)

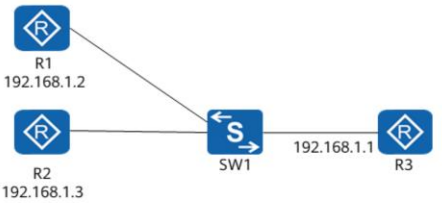


1. Configure the SSH server (R3) to generate a local key pair for secure data exchange between the server and clients (R1 and R2).
2. Configure SSH users client001 and client002 on the SSH server.
3. Enable the STelnet server function on the SSH server.
4. Set the service type of SSH users client001 and client002 to STelnet on the SSH server.
5. Customize an SSH service port number on the SSH server. Then attackers cannot connect to the server through the standard SSH service port, ensuring security.
6. Log in to the SSH server as users client001 and client002 through STelnet.

Display the public key in the RSA key pair generated by R2.
[R2] **display rsa local-key-pair public**
Key code: **30820109 02820100 CB0E88EC A1C2CFEA F97126F9 36919C08 0455127B A3A48594 69517096 35626F55 E4FAF0EB FDA2B9E9 5E417B2B E09F38B0 D26FCA73 FE2E3FC4 DFBEC8CF 4ED0C909 E8D975E6 FFC73C81 D13FE71E 759DC805 B0F0E877 4FC9288E BE1E197C 2A7186B0 B56F5573 3A5EA588 29C63E3B 20D56233 8E63278D F941734F 6B359C69 BBAE5A52 EB842179 04B4204D 5DB31D72 97F0C085 DA771F66 0AAADC28 D264CEB9 5BADA92C CDE9F116 D6D99C48 CEBA3A1D 868B053A 32941D85 CCAA9796 A4B55760 0A8108ED DB45DA12 F61634C9 59431600 341FEDEF 5379D565 A8D1953D DEA018A2 72F99FFC 63DE04BF 2A6219BD DF13D705 27D63DEF 83D556BC 5B44D983 8D5EA126 C1EB71CB 0203 010001**
==========
Time of Key pair created: 2012-08-06 17:17:44+00:00 Key name: Server Key type: RSA encryption Key
==========
Key code: 3067 0260 DF8AFF3C 28213B94 2292852E E98657EE 11DE5AF4 8A176878 CDD4BD31 55E05735 3080F367 A83A9034 47D534CA 81250C1D 35401DC3 464E9E5F A50202CF A7AD09CD AC3F531C A763F0A0 4C8E51B9 18755400 76AF4A78 225C92C3 01FE0DFF 06908363 0203 010001

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SSH Configuration Commands (5)



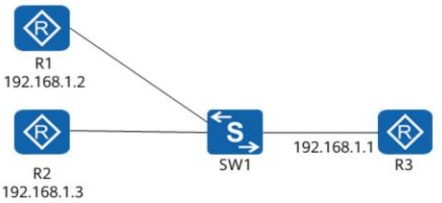
1. Configure the SSH server (R3) to generate a local key pair for secure data exchange between the server and clients (R1 and R2).
2. Configure SSH users client001 and client002 on the SSH server.
3. Enable the STelnet server function on the SSH server.
4. Set the service type of SSH users client001 and client002 to STelnet on the SSH server.
5. Customize an SSH service port number on the SSH server. Then attackers cannot connect to the server through the standard SSH service port, ensuring security.
6. Log in to the SSH server as users client001 and client002 through STelnet.

Copy the RSA public key (the information in bold in the **display** command output in the previous slide) generated on R2 to the server.
[R3] **rsa peer-public-key rsakey001**
[R3-rsa-public-key] **public-key-code begin**
[R3-rsa-key-code] **30820109**
[R3-rsa-key-code] **02820100**
[R3-rsa-key-code] **CB0E88EC A1C2CFEA F97126F9 36919C08 0455127B**
[R3-rsa-key-code] .....
[R3-rsa-key-code] **010001**
[R3-rsa-key-code] **public-key-code end**
[R3-rsa-public-key] **peer-public-key end**

Bind the RSA public key of R2 to client002 on R3.
[R3] **ssh user client002 assign rsa-key rsakey001**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Verifying the SSH Configuration



1. Configure the SSH server (R3) to generate a local key pair for secure data exchange between the server and clients (R1 and R2).
2. Configure SSH users client001 and client002 on the SSH server.
3. Enable the STelnet server function on the SSH server.
4. Set the service type of SSH users client001 and client002 to STelnet on the SSH server.
5. Customize an SSH service port number on the SSH server. Then attackers cannot connect to the server through the standard SSH service port, ensuring security.
6. Log in to the SSH server as users client001 and client002 through STelnet.

```
[R1] ssh client first-time enable
[R1] stelnet 192.168.1.1 1025
Please input the username:client001
Trying 192.168.1.1 ...
Press CTRL+K to abort
Connected to 192.168.1.1 ...
The server is not authenticated. Continue to access it?(y/n)[n]:y
Save the server's public key?(y/n)[n]:y
The server's public key will be saved with the name 192.168.1.1. Please wait...
Enter password:
<R3>                                  # The login is successful.
```

```
[R2] ssh client first-time enable
[R2] stelnet 192.168.1.1 1025
Please input the username:client002
Trying 192.168.1.1 ... Press CTRL+K to abort Connected to 192.168.1.1 ...
The server is not authenticated. Continue to access it?(y/n)[n]:y
Save the server's public key?(y/n)[n]:y
The server's public key will be saved with the name 192.168.1.1. Please wait...
<R3>                                  # The login is successful.
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Common Security Hardening Policies
2. **Examples for Security Hardening Policy Deployment**
   - STelnet Configuration
   - **Local Attack Defense Configuration**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Basic Configuration of Local Attack Defense (1)

1. Create an attack defense policy and enter the attack defense policy view.

   [Huawei] **cpu-defend policy** *policy-name*

2. Configure a blacklist.

   [Huawei-cpu-defend-policy-test] **blacklist** *blacklist-id* **acl** *acl-number*

   If multiple blacklists need to be configured in an attack defense policy, Layer 2 ACLs and basic ACLs, Layer 2 ACLs and advanced ACLs, or all the three types of ACLs cannot be used together in the blacklist group of the attack defense policy.

3. Configure the rate limit for packets sent to the CPU.

   [Huawei-cpu-defend-policy-test] **packet-type** *packet-type* **rate-limit** *rate-value*

   If a device receives attack packets or a large number of normal packets of a protocol type destined for the CPU, you can rate-limit the packets of this protocol type within a small range to reduce the impact on CPU's processing of normal services.

4. Configure the priority for packets of a specified protocol type sent to the CPU.

   [Huawei-cpu-defend-policy-test] **packet-type** *packet-type* **priority** *priority-level*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic Configuration of Local Attack Defense (2)

1. Enable active link protection.

   [Huawei-cpu-defend-policy-test] **cpu-defend application-apperceive** [**ssh** | **telnet** | **bgp** | **ftp** | **http**] **enable**

   By default, active link protection is enabled for SSH, Telnet, SSHv6, Telnetv6, FTP, BGP, and HTTP.

2. Enable attack source tracing.

   [Huawei-cpu-defend-policy-test] **auto-defend enable**

3. Set the rate threshold of attack source tracing.

   [Huawei-cpu-defend-policy-test] **auto-defend threshold** *threshold*

4. Enable the event reporting function for attack source tracing.

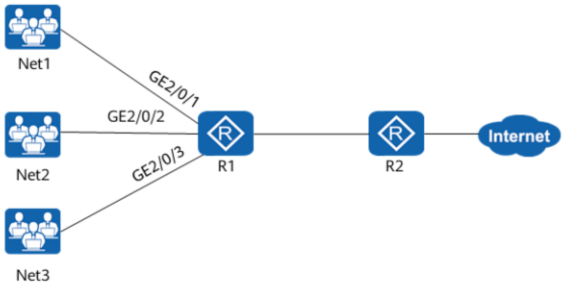
   [Huawei-cpu-defend-policy-test] **auto-defend alarm enable**

5. Apply the attack defense policy.

   [Huawei] **cpu-defend-policy** *policy-name* [**global** | **slot** *slot-id*]

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example of Local Attack Defense

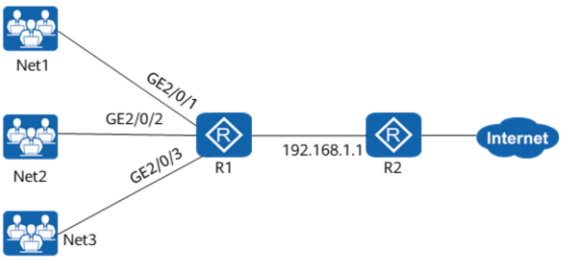


- As shown in the figure, users on different LANs access the Internet through R1. To locate attacks on R1, attack source tracing needs to be configured to trace the attack source. The existing network conditions are as follows:
  - A user on Net1 frequently initiates attacks to R1.
  - R1 receives a large number of ARP Request packets, degrading its CPU performance.
  - R1 cannot provide the FTP service.
  - LAN users obtain IP addresses through DHCP, whereas R1 does not preferentially process DHCP packets sent to the CPU.
  - R1 receives a large number of Telnet packets.
- To address these issues, proper configurations are required on R1.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Net: network

Local Attack Defense Configuration Commands (1)



1. Configure a blacklist and add the attacker (0001-c0a8-0102) on network segment Net1 to the blacklist to prevent the attacker from accessing the network.
2. Configure the rate limit for ARP Request packets sent to the CPU, reducing the impact on CPU's processing of normal services.
3. Configure active link protection for FTP so that R1 can provide the FTP service.
4. Configure a high priority for DHCP client packets so that R1 preferentially processes DHCP client packets sent to the CPU.
5. Disable the Telnet server function on R1 so that R1 discards received Telnet packets.

Configure an ACL to be referenced by a blacklist.
[R1] **acl number 4001**
[R1-acl-L2-4001] **rule 5 permit source-mac 0001-c0a8-0102**
[R1-acl-L2-4001] **quit**

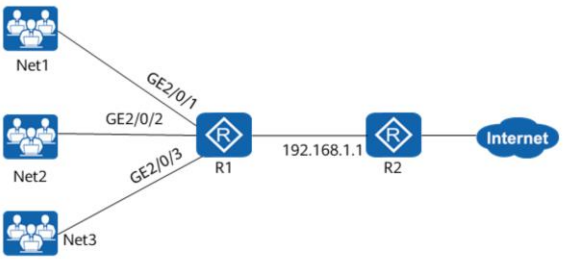
Create an attack defense policy.
[R1] **cpu-defend policy devicesafety**

Configure attack source tracing.
[R1-cpu-defend-policy-devicesafety] **auto-defend enable**
[R1-cpu-defend-policy-devicesafety] **auto-defend threshold 50**

Configure the blacklist.
[R1-cpu-defend-policy-devicesafety] **blacklist 1 acl 4001**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Local Attack Defense Configuration Commands (2)



1. Configure a blacklist and add attackers on the network segment Net1 to the blacklist to prevent users on Net1 from accessing the network.
2. Configure the rate limit for ARP Request packets sent to the CPU, reducing the impact on CPU's processing of normal services.
3. Configure active link protection for FTP so that R1 can provide the FTP service.
4. Configure a high priority for DHCP client packets so that R1 preferentially processes DHCP client packets sent to the CPU.
5. Disable the Telnet server function on R1 so that R1 discards received Telnet packets.

Set the rate limit for ARP Request packets sent to the CPU.
[R1-cpu-defend-policy-devicesafety] **packet-type arp-request rate-limit 64**

Set the rate limit for FTP packets after active link protection is enabled.
[R1-cpu-defend-policy-devicesafety] **application-apperceive packet-type ftp rate-limit 2000**
Enable active link protection for FTP.
[R1] **cpu-defend application-apperceive ftp enable**

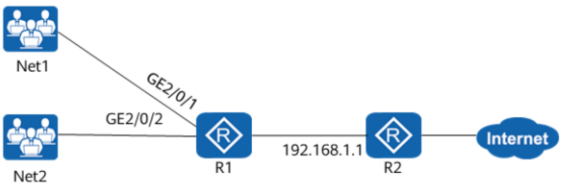
Apply the attack defense policy.
[R1] **cpu-defend-policy devicesafety**

Disable the Telnet server function.
[R1] **undo telnet server enable**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Application layer association does not need to be enabled. You only need to disable the Telnet server function on the router so that the router discards received Telnet packets.

Verifying the Local Attack Defense Configuration



Check the statistics on packets sent to the MPU. That some packets are discarded indicates that the rate limit is set for ARP Request packets.

<R1> **display cpu-defend statistics**

Packet Type	Pass Packets	Drop Packets
8021X	0	0
arp-miss	5	0
arp-reply	8090	0
arp-request	1446576	127773
bfd	0	0
bgp	0	0
bgp4plus	0	0
dhcp-client	879	0
dhcp-server	0	0

Check the configured attack defense policy.

[R1] **display cpu-defend policy devicesafety**

```
Related slot    : <0>
...
Slot<0>         : Success
Configuration :
        Blacklist 1 ACL number                    : 4001
        Packet-type arp-request rate-limit        : 64(pps)
        Packet-type dhcp-client priority          : 3
        Rate-limit all-packets                    : 2000(pps)(default)
        Application-apperceive packet-type ftp    : 2000(pps)
        Application-apperceive packet-type tftp   : 2000(pps)
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



Quiz

1. (TorF) STelnet is more secure than Telnet for remote login. This is because STelnet uses TCP encapsulation while Telnet uses UDP encapsulation.
2. (TorF) Local attack defense includes CPU attack defense and attack source tracing.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. False
2. True

Summary

- When performing security hardening on a device, take all possible dimensions into consideration, including the management plane, control plane, and forwarding plane. Based on the innate vulnerabilities of the device, assess network security risks and sort out the overall security architecture and security hardening policies for the device.
- Use secure protocols such as SSH and SFTP for device access and data transmission, unless the device does not support these. The CPU of a device is vulnerable to attacks. Therefore, local attack defense is of vital importance.

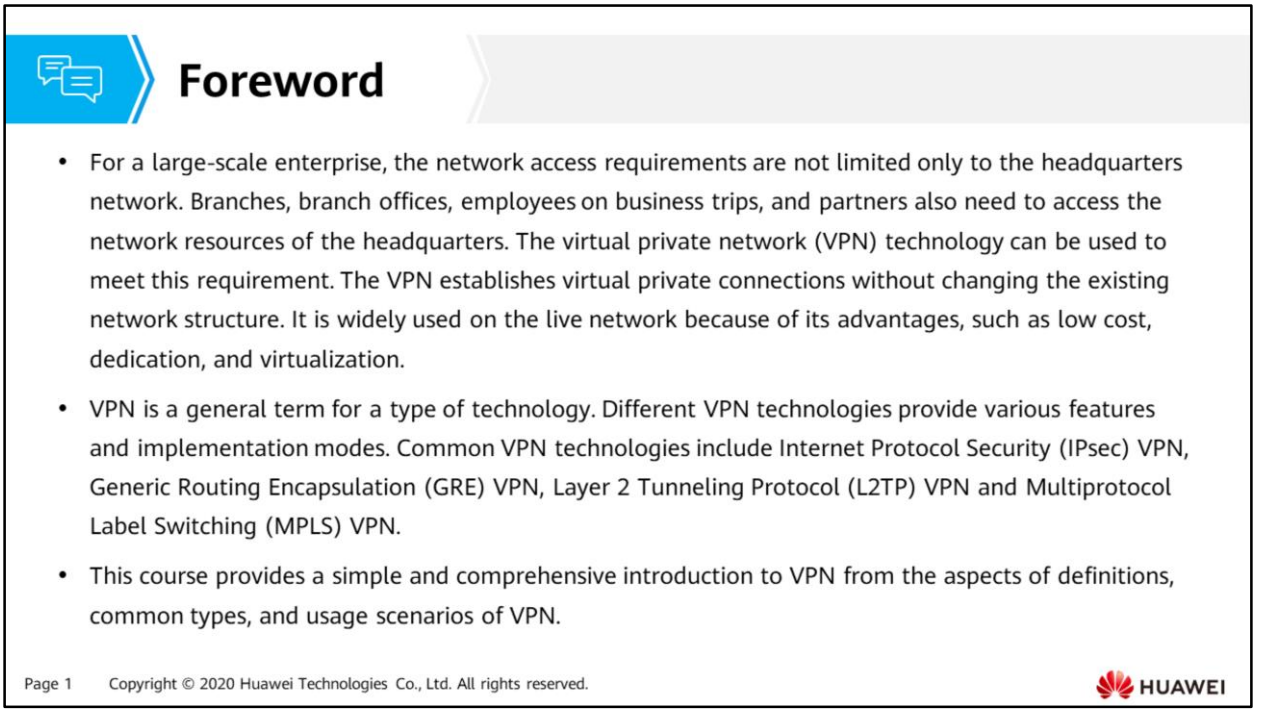
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Introduction to VPN Technology

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- IPsec: Internet Protocol Security
- GRE: Generic Routing Encapsulation
- L2TP: Layer 2 Tunneling Protocol.
- MPLS: Multiprotocol Label Switching

Objectives

- Upon completion of this course, you will be able to:
  - Understand VPN classification.
  - Understand VPN implementation modes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **What Is VPN?**
2. Common VPN Technologies

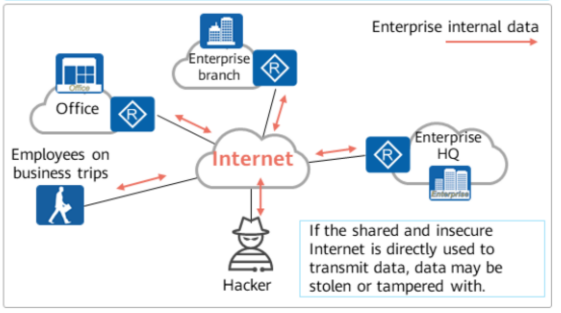
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

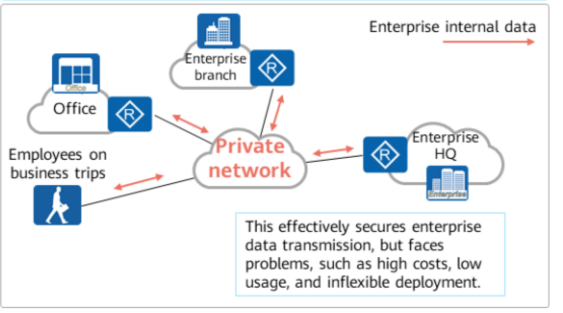
Technical Background

- Before VPN technology is used, data transmission between enterprise branches depends on the existing physical network (for example, the Internet). Packets are prone to being stolen or tampered with by hackers due to various insecure factors on the Internet. As a result, data may be disclosed or important data may be damaged.
- In addition to the Internet, a physical private network connection can be created to secure data transmission. However, the cost is high, and it is difficult to build and maintain a private network.

Using the Internet to transmit data



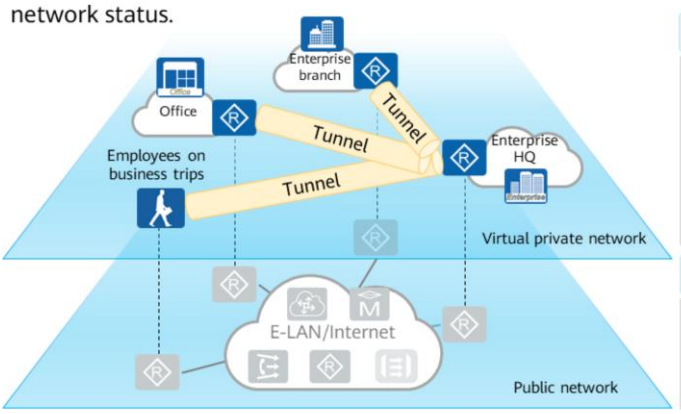
Building a private network for data transmission



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VPN Overview

A VPN is constructed on a public network using a VPN technology. VPN users transmit private network traffic on the virtual network, implementing secure and reliable connections without changing the current network status.



Private

A VPN is a network dedicated to VPN users. For VPN users, using a VPN is the same as using a traditional dedicated network. The VPN can provide sufficient security guarantee to ensure that the internal information of the VPN is not interfered by external networks. Resources of a VPN are independent of the transport network (usually an IP network). That is, VPN resources are not used by non-VPN users.

Virtual

VPN users communicate with one another through the public network. The public network can also be used by non-VPN users. VPN users obtain only a logical private network.

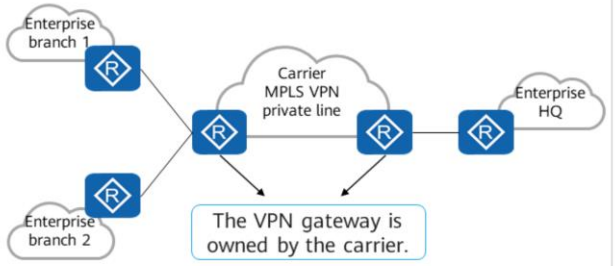
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Compared with the traditional data private network, the VPN has the following advantages:
  - Security: Reliable connections are created between the HQ and remote users, regional offices, partners, and suppliers to secure data transmission. This is particularly important for the integration of e-commerce or financial networks with communications networks.
  - Cost-effective: Public networks are used for information communication, allowing enterprises to connect remote offices, employees on business trips, and business partners at lower costs.
  - Support for mobile services: VPN users can access the network anytime and anywhere, meeting the increasing mobile service requirements.
  - Scalability: A VPN is a logical network. Adding or modifying nodes on a physical network does not affect VPN deployment.
- A public network is also called a VPN backbone network. The public network can be the Internet, a private network built by an enterprise, or a private network leased out by a carrier.

VPN Classification — Based on Construction Organizations

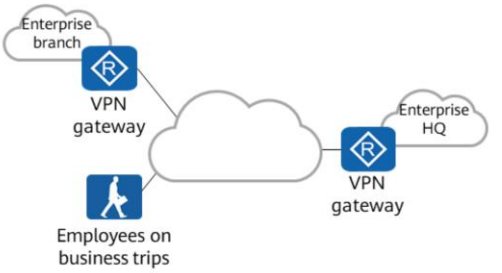
Leasing a carrier's VPN leased line to build an enterprise VPN

The most common scenario is leasing carriers' MPLS VPN private lines.



Self-built enterprise VPN network

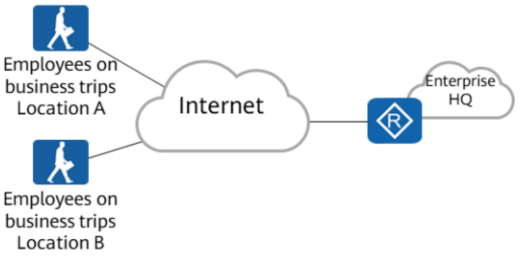
Enterprise VPN networks are established based on Internet and include the IPsec VPN, L2TP VPN, and Secure Sockets Layer (SSL) VPN.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VPN Classification — Based on Networking Modes

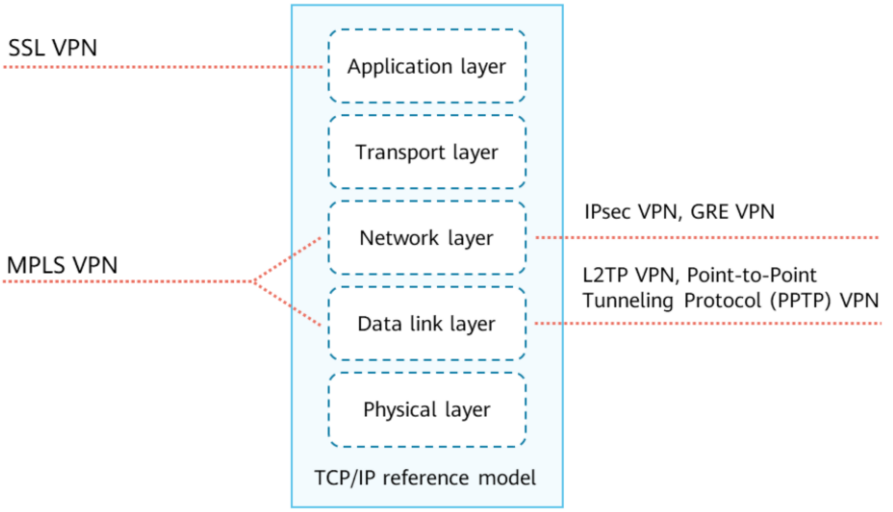
Remote access VPN	LAN-to-LAN VPN (site-to-site VPN)
This mode applies to VPN dial-up access of employees on business trips. Employees can access enterprise intranet resources through the VPN anywhere they can access the Internet. Such VPNs usually include the L2TP VPN and SSL VPN.	It is applicable to the interconnection of LANs of two geographically dispersed divisions of a company. Such VPNs usually include the MPLS VPN and IPsec VPN.





Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VPN Classification — Based on Network Layers

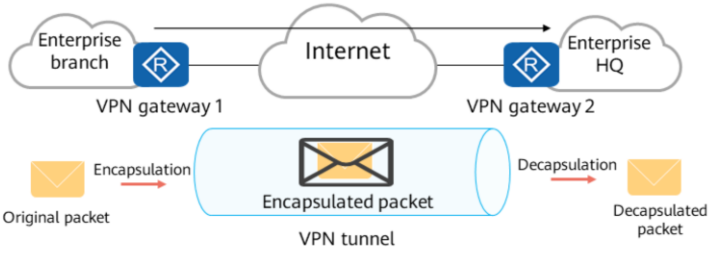


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- VPNs working at the network layer and data link layer are also called Layer 3 and Layer 2 VPNs, respectively.

Key VPN Technology — Tunneling Technologies

- The VPN technology is to use a tunneling technology to encapsulate packets, and establish a dedicated data transmission channel over a VPN backbone network to implement secure packet transmission.
- The VPN gateways on both ends establish a point-to-point virtual communication tunnel by encapsulating and decapsulating original packets.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A tunnel provides a path between two nodes so that data can be transparently transmitted along the path. A VPN tunnel is a virtual connection established between VPN nodes on a VPN backbone network to transmit VPN data. A tunnel is an indispensable part for constructing a VPN, and is used to transparently transmit VPN data from one VPN node to another.
- The tunnel is established using a tunneling protocol. Currently, there are many tunneling protocols, such as GRE and L2TP. The tunneling protocol adds a tunneling protocol header to the data on one end of the tunnel for encapsulation, so that the encapsulated data can be transmitted on a network. The tunneling protocol then removes the tunneling protocol header carried in the data on the other end of the tunnel for decapsulation. Packets are encapsulated and decapsulated before and after being transmitted within a tunnel, respectively.
- Some tunnels can be used together, for example, forming a GRE over IPsec tunnel.

Key VPN Technologies — Identity Authentication, Data Encryption, and Verification

Identity authentication, data encryption, and authentication technologies can effectively ensure the security of VPN networks and data.

- Identity authentication: This function can be used in remote VPN access scenarios. The VPN gateway authenticates user identities to ensure that only authorized users can access the network. It can also be used by VPN gateways to authenticate each other's identity.
- Data encryption: The clear text data is encrypted into ciphertext. Even if the data is intercepted by hackers, the hackers cannot obtain information in the data.
- Data verification: The data verification technology is used to check the integrity and authenticity of packets and discard the packets that are forged and tampered with.

VPN	User Identity Authentication	Encryption of Sensitive Data	Remarks
GRE	Not supported	Supports simple keyword verification, verification, and verification.	Can be used together with IPsec to encrypt and verify data.
L2TP	PPP-based Challenge-Handshake Authentication Protocol (CHAP), Password Authentication Protocol (PAP), and Extensible Authentication Protocol (EAP) authentication	Not supported	
IPsec	Supported	Supported	Supports pre-shared key authentication or certificate authentication. Internet Key Exchange version 2 (IKEv2) EAP authentication is supported.
SSL	Supported	Supported	Supports username/password or certificate authentication.
MPLS	Not supported	Not supported	Generally, it runs on a dedicated VPN backbone network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

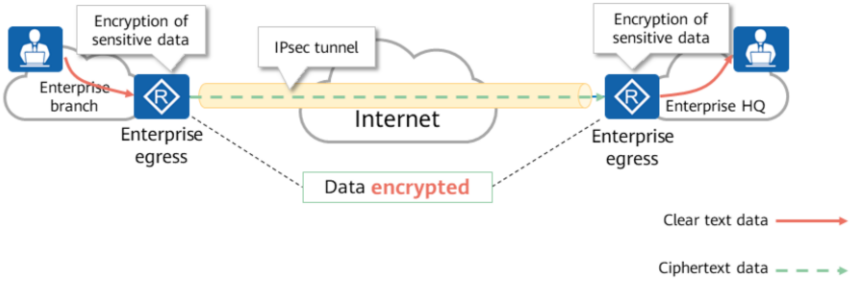
1. What Is VPN?
2. **Common VPN Technologies**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

IPsec Overview

IPsec VPN is usually deployed between egress devices of an enterprise. Through encryption and authentication, IPsec VPN implements various functions, such as data origin authentication data encryption, data integrity protection, and anti-replay.

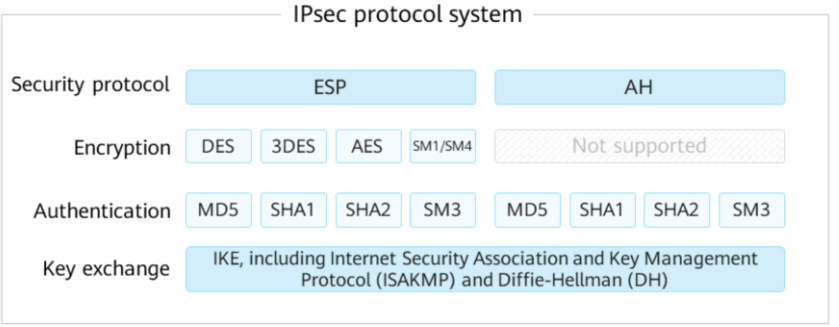


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Data origin authentication: The receiver verifies the identity of the sender.
- Data encryption: The sender encrypts data and transmits the data in ciphertext on the Internet. The receiver decrypts the received encrypted data for processing or directly forwards the data.
- Data integrity: The receiver verifies the received data to determine whether the packet has been tampered with.
- Anti-replay: The receiver rejects old or repetitive data packets to prevent malicious users from repeatedly sending obtained data.

IPsec Protocol System

IPsec is not an independent protocol. It provides a complete set of data security architecture on the IP network, including Authentication Header (AH), Encapsulating Security Payload (ESP) and Internet Key Exchange (IKE).

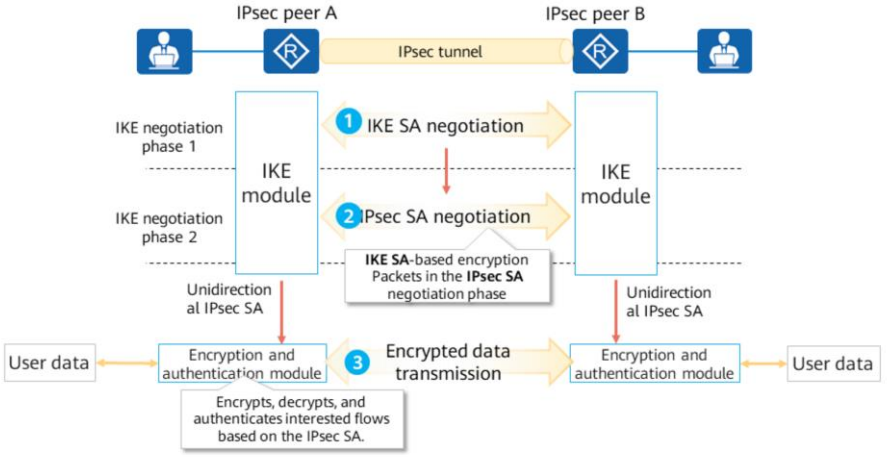


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- IPsec uses two security protocols, AH and ESP, to transmit and encapsulate data and provide security services, such as authentication and encryption.
  - The security functions provided by AH and ESP depend on the authentication and encryption algorithms used by these protocols.
  - AH supports only authentication but not encryption. ESP supports both authentication and encryption.
  - Keys are required by the security protocol that provides security services, such as authentication or encryption.
- There are two key exchange modes:
  - Out-of-band shared key: Static encryption and verification key are manually configured on the transmit and receive devices. Both parties maintain key consistency through out-of-band sharing (for example, by phone or email). The disadvantages of this mode are that poor scalability is provided and that the workload of configuring keys on a P2MP network doubles. In addition, this mode makes it difficult to change keys periodically to improve network security.
  - Automatic key negotiation through IKE: IKE is built based on the framework defined by the Internet security association (SA) and the key management protocol ISAKMP. IKE uses the DH algorithm to securely distribute keys on insecure networks. This mode is easy to configure and has good scalability, especially on large-scale dynamic networks. In addition, both communication parties exchange key exchange materials to calculate the shared key. Even if a third party intercepts all exchanged data used to calculate the key, the real key cannot be calculated.

IPsec Principles

During the establishment of an IPsec tunnel, IPsec SAs must be negotiated. Generally, an IPsec SA is generated through IKE negotiation.

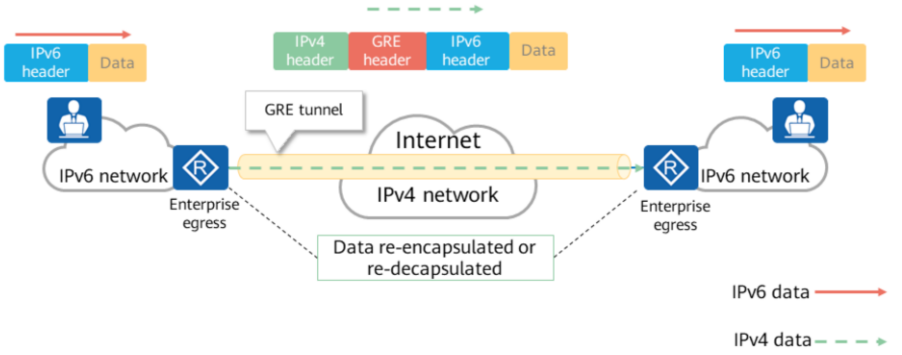


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- An SA is uniquely identified by a triplet, which consists of a security parameter index (SPI), destination IP address, and security protocol ID (AH or ESP). The SPI is a 32-bit value generated to uniquely identify an SA, and is transmitted in an AH header and an ESP header. When manually configuring the SA, you have to manually specify the SPI value. When the SA is generated through IKE negotiation, the SPI is randomly generated.
- An SA is a unidirectional logical connection. Therefore, at least two SAs must be established to protect data flows in opposite directions.
- As a key negotiation protocol, IKE has two versions: IKEv1 and IKEv2. This course uses IKEv1 as an example. For details about IKEv2, see product documentation.
    - IKEv1 negotiation phase 1 is to establish an IKE SA. After an IKE SA is established, all ISAKMP messages exchanged between IKE peers are encrypted and authenticated. This secure tunnel ensures that IKEv1 negotiation in phase 2 can be performed securely. An IKE SA is a bidirectional logical connection. Only one IKE SA is established between two IPsec peers.
    - IKEv1 negotiation phase 2 is to establish an IPsec SA for secure data transmission and derive a key for data transmission. In this phase, the key generated in phase 1 of IKEv1 negotiation is used to authenticate the integrity and identity of ISAKMP messages and encrypt these messages, securing the message exchange.
- Successful IKE negotiation indicates that a bidirectional IPsec tunnel has been established. You can define an IPsec interested flow using an ACL or an IPsec profile. All data that matches the characteristics of the interested flow is forwarded to the IPsec tunnel for processing.
- Interested flows: data flows that need to be protected by IPsec.

GRE Overview

GRE is a Layer 3 VPN encapsulation technology. GRE encapsulates packets of certain network layer protocols, such as Internetwork Packet Exchange (IPX), IPv4, and IPv6, so that the encapsulated packets can be transmitted over another network, for example, an IPv4 network. This solves the problem of packet transmission over heterogeneous networks.



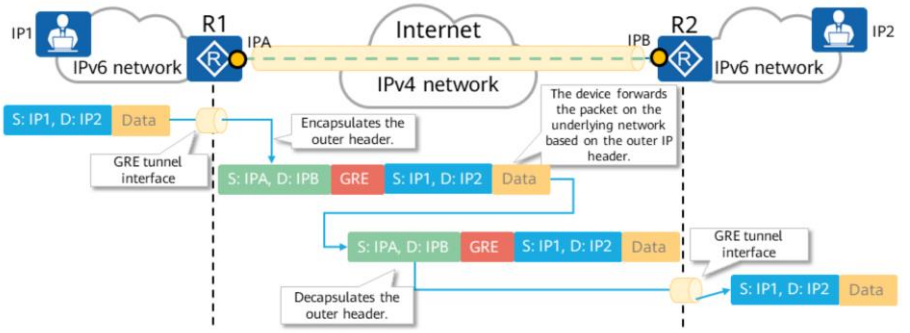
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- As shown in the figure, a GRE tunnel is established on the IPv4 network to enable communication between two IPv6 networks.
- GRE can also encapsulate multicast packets. Dynamic routing protocols use multicast packets. Therefore, GRE is often used in scenarios where multicast routing data needs to be transmitted. This is where GRE's name comes from.

GRE Principles

GRE consists of three parts: passenger protocol, encapsulation protocol, and transport protocol.

- A passenger protocol refers to an original network protocol for data transmission.
- An encapsulation protocol encapsulates the packets of the passenger protocol so that the original packets can be transmitted on the new network.
- A transport protocol transmits an encapsulated packet on the new network.

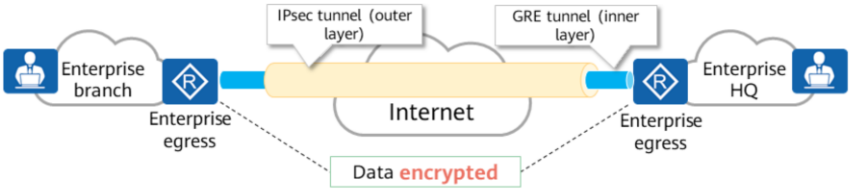


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A tunnel interface is a point-to-point virtual interface used to encapsulate packets. Similar to a loopback interface, a tunnel interface is a logical interface.
- As shown in the figure, the passenger protocol is IPv6, the encapsulation protocol is GRE, and the transport protocol is IPv4. The overall forwarding process is as follows:
    1. When R1 receives an IPv6 packet from IP1, R1 searches the routing table and finds that the outbound interface is a tunnel interface. R1 then forwards the packet to the tunnel interface.
    2. The tunnel interface adds a GRE header to the original packet and then adds an IP header to the packet based on the configuration. The source address of the IP header is the source address of the tunnel, and the destination address of the IP header is the destination address of the tunnel.
    3. The encapsulated packet is forwarded on the IPv4 network through a common IPv4 route and finally reaches the destination R2.
    4. The decapsulation process is opposite to the encapsulation process, and is not described here.

GRE over IPsec

- GRE does not support encryption or authentication, and GRE data transmission cannot be secured.
- The major disadvantage of IPsec is that it supports only the IP protocols and does not support multicast.
- GRE over IPsec can be deployed to combine the advantages of the two VPN technologies.

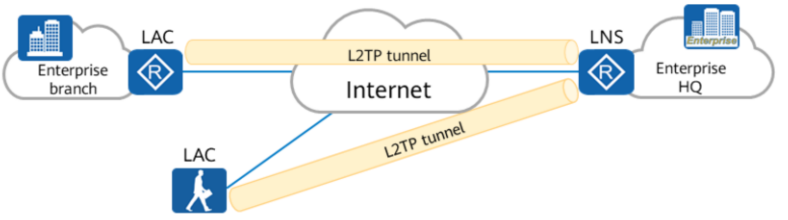


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

IPsec GRE L2TP MPLS

L2TP Overview

- L2TP is a virtual private dial-up network (VPDN) tunneling protocol that extends point-to-point (P2P) applications. L2TP provides access services for employees on business trips or enterprise branches to remotely access intranet resources.
- An L2TP network consists of L2TP access concentrators (LACs) and L2TP network servers (LNSs).



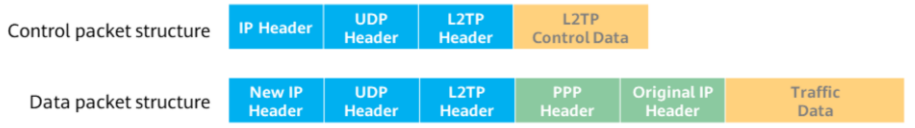
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The VPDN uses the dial-up function of the public network (such as the ISDN and PSTN) and the access network to implement the VPN and provide access services for enterprises, ISPs, and mobile office personnel. The VPDN uses the dedicated encryption-capable communication protocol to establish a secure virtual private network for enterprises on the public network. Geographically dispersed divisions and employees on business trips can remotely connect to the headquarters through virtual encrypted tunnels over the public network. However, other users on the public network cannot access the internal resources of the enterprise network through the virtual encryption tunnels. There are multiple VPDN tunneling protocols, among which L2TP is the most widely used.
- An LAC is a device that can process PPP and L2TP packets. The LAC establishes an L2TP tunnel with the LNS. The types of devices function as the LAC vary according to the networking environment. For instance, a gateway or terminal can function as the LAC. The LAC can initiate the establishment of multiple L2TP tunnels to isolate data flows.
- The LNS is the peer of the LAC, and an L2TP tunnel is established between them. The LNS is located on the border between the private and public networks of the enterprise headquarters and is usually the gateway of the enterprise headquarters.

L2TP Messages

The L2TP protocol uses two types of messages: control and data messages, which are transmitted between the LAC and LNS.

- Control messages are used to establish, maintain, and tear down L2TP tunnels and session connections.
- Data messages are used to encapsulate PPP data frames and are transmitted over tunnels.



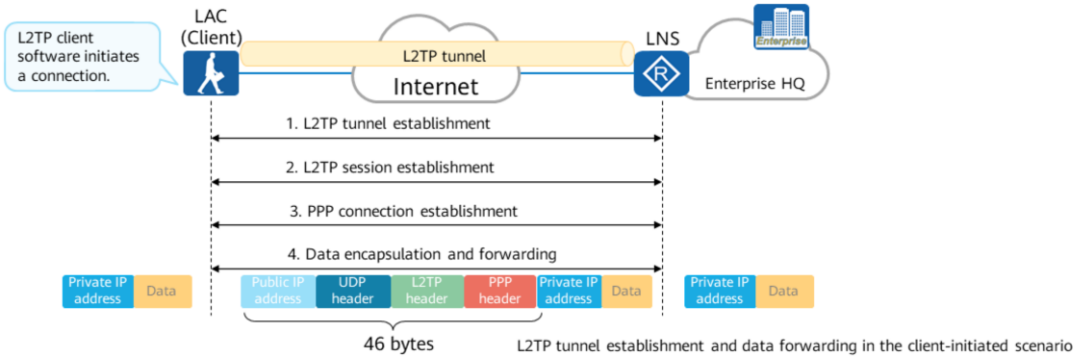
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Control message
  - It is used to establish, maintain, and tear down L2TP tunnels and session connections. During the transmission of control messages, mechanisms, such as retransmission of dropped messages and periodic detection of tunnel connectivity, are used to ensure the reliability of control message transmission. Traffic control and congestion control of control messages are supported.
  - Control messages are transmitted over the L2TP control channel. The control channel encapsulates control messages into L2TP headers and transmits them over the IP network.
- Data message
  - Encapsulates PPP data frames and are transmitted over tunnels. Data messages are not transmitted reliably. Dropped data packets are not retransmitted, and flow control and congestion control are not supported for the data messages.
  - PPP frames carried in data messages are transmitted over unreliable data channels. The PPP frames are encapsulated using L2TP and then transmitted over the IP network.

L2TP Working Process

L2TP works in the following scenarios:

- NAS-initiated scenario: Dial-up users access the enterprise intranet through the NAS.
- Client-initiated scenario: Mobile office users access the enterprise intranet.
- Call-LNS scenario: The LAC automatically dials up for enterprise intranet interconnection.



L2TP tunnel establishment and data forwarding in the client-initiated scenario

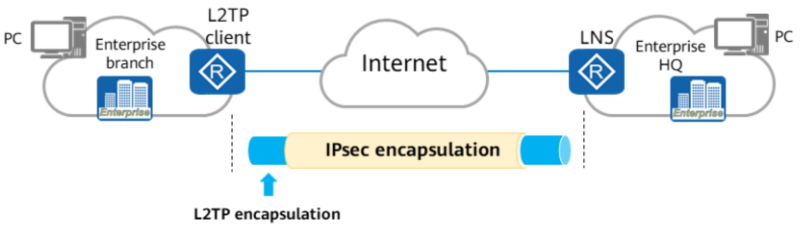
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- NAS-initiated scenario: A remote dial-up user initiates a request to establish a tunnel. The remote system dials up to log in to the LAC through the PSTN/ISDN. The LAC then initiates a request to establish a tunnel to the LNS through the Internet. The LNS assigns IP addresses to dial-up users. The authentication and accounting for remote dial-up users can be performed by the LAC proxy or LNS.
  - Users must access the Internet through PPP or PPPoE.
  - The carrier's access device (mainly a BAS device) needs to have the corresponding VPN service enabled. Users need to apply for the service from the carrier.
  - The two ends of an L2TP tunnel reside on the LAC and LNS. An L2TP tunnel can carry multiple sessions.
- Client-initialized scenario: The LAC client (local user supporting L2TP) initiates a request to establish a tunnel. The LAC client needs to know the IP address of the LNS. The LAC client can directly initiate a tunnel connection request to the LNS without passing through an independent LAC. After receiving the request from the LAC client, the LNS authenticates the LAC client based on the username and password, and assigns a private IP address to the LAC client.
  - You have to install L2TP dial-up software. Some operating systems have built-in L2TP client software.
  - There is no restriction on the Internet access mode or location, which eliminates the need of the ISP to be involved.
  - The two ends of an L2TP tunnel reside on the user and LNS sides. An L2TP tunnel carries a single L2TP session.

L2TP over IPsec

When an enterprise has high requirements on data and network security, L2TP cannot provide sufficient protection for packet transmission. In this case, L2TP can be used together with IPsec to protect transmitted data and prevent data interception or attacks.



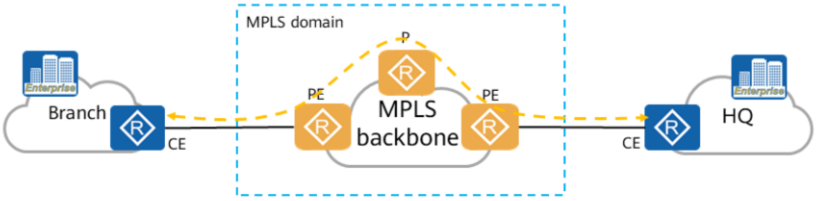
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Users on business trips communicate with the headquarters. L2TP is used to establish VPN connections, and the LNS is deployed in the headquarters to authenticate access users. When traveling users need to transmit confidential information to the headquarters, L2TP cannot provide sufficient protection for packet transmission. In this case, L2TP can be used together with IPsec to protect transmitted data. Dial-up software can be installed on the PC of a user on a business trip to encapsulate data packets through L2TP and then IPsec, and then send the packets to the headquarters. IPsec policies are deployed on the headquarters gateway to restore data. In this mode, IPsec protects all packets that originate at the LAC and are destined for the LNS.

IPsec GRE L2TP MPLS

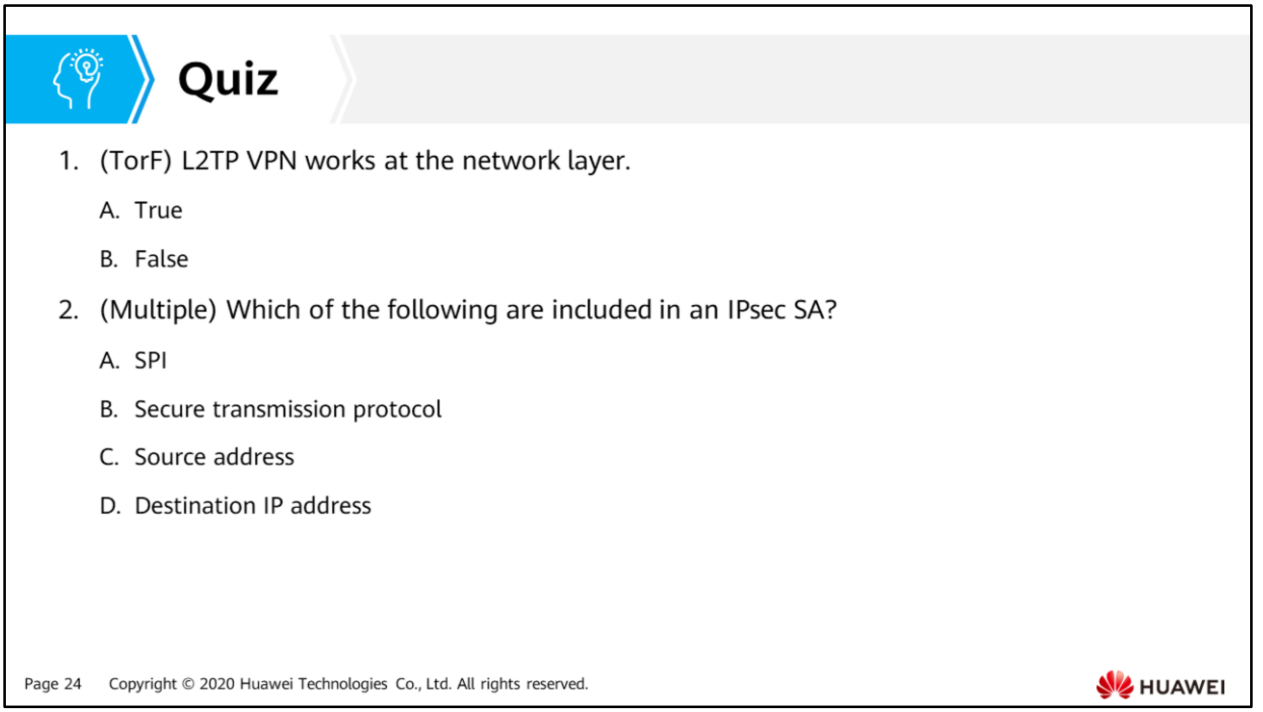
MPLS VPN Overview

- Multiprotocol Label Switching (MPLS) uses labels to forward packets. It was originally proposed to improve the forwarding rate of IP packets. Currently, MPLS is mainly used in VPN, traffic engineering, and QoS scenarios.
- MPLS VPNs are classified into the MPLS L2VPN and MPLS L3VPN.
- Enterprises can build their own MPLS networks or lease carriers' MPLS networks to obtain MPLS VPN access services.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- An MPLS VPN is usually constructed by a carrier. VPN users purchase VPN services to implement route transmission and data exchange between user networks (branches and headquarters shown in the figure).
- A basic MPLS VPN consists of customer edge (CE), provider edge (PE), and provider (P) devices.
  - CE: an edge device on the user network. A CE has interfaces that are directly connected to a carrier network. A CE can be a router, switch, or host. Generally, CEs are unaware of VPNs and do not need to support MPLS.
  - PE: an edge device on a carrier network and is directly connected to a CE. On an MPLS network, VPN processing is performed on PEs, which poses high requirements on PE performance.
  - P: a backbone router on the carrier's network and is not directly connected to a CE. Ps only need to have basic MPLS forwarding capabilities and do not need to maintain VPN information.
- For more information about BGP/MPLS IP VPN, see materials of related HCIP-Datacom-Advance courses.



- Answers:
  - 1. B
  - 2. ABD

Summary

- The VPN technology provides a series of advantages, such as security, low cost, support for mobile services, and flexibility, and has become the most widely deployed technology on the live network.
- This course briefly introduces the following common VPN technologies based on the basic concepts and classifications of VPNs:
  - IPsec VPN: a set of protocols and services that provide secure transmission for IP networks.
  - GRE VPN: provides a mechanism to encapsulate packets of a protocol into packets of another protocol, solving the transmission problem of heterogeneous networks.
  - L2TP VPN: a Layer 2 VPN technology that provides remote access services for mobile users.
  - MPLS VPN: a VPN technology that implements interconnection between sites through label switching.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Introduction to VRF

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- Looking back on the previous courses, would you recall certain protocols or technologies whose names contain the letter "V"? What are the characteristics or similarities of these protocols or technologies?
- Virtual routing and forwarding (VRF) technology creates multiple routing tables on a Layer 3 forwarding device to isolate data or services. This technology is often used in isolation-related scenarios, such as when Multiprotocol Label Switching (MPLS) VPN and firewalls are deployed.
- This course describes basic concepts and configurations of VRF technology.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Understand basic VRF concepts.
  - Understand typical VRF applications.
  - Implement basic VRF configurations.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

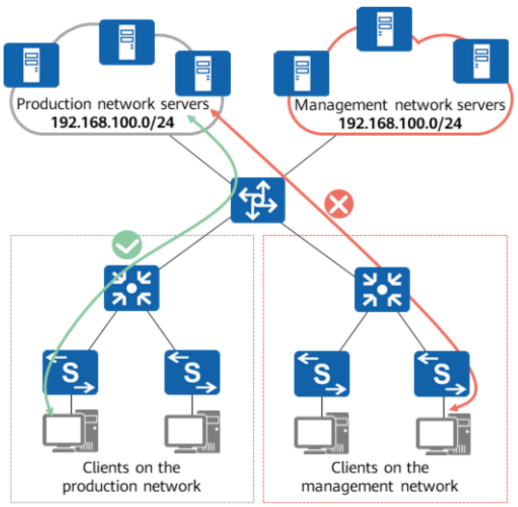
Contents

1. **Basic Concepts of VRF**
2. Typical VRF Configuration Examples

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

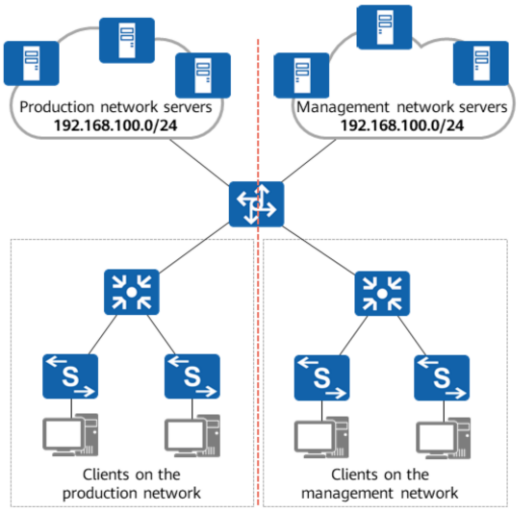
Network Requirements



- An enterprise network has a production network and a management network. Each network exclusively uses an aggregation switch and access switches and shares a core switch with the other one.
- The core switch connects to a server cluster on each of the production and management networks. The two network segments are 192.168.100.0/24.
- Requirements: Implement data communication within the production and management networks and isolate the communication between the two networks.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

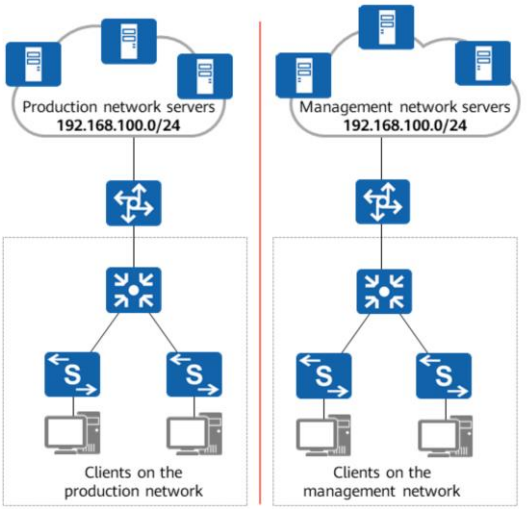
Implementation by Deploying ACLs



- Configure an ACL on the core switch to prevent mutual access between the production and management networks.
- Disadvantages:
  - The configuration is complex, and scalability is poor.
  - The problem that two networks use overlapping network segments cannot be resolved. Additional configurations must be performed to prevent the use of overlapping network segments.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

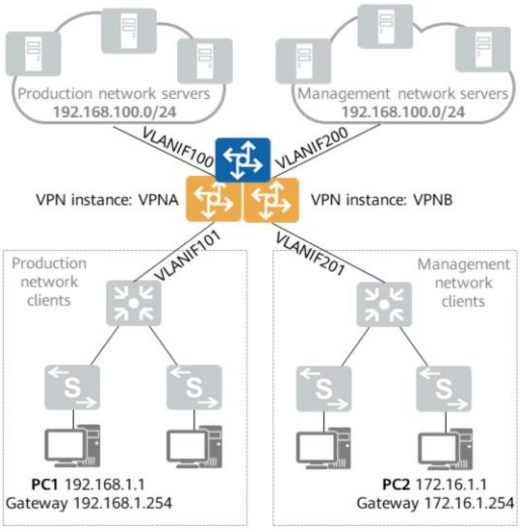
Adding a Core Switch



- Add a core switch so that an exclusive core switch serves each network to physically isolate the two networks.
- Disadvantage: Additional equipment expenditures are required.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Using VRF Technology



- VRF, also called VPN instance, is a virtualization technology. Multiple VPN instances are created on a physical device. Each VPN instance has its own interfaces, a routing table, and routing protocol processes.

VPNA's routing table

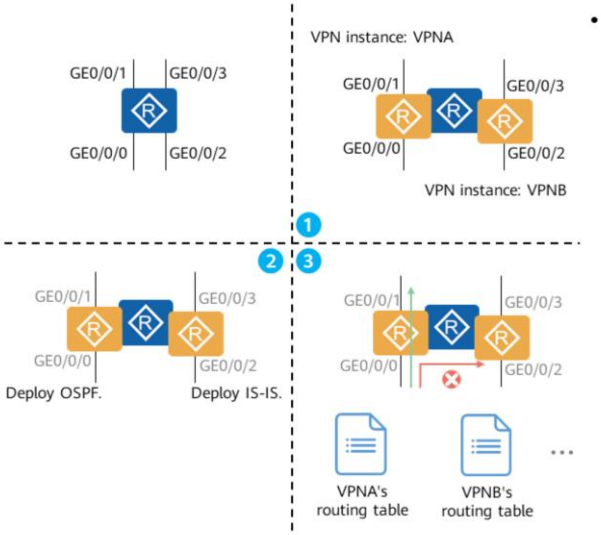
Destination/ Mask	Protocol	Interface	NextHop
192.168.100.0/24	OSPF	VLANIF100	192.168.100.1
192.168.1.0/24	OSPF	VLANIF101	192.168.101.1

VPNB's routing table

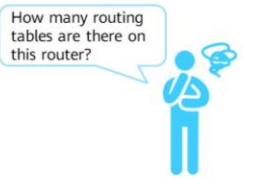
Destination/ Mask	Protocol	Interface	NextHop
192.168.100.0/24	OSPF	VLANIF200	192.168.100.1
172.16.1.0/24	Static	VLANIF201	192.168.201.1

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VRF Implementation



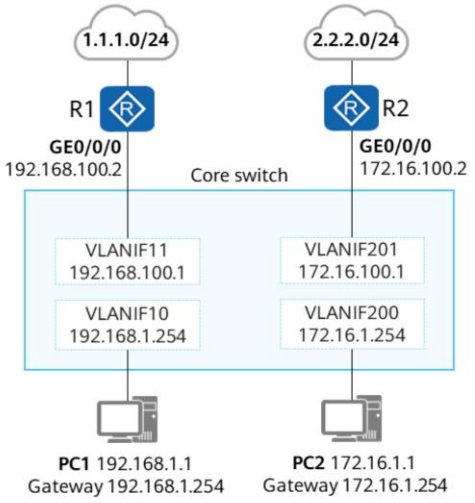
- A VRF is a logical unit on a physical device. Each logical unit is called a VPN instance, and instances are isolated on the routing plane. VRF implementation is as follows:
  1. Create an instance and bind each involved Layer 3 interface (physical interface, sub-interface, or VLANIF interface) to the instance.
  2. (Optional) Configure a routing protocol or static route and bind it to the VPN instance.
  3. A VPN instance routing table is created using information about the interfaces and routing protocols bound to the instance. Then data is forwarded using entries contained in the instance routing table, and instance isolation is implemented.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- By default, all interfaces on a network device belong to the same forwarding instance, that is, the root instance of the device.

Example: Before VRF Deployment



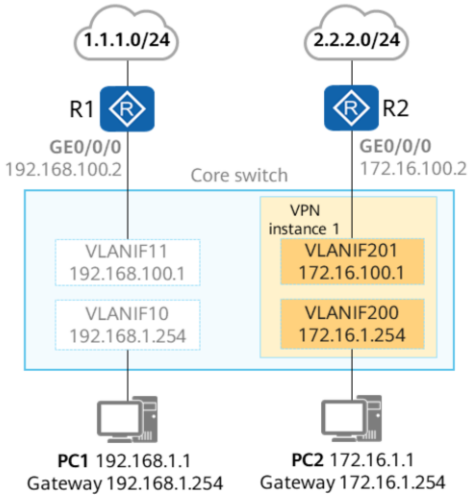
Core switch's routing table

Destination/Mask	Protocol	Interface	NextHop
192.168.1.0/24	Direct	VLANIF10	192.168.1.254
172.16.1.0/24	Direct	VLANIF200	172.16.1.254
192.168.100.0/24	Direct	VLANIF11	192.168.100.1
172.16.100.0/24	Direct	VLANIF201	172.16.100.1
1.1.1.0/24	Static	VLANIF11	192.168.100.2
2.2.2.0/24	Static	VLANIF201	172.16.100.2

Background: PC1, R1, and network segment 1.1.1.0/24 connected R1 belongs to a service network.
PC2, R2, and network segment 2.2.2.0/24 connected to R2 belong to the management network. All direct routes on the core switch and routes destined for remote networks and discovered by the core switch are stored in the core switch's routing table (also called the global routing table). With such a routing table, the core switch enables the service and management networks to communicate.
Requirements: Configure the core switches to isolate the service and management networks.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Example: Creating VPN Instances



Core switch's global routing table
(Root device's routing table)

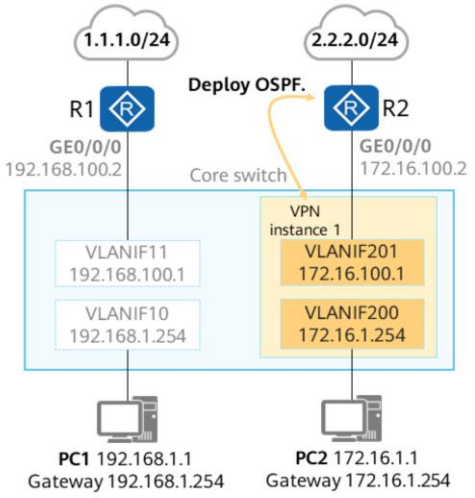
Destination/Mask	Protocol	Interface	NextHop
192.168.1.0/24	Direct	VLANIF10	192.168.1.254
192.168.100.0/24	Direct	VLANIF11	192.168.100.1
1.1.1.0/24	Static	VLANIF11	192.168.100.2

VPN instance 1's routing table on the core switch

Destination/Mask	Protocol	Interface	NextHop
172.16.1.0/24	Direct	VLANIF200	172.16.1.254
172.16.100.0/24	Direct	VLANIF201	172.16.100.1

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Example: Configuring a Dynamic Routing Protocol



Core switch's global routing table
(Root device's routing table)

Destination/Mask	Protocol	Interface	NextHop
192.168.1.0/24	Direct	VLANIF10	192.168.1.254
192.168.100.0/24	Direct	VLANIF11	192.168.100.1
1.1.1.0/24	Static	VLANIF11	192.168.100.2

VPN instance 1's routing table on the core switch

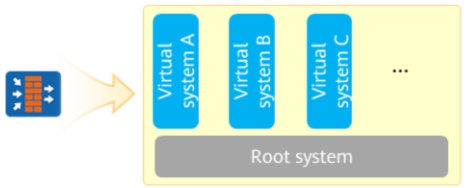
Destination/Mask	Protocol	Interface	NextHop
172.16.1.0/24	Direct	VLANIF200	172.16.1.254
172.16.100.0/24	Direct	VLANIF201	172.16.100.1
2.2.2.0/24	OSPF	VLANIF201	172.16.100.2

Either multiple dynamic routing protocols or multiple processes of a specific dynamic routing protocol can run on a device. A dynamic routing protocol process associated with a VPN instance serves the instance, and the device installs the routes learned from the process to the routing table of the VPN instance.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Common Usage Scenarios

Firewall Virtual System

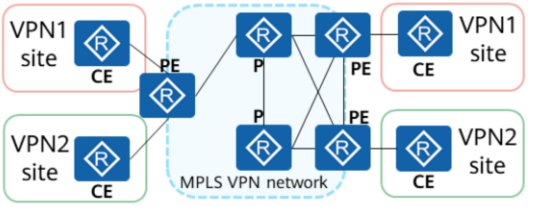


Virtual systems (VSs) are logical devices divided from a physical device and are working separately. A virtual system has the following characteristics:

- Resource virtualization: Each VM has exclusive resources, including interfaces, VLANs, policies, and sessions.
- Route virtualization: Each VM has its own routing table and is isolated from the other VMs.

Route virtualization is implemented by creating VPN instances.

BGP/MPLS IP VPN



BGP/MPLS IP VPN is a PE-based L3VPN technology. It runs BGP to advertise VPN routes and MPLS to forward VPN packets over the backbone network of a service provider (SP).
VPN instances are created on PEs to distinguish routes of different VPNs.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- For more information about BGP/MPLS IP VPN, see the related HCIP-Datacom-Advance courses.

Contents

1. Basic Concepts of VRF
2. **Typical VRF Configuration Examples**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Basic VRF Configuration Commands (1)

1. Create a VPN instance and enter the VPN instance view.

   [Huawei] **ip vpn-instance** *vpn-instance-name*

   The **ip vpn-instance** command creates a VPN instance and displays the VPN instance view. By default, no VPN instance is configured.

2. Enable IPv4 route advertisement and data forwarding for the VPN instance.

   [Huawei-vpn-instance-InstanceName] **ipv4-family**

   The **ipv4-family** command creates an IPv4 address family for a VPN instance and displays the VPN instance IPv4 address family view. By default, no IPv4 address family is created for a VPN instance. An interface cannot be bound to a VPN instance that has no address family created.

3. Bind an interface to the VPN instance.

   [Huawei-GigabitEthernet0/0/0]**ip binding vpn-instance** *vpn-instance-name*

   The **ip binding vpn-instance** command binds an interface on a PE to a VPN instance. By default, an interface is not bound to any VPN instance and belongs to the root instance. After an interface is bound to a VPN instance or unbound from a VPN instance, the IP address, Layer 3 features, and IP-related routing protocols on the interface are deleted. You have to reconfigure them if needed.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic VRF Configuration Commands (2)

4. Add a static route to the routing table of the VPN instance.

   [Huawei] **ip route-static vpn-instance** *vpn-instance-name ip-address* { *mask* | *mask-length* } { *nexthop-address* | *interface-type interface-number* }

5. Create a dynamic routing protocol process (for example, OSPF) bound to the VPN instance.

   [Huawei] **ospf** [*process-id* | **router-id** *router-id*] **vpn-instance** *vpn-instance-name*

   Note: The process IDs of different VPN instances must be different.

6. Use the following commands to maintain VPN instances.

   [Huawei] **display ip routing-table vpn-instance** *vpn-instance-name*

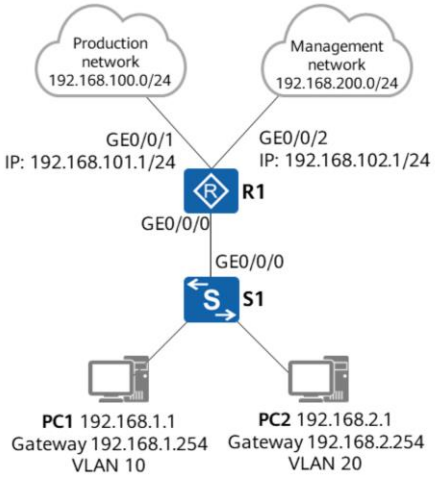
   [Huawei] **ping -vpn-instance** *vpn-instance-name host*

   [Huawei] **tracert -vpn-instance** *vpn-instance-name host*

   Note: If the **ping** command does not carry the **vpn-instance** keyword, the ping is performed on the root device by default, and the generated ICMP messages are forwarded using entries in the global routing table, which also applies to the tracert operation.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Case — Background and Requirement Analysis



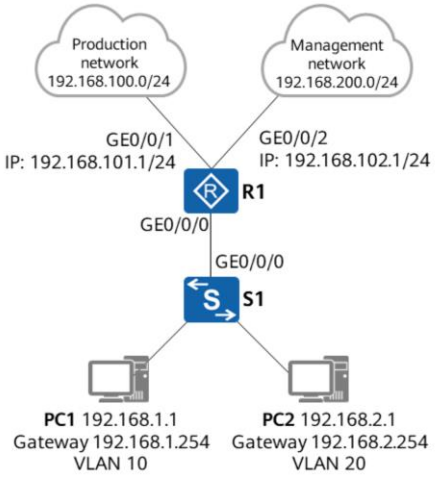
- **Background**: An enterprise has a production network and a management network, shown on the left figure.
  - PC1 belongs to the production network, PC2 belongs to the management network, and S1 is the access switch for the two PCs. PC1 is added to VLAN 10, and PC2 is added to VLAN 20.
  - Sub-interfaces are configured on R1 functioning as a gateway for PC1 and PC2.
- **Requirements**: Isolate the production network from the management network and use static routes for communication within each network.

Destination/Mask	Protocol	Interface	NextHop
192.168.100.0/24	Static	GE0/0/1	192.168.101.254
192.168.200.0/24	Static	GE0/0/2	192.168.102.254
192.168.1.0/24	Direct	GE0/0/0.1	192.168.1.254
192.168.2.0/24	Direct	GE0/0/0.1	192.168.2.254

Routing table before VRF is deployed on R1

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example — Procedure (1)



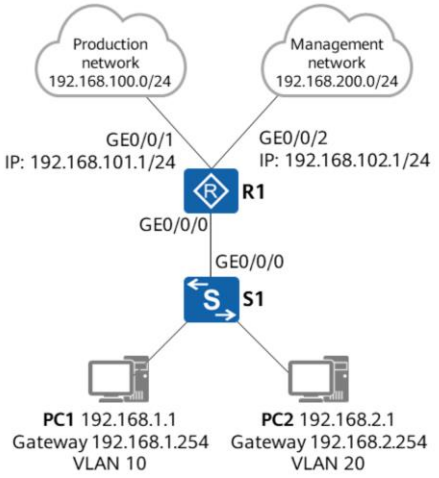
1. Configure VLANs on the switch. The configuration details are not provided.

2. Create VPN instances for the production and management networks and configure the IPv4 address family for each VPN instance.

```
[R1]ip vpn-instance production
[R1-vpn-instance-production]ipv4-family
[R1-vpn-instance-production-af-ipv4]quit
[R1-vpn-instance-production]quit
[R1]ip vpn-instance management
[R1-vpn-instance-management]ipv4-family
[R1-vpn-instance-management-af-ipv4]quit
[R1-vpn-instance-management]quit
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example — Procedure (2)

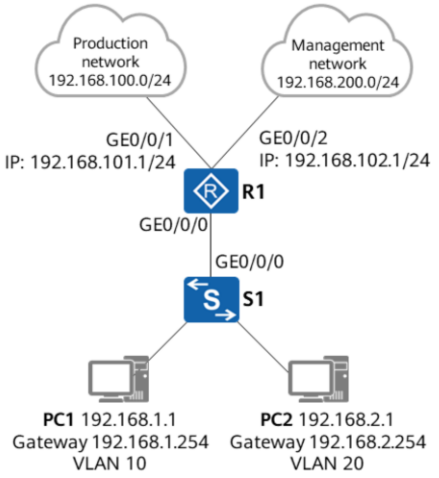


3. Bind involved interfaces to the VPN instances.

```
[Huawei]interface GigabitEthernet 0/0/0.1
[Huawei-GigabitEthernet0/0/0.1]ip binding vpn-instance production
Info: All IPv4 related configurations on this interface are removed!
//After an interface is bound to an instance, the IP address of the interface is deleted and has to be reconfigured.
[Huawei-GigabitEthernet0/0/0.1]ip address 192.168.1.254 24
[Huawei-GigabitEthernet0/0/0.1]quit
[Huawei]interface GigabitEthernet 0/0/0.2
[Huawei-GigabitEthernet0/0/0.2]ip binding vpn-instance management
Info: All IPv4 related configurations on this interface are removed!
[Huawei-GigabitEthernet0/0/0.2]ip address 192.168.2.254 24
[Huawei-GigabitEthernet0/0/0.2]quit
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example — Procedure (3)



4. Configure static routes and bind them to the instance.

```
[R1]ip route-static vpn-instance production 192.168.100.0 24
192.168.101.254
[R1]ip route-static vpn-instance management 192.168.200.0 24
192.168.102.254
```

5. Verify the configuration.

```
[R1] display ip routing-table vpn-instance production
Route Flags: R - relay, D - download to fib
------------------------------------------------------------------------------
Routing Tables: production
         Destinations : 5        Routes : 5
Destination/Mask  Proto  Pre Cost      Flags NextHop         Interface
192.168.100.0/24  Static 60  0           RD    192.168.101.254 GigabitEthernet0/0/1
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Single) At which of the following layers can VRF technology be used to implement isolation?
   A. Physical layer
   B. Network layer
   C. Data link layer
   D. Application layer
2. (TorF) By default, all Layer 3 interfaces on Huawei data communication products belong to the root instance.
   A. True
   B. False

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. B
2. A

Summary

- VRF technology implements logical isolation between different networks on the same physical device. When multiple VRF instances are deployed on the physical device, each VRF instance is equivalent to a virtual network device. The interfaces and routes between VRF instances are isolated. When a physical device is connected to the same network segment that belongs to different VPN instances, you have no worries about IP address conflicts.
- The VRF technology is widely used in firewall virtual systems and BGP/MPLS IP VPN scenarios.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



BFD Implementation and Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- As network applications are widely deployed, network faults may cause service anomalies. To minimize the impact of device faults on services and improve network reliability, a network device must be able to quickly detect faults in communication with adjacent devices. Measures can then be taken to promptly rectify the faults to ensure service continuity.
- Bidirectional Forwarding Detection (BFD) is a fast fault detection mechanism that rapidly detects link faults and monitors IP connectivity on the entire network independent of media and routing protocols.
- This document describes the implementation and typical application scenarios of BFD.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe functions of BFD.
  - Describe BFD implementation.
  - Configure and apply BFD.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

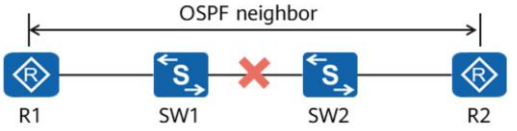
1. **Introduction to BFD**
2. BFD Implementation
3. BFD Application Scenarios
4. Basic BFD Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

You May Encounter in Problems in Network Fault Detection

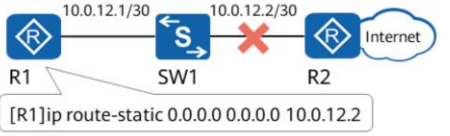
- In a system where hardware signals cannot be used to detect faults, applications usually use the Hello mechanism of an upper-layer routing protocol to detect network faults.
- The detection duration is more than 1 second, which is too long for some applications. When data transmitted on the network exceeds 1 Gbit/s, the second-level detection will cause a large amount of data to be lost.
- On a Layer 3 network, static routes do not have a fault detection mechanism.

Dynamic Route Fault Detection



R1 and R2 establish an OSPF neighbor relationship through a Layer 2 switch. Can R1 and R2 quickly detect a fault on the link between SW1 and SW2?

Static Route Fault Detection



On R1, a default route to the gateway R2 is configured. Can R1 quickly detect the fault on the link between SW1 and R2?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BFD Overview

- BFD provides a universal, standard, media-independent, and protocol-independent fast fault detection mechanism. It has the following advantages:
  - Uses a low-overhead, short-duration method to detect faults in a path between adjacent forwarding engines.
  - Provides uniform detection for all media and protocol layers in real time.
- BFD is a simple Hello protocol. A BFD session channel is established between two systems and BFD packets are periodically sent over the channel. If a system does not receive BFD packets from the remote end within a specified period, the system considers that a fault occurs on the channel.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Only one BFD session can be established on a data path. If different applications need to use different BFD protocol parameters on the same data path, you can configure a unique BFD session by using the parameters that match needs of all applications. The BFD session status change is reported to all bound applications.

Contents

1. Introduction to BFD
2. **BFD Implementation**
3. BFD Application Scenarios
4. Basic BFD Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BFD Packet Format

- BFD is implemented by maintaining the BFD session between two systems. The system establishes a session by sending BFD packets to each other.
- BFD control packets are encapsulated differently based on scenarios. A BFD control packet consists of mandatory fields and optional authentication fields.

Mandatory fields

Ver	Diag	Sta	P	F	C	A	D	M	Detect Mult	Length
My discriminator										
Your discriminator										
Desired Min TX Interval										
Required Min RX Interval										
Required Min Echo RX Interval										

Optional authentication fields

Auth-Type	Auth-Len	Authentication Data

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

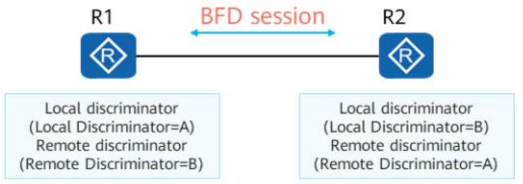
- Sta: indicates the status of the local BFD system.
- Detect Mult: indicates the detection multiplier flag. It is used by the detector to calculate the detection timeout interval.
- My Discriminator: indicates the local discriminator of the BFD session. It is a unique non-zero value generated by the transmitting system. Local discriminators are used to distinguish multiple BFD sessions in a system.
- Your Discriminator: indicates the remote discriminator of the BFD session. If this value is received from the remote system, the value of the received My Discriminator field is used. If this value is unknown, the system returns value 0.
- Desired Min Tx Interval: indicates the minimum interval for sending BFD packets on the local end.
- Required Min RX Interval: indicates the minimum interval for receiving BFD packets on the local end.
- Required Min Echo RX Interval: indicates the minimum interval for receiving Echo packets on the local end. If the local end does not support the Echo function, the field is set to 0.

BFD Session Setup

BFD sessions can be established in either static or dynamic mode. BFD distinguishes sessions according to local and remote discriminators of control packets. Static and dynamic BFD sessions differ in that local and remote discriminators are configured in different modes.

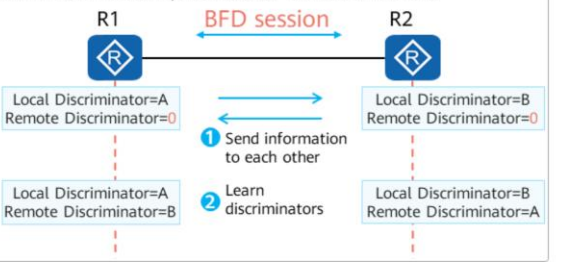
Static Establishment

BFD session parameters, including the local and remote discriminators, are manually specified. A request for BFD session establishment is distributed manually.



Dynamic Establishment

The local discriminator for dynamically establishing a BFD session is dynamically allocated by the system that triggers the establishment of the BFD session. The remote discriminator is learned from the value of Local Discriminator in the BFD packet received from the remote end.

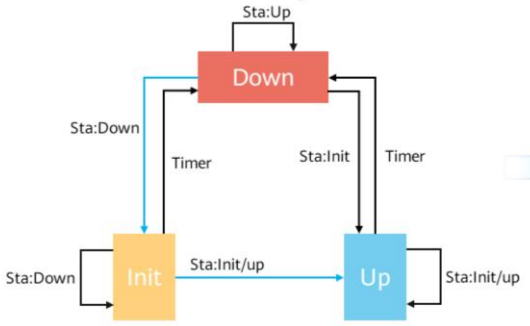


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

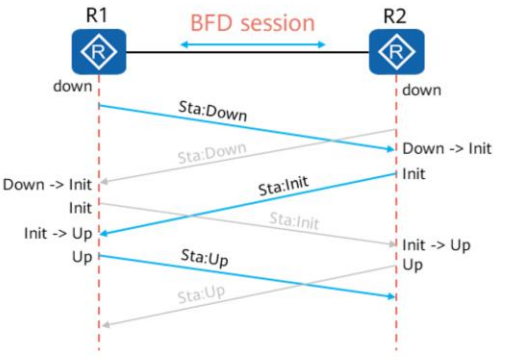
- When a BFD session is set up dynamically, the system processes the local and remote discriminators as follows:
  - Dynamic allocation of the local discriminator: When an application triggers setup of a dynamic BFD session, the system allocates a dynamic local discriminator within a specified range to the BFD session. Then the local system sends a BFD control packet with the remote discriminator of 0 to the remote system for BFD session negotiation.
  - Automatically learning the remote discriminator: When one end of a BFD session receives a BFD control packet with the remote discriminator of 0, this end checks whether the packet matches parameters the BFD session. If the packet matches parameters of the BFD session, this end learns the value of Local Discriminator in the received control BFD packet and obtains the remote discriminator.

States of BFD Sessions

A BFD session has the following states: Down, Init, Up, and AdminDown. The State field of a BFD control packet shows the session status. The system changes the BFD session status based on the local session status and the received BFD session status of the remote end. The BFD state machine implements a three-way handshake for BFD session setup or deletion to ensure that the two systems detect the status change.



State transition of a BFD session



State transition during BFD session establishment

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A BFD session often involves three states. The BFD sessions are established in Init and Up states, and the Down state indicates that the BFD session is terminated. A three-way handshake is required for establishing and disconnecting a BFD session to ensure that both systems can detect the setup or disconnection. The AdminDown state is special and means that the **shutdown** command is configured in the BFD session view. Each system sends the local status through the State field in outgoing BFD control packets, and learns the remote end status through the State field in the received BFD control packets.
- The Down state indicates that the BFD session is Down. A BFD session remains in Down state until the local end receives a packet from the remote end, where the State field indicates that the remote end is not in Up state. If a BFD control packet with the State field set to Down is received, the state machine transits from the Down state to the Init state. If a BFD control packet with the State field set to Init is received, the state machine transits from the Down state to the Up state. If a BFD control packet with the State field set to Up is received, the state machine maintains the Down state.
- The Init state indicates that the local end is communicating with the remote end and the local session is expected to go Up, but the remote end does not respond. A BFD session in Init state transitions to the Up state until it receives a BFD control packet with the State field set to Init or Up from the remote end. Otherwise, the BFD session enters the Down state after the detection timer expires, indicating that the communication with the remote end is terminated.
- The Up state indicates that the BFD session is successfully established and the link

connectivity is being checked. The BFD session remains Up until the link is faulty or the **shutdown** command is configured in the BFD session view. If the local end receives a BFD control packet with the State field set to Down from the remote end or the detection timer expires, the BFD session changes from Up to Down.

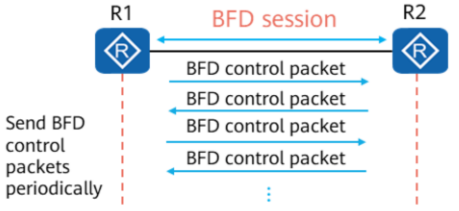
- The AdminDown state indicates that the remote system enters the Down state and remains in Down state until the local system exits the AdminDown state. The AdminDown state does not mean that the forwarding path is unreachable.

BFD Detection Modes

Two systems establish a BFD session and periodically send BFD control packets along the path between them. If one system does not receive BFD control packets within a specified period, the system considers the path faulty. BFD detection modes include asynchronous and demand modes.

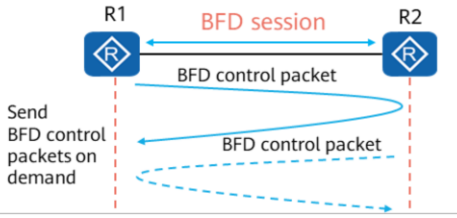
Asynchronous Mode

In asynchronous mode, two systems periodically send BFD control packets to each other. If a system does not receive any BFD control packet from the remote end within the detection period, the system advertises that the BFD session is Down.



Demand Mode

In a scenario where connectivity needs to be verified, the system sends multiple BFD control packets consecutively. If it does not receive any return packets within a specific period of time, it considers the BFD session Down.

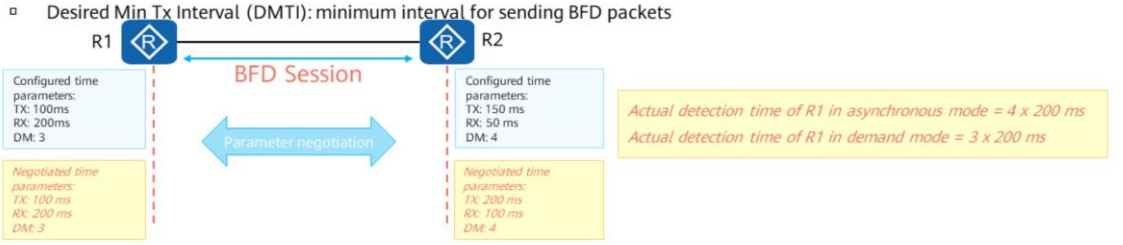


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The asynchronous mode differs from the demand mode in the detection location. In asynchronous mode, the local end sends BFD control packets at a given period of time. The detection location is the remote end. The remote end detects whether the local end periodically sends BFD control packets. In demand mode, the local end checks whether there is a response packet for the BFD control packet sent by itself.

BFD Detection Time

- The detection time of a BFD session is determined by the local detection multiplier (Detect Multi), minimum interval for receiving BFD packets (Required Min RX Interval), and minimum interval for sending BFD packets (Desired Min TX Interval). The actual interval at which BFD packets are received or sent depend on BFD session negotiation.
- Detection time = Received Detect Multi of the remote system x Max (Local RMRI/Received DMTI)
  - Detect Multi: local detection multiplier
  - Required Min Rx Interval (RMRI): minimum interval for receiving BFD packets
  - Desired Min Tx Interval (DMTI): minimum interval for sending BFD packets

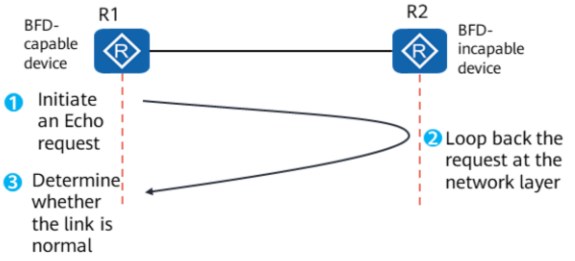


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Default BFD time parameters
  - By default, the interval for sending BFD packets is 1000 ms, the interval for receiving BFD packets is 1000 ms, and the local detection multiplier is 3.
  - The WTR time of a BFD session is 0, and the delay for the BFD session to go Up is 0.
- The detection timeout multiplier, used by the detecting party to calculate the detection timeout interval:
  - Demand mode: The local detection multiplier takes effect.
  - Asynchronous mode: The remote detection multiplier takes effect.

BFD Echo

- BFD Echo is a detection mechanism in which the local system sends BFD echo packets and the remote system loops back the packets.
- One of the two directly connected devices supports BFD (R1), and the other device does not support BFD (R2) and supports only basic forwarding at the network layer. To quickly detect faults between the two devices, you can create a one-arm BFD Echo session on the BFD-capable device. The BFD-capable device initiates an Echo request. After receiving the Echo request, the BFD-incapable device loops back the Echo request to implement connectivity detection on forwarding links.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

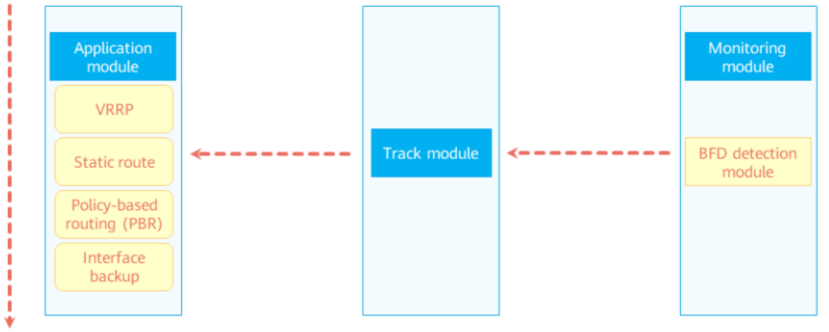
1. Introduction to BFD
2. BFD Implementation
3. BFD Application Scenarios
4. Basic BFD Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BFD Association

Association functions involve the detection module, track module, and application module.

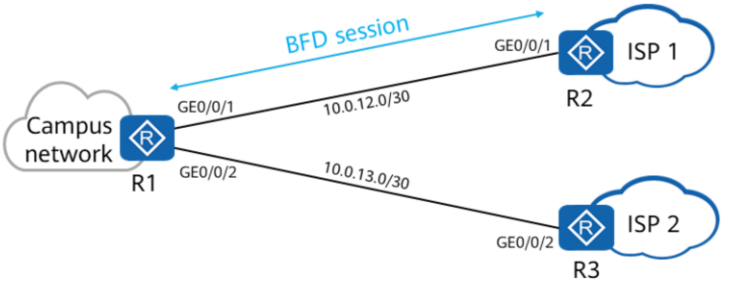


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The monitoring module monitors the link status and network performance, and notifies the track module of the detection result.
- After receiving the detection result from the monitoring module, the track module changes the status of the track item immediately and notifies the application module.
- The application module performs corresponding processing according to the status of the track item.

Association Between the Static Route and BFD

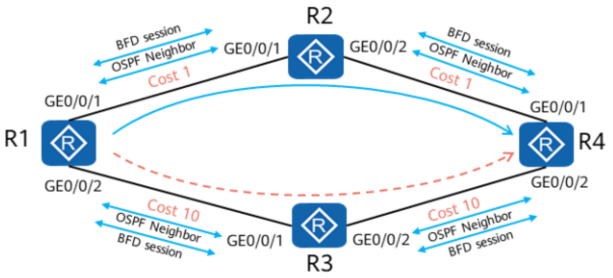
- A static route does not have a detection mechanism. When a static route is associated with BFD, the BFD session is used to detect the status of the link where the static route resides.
- Association between static routes and BFD is widely used. As shown in the figure, R1 connects to ISP 1 and ISP 2 through two links. In normal cases, the default route points to the link connected to ISP 1. When the link connected to ISP 1 fails, the BFD session can rapidly detect the fault and instructs R1 to switch traffic to the link connected to ISP 2.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

BFD for OSPF (1)

- If OSPF is not bound to BFD, the link fault detection time is determined by the Hello mechanism and is usually in seconds. BFD can be bound to implement millisecond-level fault detection.
- BFD for OSPF associates BFD with OSPF. BFD notifies OSPF of the fast detection result of a link fault.

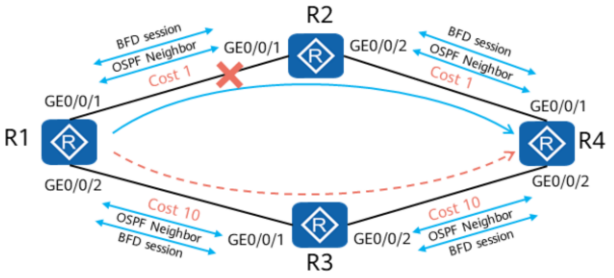


1. OSPF uses the Hello mechanism to discover neighbors and establishes a neighbor relationship.
2. OSPF notifies BFD of neighbor information including source and destination addresses.
3. BFD establishes a session based on the received neighbor information. After the session is established, BFD starts to detect link faults.
4. In normal situations, R1 selects the path R2 -> R4 based on the OSPF path cost.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BFD for OSPF (2)

After the BFD session is set up, devices periodically send BFD packets. If a device does not receive a response within a specified detection period, the device considers the forwarding path faulty. BFD will then notify the upper-layer application for processing.



1. When the link between R1 and R2 fails, BFD quickly detects the fault. The BFD session goes Down and BFD notifies R1.
2. R1 processes the neighbor Down event and notifies the local OSPF process that the neighbor is unreachable. Then R1 recalculates the route and sends the route to R4 through R3.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Introduction to BFD
2. BFD Implementation
3. BFD Application Scenarios
4. **Basic BFD Configurations**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

BFD Configuration Commands (1)

1. Create a BFD session and enter the BFD session view.

   [Huawei] **bfd** *session-name* **bind peer-ip** *ip-address* [**vpn-instance** *vpn-name*] **interface** *interface-type interface-number* [**source-ip** *ip-address*]

   By default, no BFD session is created. When creating a single-hop BFD session for the first time, you must bind the single-hop BFD session to the peer IP address and local interface. In addition, the configuration of the single-hop BFD session cannot be changed after being created. To change a configured BFD session, delete it and recreate a BFD session.

2. Create a BFD session using the multicast address as the peer address, and enter the BFD session view.

   [Huawei] **bfd** *session-name* **bind peer-ip default-ip interface** *interface-type interface-number* **[source-ip** *ip-address* **]**

3. Create a BFD6 session and enter the BFD6 session view.

   [Huawei] **bfd** *session-name* **bind peer-ipv6** *ip-address* [**vpn-instance** *vpn-name*] **interface** *interface-type interface-number* [**source-ipv6** *ip-address*]

   When creating a single-hop BFD6 session for the first time, bind the single-hop BFD6 session to the peer IPv6 address and local interface. In addition, the configuration of the single-hop BFD6 session cannot be changed after being created.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

BFD Configuration Commands (2)

4. Create a static BFD session with automatically negotiated discriminators.

   [Huawei] **bfd** *session-name* **bind peer-ip** *ip-address* [**vpn-instance** *vpn-name*] **interface** *interface-type interface-number* [**source-ip** *ip-address*] **auto**

5. Create a one-arm BFD echo session.

   [Huawei] **bfd** *session-name* **bind peer-ip** *ip-address* [**vpn-instance** *vpn-name*] **interface** *interface-type interface-number* [**source-ip** *ip-address*] **one-arm-echo**

6. Configure a local discriminator for the BFD session.

   [Huawei-bfd-session-test] **discriminator local** *discr-value*

   Assume that the BFD session name is **test**.

7. Configure a remote discriminator for the BFD session.

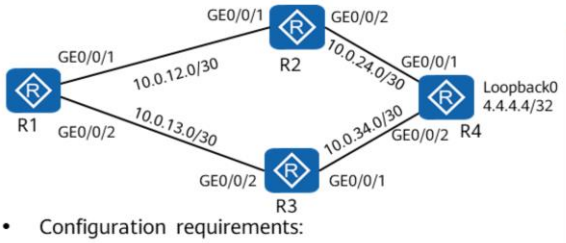
   [Huawei-bfd-session-test] **discriminator remote** *discr-value*

   The local discriminator of the local system must be the same as the remote discriminator of the remote system; otherwise, BFD sessions cannot be established. In addition, the local and remote discriminators cannot be modified after being configured.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Use the **commit** command to commit the configuration so that the configuration takes effect.

Configuring Association Between the Static Route and BFD



- Configuration requirements:
  - As shown in the figure, a floating static route to the network segment 4.4.4.4/32 of Loopback0 on R4 is configured on R1. In normal cases, R1 accesses R4 through R2. When R2 fails, R1 automatically selects the route to access Loopback0 on R4 through R3.
  - A BFD session needs to be established between R1 and R2 and is bound to the static route to implement fast fault detection and fast convergence.

Establish a static BFD session between R1 and R2.

```
[R1]bfd
[R1]bfd 12 bind peer 10.0.12.2 interface GigabitEthernet 0/0/1
[R1-bfd-session-12]discriminator local 10
[R1-bfd-session-12]discriminator remote 20
[R1-bfd-session-12]commit
[R2]bfd
[R2]bfd 21 bind peer 10.0.12.1 interface GigabitEthernet 0/0/1
[R2-bfd-session-21]discriminator local 20
[R2-bfd-session-21]discriminator remote 10
[R2-bfd-session-21]commit
```

On R1, configure a static route and bind the static route to a BFD session.

```
[R1] ip route-static 4.4.4.4 32 10.0.12.2 track bfd-session 12
[R1] ip route-static 4.4.4.4 32 10.0.13.2 preference 100
```

Other configurations are not provided here.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Verifying the BFD Session Configuration

```
[R1]display bfd session all  verbose
Session MIndex : 256       (One Hop) State : Up        Name : 12
Local Discriminator   : 1              Remote Discriminator  : 2
Session Detect Mode  : Asynchronous Mode Without Echo Function
  BFD Bind Type         : Interface(Vlanif10)
  Bind Session Type     : Static
  Bind Peer IP Address  : 10.0.12.2
  NextHop Ip Address    : 10.0.12.2
  Bind Interface        : GigabitEthernet0/0/1
  FSM Board Id          : 0              TOS-EXP             : 7
  Min Tx Interval (ms)  : 1000           Min Rx Interval (ms)  : 1000
  Actual Tx Interval (ms): 1000          Actual Rx Interval (ms): 1000
  Local Detect Multi    : 3              Detect Interval (ms)  : 3000
 Echo Passive          : Disable        Acl Number          : -
 Destination Port      : 3784           TTL                 : 255
----more----
```

BFD session in Up state

Static BFD session

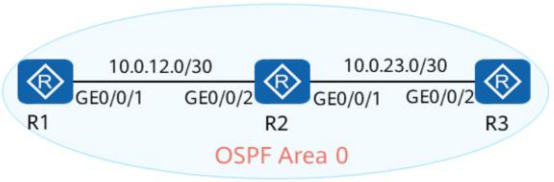
Minimum intervals for receiving and sending BFD control packets

Actual minimum intervals for receiving and sending BFD control packets after system negotiation

System detection count and fault detection interval

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuring BFD for OSPF



- Configuration requirements:
  - R1, R2, and R3 run OSPF and belong to area 0.
  - Configure BFD for OSPF and set BFD session parameters for all OSPF interfaces to further speed up OSPF convergence when the link status changes.
  - Set the maximum intervals for receiving and sending BFD control packets to 100 ms, and retain the default detection count.

R1 configuration:

```
[R1]bfd
[R1]interface GigabitEthernet 0/0/1
[R1-GigabitEthernet0/0/1]ip address 10.0.12.1 30
[R1]ospf 1
[R1-ospf-1]area 0
[R1-ospf-1-area-0.0.0.0]network 10.0.12.0 0.0.0.3
[R1-ospf-1-area-0.0.0.0]quit
[R1-ospf-1]bfd all-interfaces enable
[R1-ospf-1]bfd all-interfaces min-tx-interval 100 min-rx-interval
100 detect-multiplier 3
```

The configurations of R2 and R3 are similar to the configuration of R1, and are not provided.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Verifying the BFD Configuration

```
[R1]display bfd session all verbose
Session Mince : 256      (One Hop) State : Up       Name :
dyn_8192
Local Discriminator  : 8192          Remote Discriminator  : 8192
Session Detect Mode : Asynchronous Mode Without Echo Function
 BFD Bind Type       : Interface(GigabitEthernet0/0/0)
 Bind Session Type   : Dynamic
 Bind Peer IP Address  : 10.0.12.2
 NextHop Ip Address    : 10.0.12.2
 Bind Interface      : GigabitEthernet0/0/0
 FSM Board Id        : 0              TOS-EXP             : 7
 Min Tx Interval (ms)  : 100          Min Rx Interval (ms)  : 100
 Actual Tx Interval (ms): 100         Actual Rx Interval (ms): 100
 Local Detect Multi    : 3            Detect Interval (ms)  : 300
 Echo Passive        : Disable        Acl Number          : -
```

```
[R1]display ospf 1 bfd session all

        OSPF Process 1 with Router ID 10.0.12.1
 Area 0.0.0.0 interface 10.0.12.1(GigabitEthernet0/0/0)'s BFD
Sessions

NeighborId:10.0.12.2
AreaId:0.0.0.0
Interface:GigabitEthernet0/0/0
 BFDState:up          rx   :100          tx     :100
 Multiplier:3
BFD Local Dis:8192
LocalIpAdd: 10.0.12.1
RemoteIpAdd:10.0.12.2
Diagnostic Info:No  diagnostic information
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

1. (Multiple) Which of the following states are involved in the establishment of a BFD session?
   A. Down
   B. Init
   C. Up
   D. Establish
2. (Multiple) Which of the following BFD detection modes are available?
   A. Asynchronous mode
   B. Synchronous mode
   C. Demand mode
   D. Echo mode

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Answer 1: ABC

Answer 2: ACD

Summary

- This document describes the working principles of BFD, including the session establishment process, detection mode, detection time, and detection process.
- A BFD session experiences three states: Down, Init, and Up. The state of the BFD session is changed according to the State field in the received BFD packet. The BFD detection modes are classified into asynchronous and demand modes. The two modes differ in the detection packet sending mode, detection position, detection time, and detection process.
- BFD can quickly detect link faults. However, only application modules, such as OSPF and static routes, can invoke BFD to implement fast traffic path switching. Therefore, BFD is widely used on the live network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



VRRP Implementation and Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- User terminals on a LAN usually use a default gateway to access an external network. If the default gateway fails, all user terminals' traffic to the external network will be interrupted. Multiple gateways can be deployed to prevent single points of failure. To this end, the conflict between multiple gateways must be resolved.
- The Virtual Router Redundancy Protocol (VRRP) can implement gateway backup and solve the conflict between multiple gateways, thus improving network reliability.
- This course describes the working mechanism, basic configuration, and typical application scenarios of VRRP on a network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Analyze problems faced by a single gateway on a LAN.
  - Explain the basic concepts and working mechanism of VRRP.
  - Describe the active/standby VRRP switchover process.
  - Perform basic VRRP configurations.

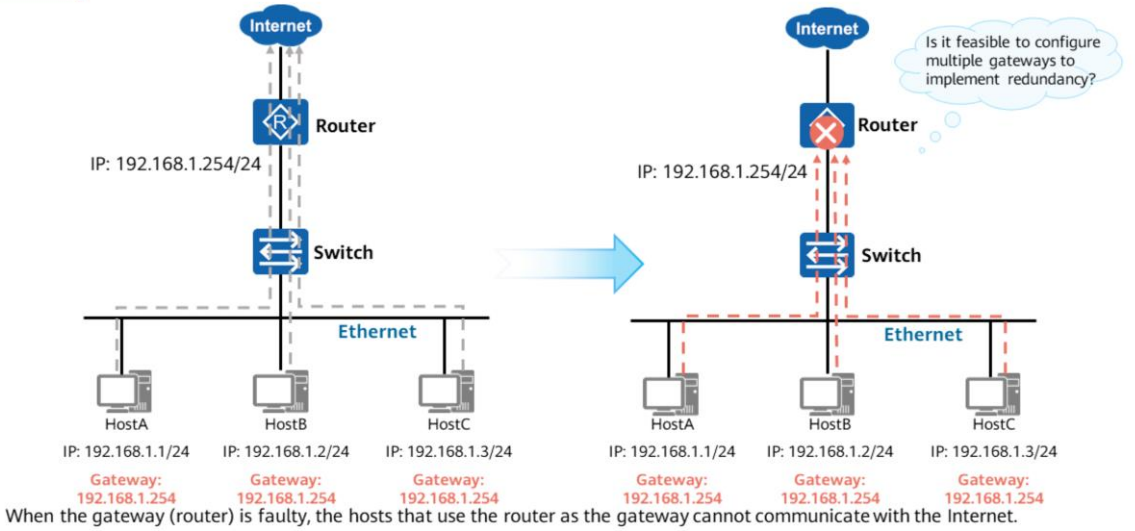
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Introduction to VRRP**
2. VRRP Implementation
3. Typical Application Scenarios of VRRP
4. Basic VRRP Configurations

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Problems Faced by a Single Gateway

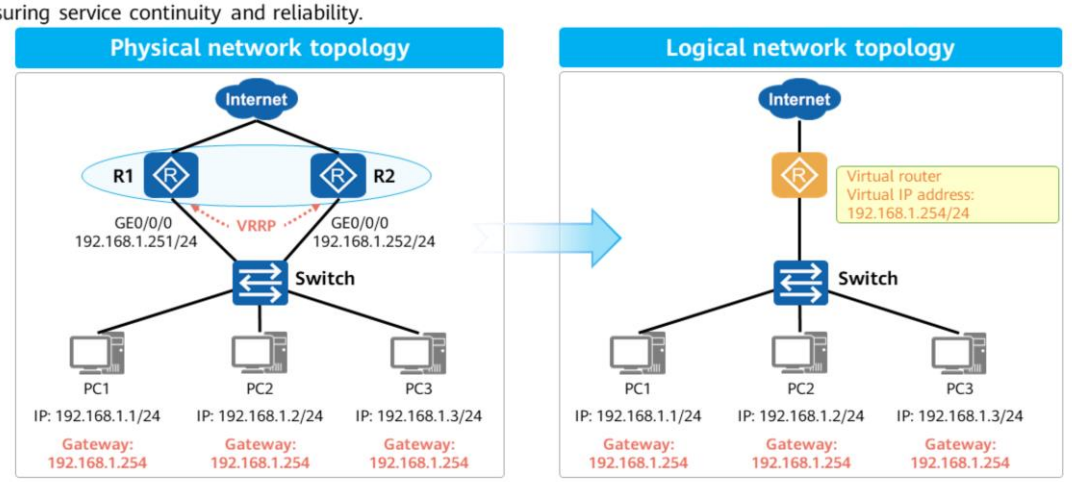


When the gateway (router) is faulty, the hosts that use the router as the gateway cannot communicate with the Internet.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Overview of VRRP

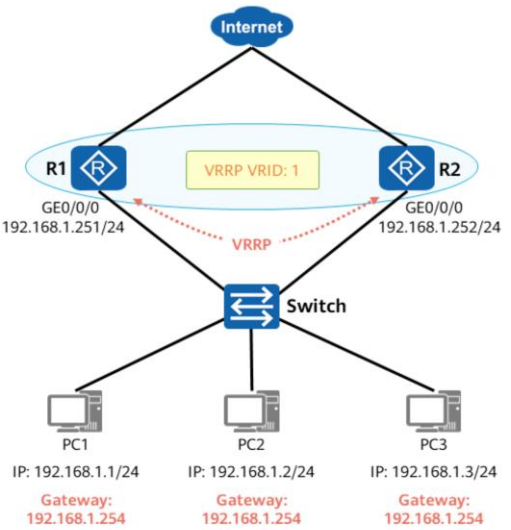
VRRP groups several routers into a virtual router. If one of routers fails, traffic can be switched to another router, ensuring service continuity and reliability.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- VRRP sets up a virtual router on a LAN.
- In this example:
  - There are two routers on the LAN: R1 and R2. The IP addresses of R1 and R2 are 192.168.1.251/24 and 192.168.1.252/24, respectively.
  - Configure R1 and R2 to constitute a virtual router. The virtual router uses IP address 192.168.1.254.
  - All PCs use IP address 192.168.1.254 as the default gateway address.

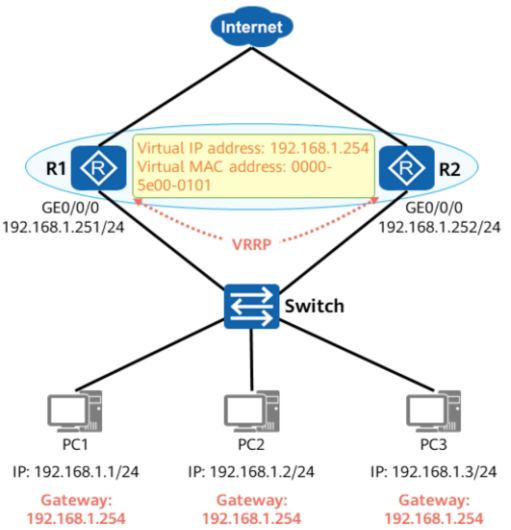
Basic VRRP Concepts (1)



- **VRRP router**: A router running VRRP, such as R1 and R2. VRRP is configured on interfaces of routers and works based on interfaces.
- **Virtual router ID (VRID)**: A VRRP group consists of multiple routers (interfaces) that work together and are identified by the same VRID. Routers in the same VRRP group exchange VRRP packets and form a virtual router. Only one master router can exist in a VRRP group.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

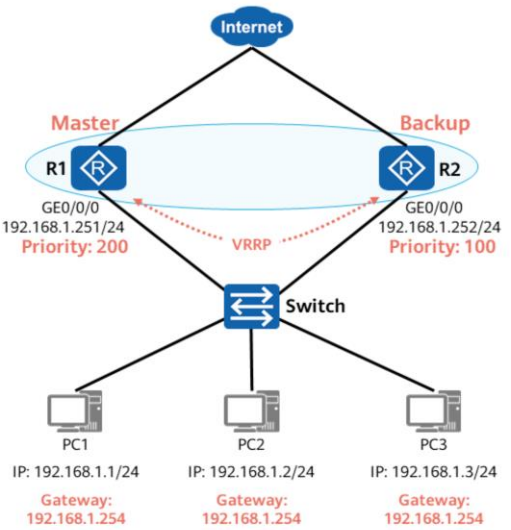
Basic VRRP Concepts (2)



- **Virtual router**: Each VRRP group forms a virtual router that is a logical device. A VRRP group forms only one virtual router.
- **Virtual IP address and virtual MAC address**: A virtual router has its own IP address and MAC address. The IP address is specified by the network administrator during VRRP configuration. A virtual router can have one or more IP addresses. Generally, this IP address is used as the gateway address. The format of the virtual MAC address is 0000-5e00-01xx, where xx is the VRID.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic VRRP Concepts (3)

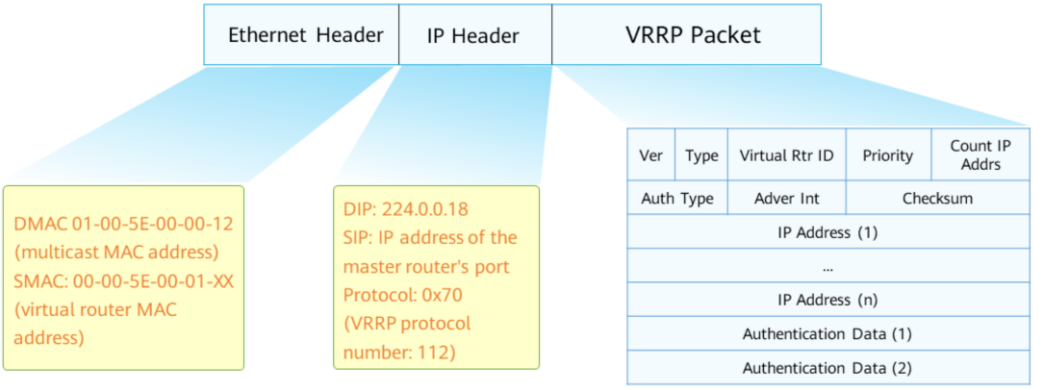


- **Master router**: The master router forwards packets in a VRRP group. In each VRRP group, only the master router responds to the ARP Request packet destined for the virtual IP address. The master router periodically sends VRRP packets to notify the backup router in the same VRRP group of its own status.
- **Backup router**: The backup router listens to the VRRP packets sent by the master router in real time and is ready to take over services of the master router.
- **Priority**: The priority value is used to elect the master and backup routers. The priority value ranges from 0 to 255. A larger value indicates a higher priority. If the values are the same, the interface IP addresses are compared. The interface with the largest IP address is preferred.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VRRP Packet Format

VRRP has only one packet type — Advertisement packets. Advertisement packets are multicast packets and can be transmitted only in the same broadcast domain. The destination multicast address of Advertisement packets is 224.0.0.18.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Fields in a VRRP Advertisement packet:
  - Ver: VRRP has two versions. VRRPv2 applies only to IPv4 networks, and VRRPv3 applies to both IPv4 and IPv6 networks.
  - Virtual Rtr ID: Virtual router ID associated with the packet.
  - Priority: Priority of the VRRP router that sends the VRRP packet.
  - Count IP Addrs: Number of virtual IP addresses contained in the VRRP packet.
  - Auth Type: VRRP supports non-authentication, plain-text password authentication, and MD5 authentication, corresponding to values 0, 1, and 2, respectively.
  - Adver Int: Interval for sending VRRP Advertisement packets. The default value is 1s.
  - IP Address: Virtual IP address of the associated virtual router. Multiple IP addresses can be configured.
  - Authentication Data: Password required for authentication.

VRRP Timers

- VRRP defines two timers:
  - ADVER_INTERVAL timer: specifies the interval at which the master router sends VRRP Advertisement packets. The default value is 1s.
  - MASTER_DOWN timer: indicates that the backup router preempts the Master state after the MASTER_DOWN timer expires. The MASTER_DOWN timer is calculated as follows:
    - MASTER_DOWN = (3*ADVER_INTERVAL) + Skew_time (offset time)
    - Skew_Time = (256 – Priority)/256

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

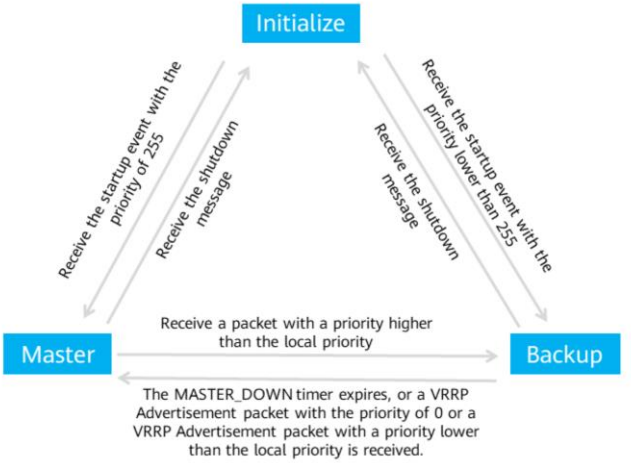
1. Introduction to VRRP
2. **VRRP Implementation**
3. Typical Application Scenarios of VRRP
4. Basic VRRP Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

VRRP State Machine

The VRRP state machine has three states: Initialize, Master, and Backup.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- A startup event can be automatically triggered by the system after VRRP is configured or triggered by the change of the lower-layer link from unavailable to available on the interface configured with VRRP.

VRRP States

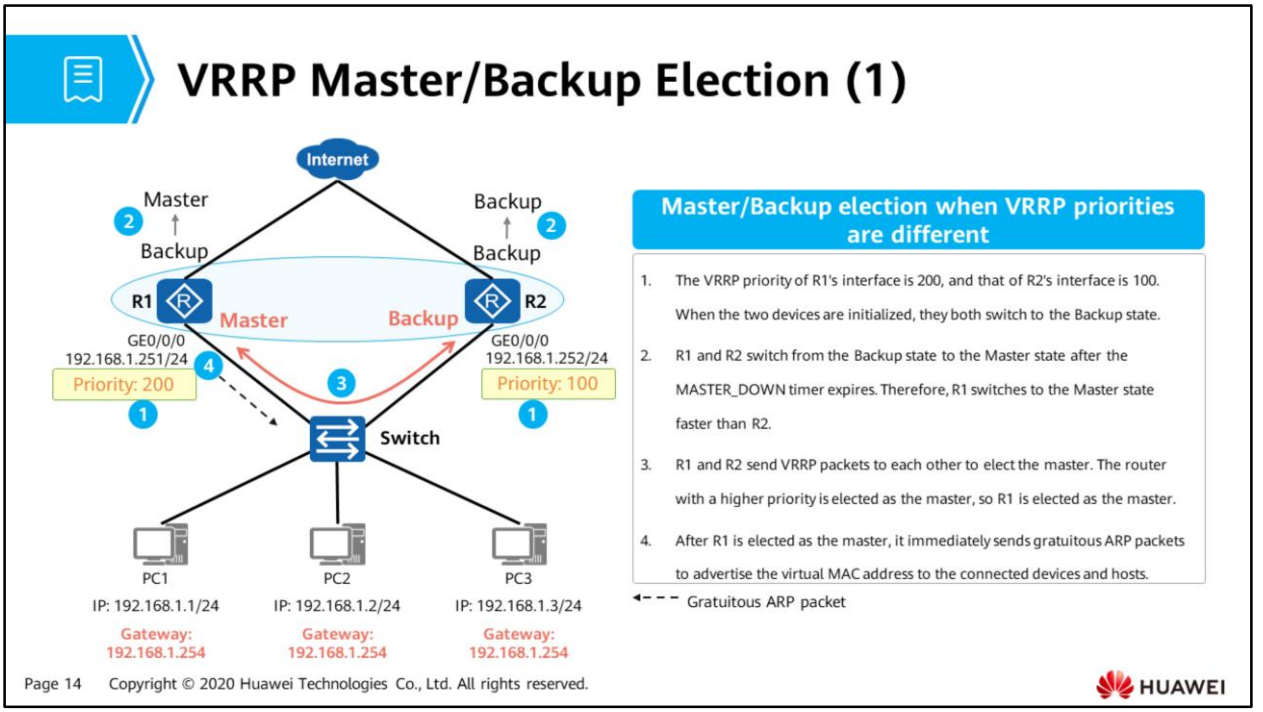
Device in Master State

1. Sends a VRRP Advertisement packet each time the ADVER_INTERVAL timer expires.
2. Uses the virtual MAC address to respond to ARP Request packets destined for the virtual IP address.
3. Forwards IP packets sent to the virtual MAC address.
4. Allows a virtual IP address to be pinged by default.
5. When multiple devices are in Master state, they compare their IP addresses in received packets with the same priority. If the IP address in the packet is greater than the local IP address, the device switches to the Backup state. If the IP address in the packet is less than or equal to the local IP address, the device remains in the Master state.

Device in Backup State

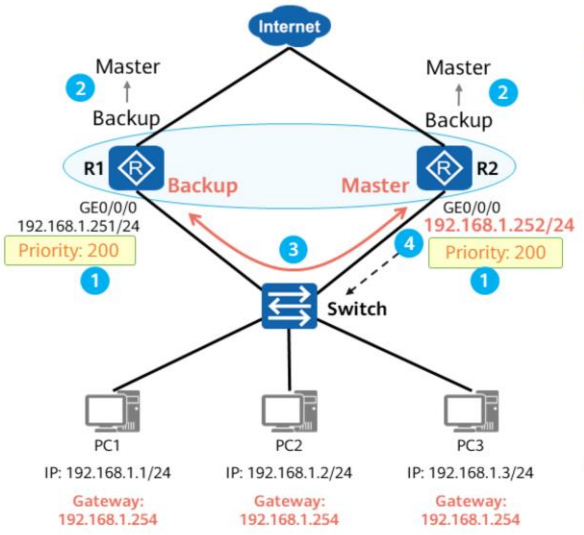
1. Receives VRRP Advertisement packets from the master router and checks whether the master router is working properly based on information in the packets.
2. Does not respond to an ARP Request packet carrying a virtual IP address.
3. Discards IP packets sent to the virtual MAC address.
4. Discards IP packets sent to the virtual IP address.
5. Resets the MASTER_DOWN timer but does not compare IP addresses if it receives a VRRP Advertisement packet carrying a VRRP priority higher than or equal to the local VRRP priority.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- If a VRRP-enabled device in Initialize state receives an interface Up message and its priority is lower than 255, it switches to the Backup state. The device switches to the Master state when the MASTER_DOWN timer expires.
- If the device with a higher priority and the device with a lower priority start in sequence, the device with a higher priority enters the Master state first. After receiving a VRRP Advertisement packet with a higher priority, the device with a lower priority remains in Backup state.
- If the device with a lower priority and the device with a higher priority start in sequence, the device with a lower priority switches from the Backup state to the Master state first. After receiving the VRRP Advertisement packet with a lower priority, the device with a higher priority switches to the Master state.

VRRP Master/Backup Election (2)



Master/Backup election when VRRP priorities are the same

1. The VRRP priorities of GE0/0/0 interfaces on R1 and R2 are both 200. When the two devices are initialized, they both switch to the Backup state.
2. Because the priorities of R1 and R2 are the same, R1 and R2 switch from the Backup state to the Master state after the MASTER_DOWN timer expires.
3. R1 and R2 exchange VRRP packets with the same priority. R1 and R2 compare their interface IP addresses to select the master. Because the interface IP address of R2 is greater than that of R1, R2 is selected as the master.
4. After R2 is elected as the master, it immediately sends gratuitous ARP packets to advertise the virtual MAC address to the connected devices and hosts.

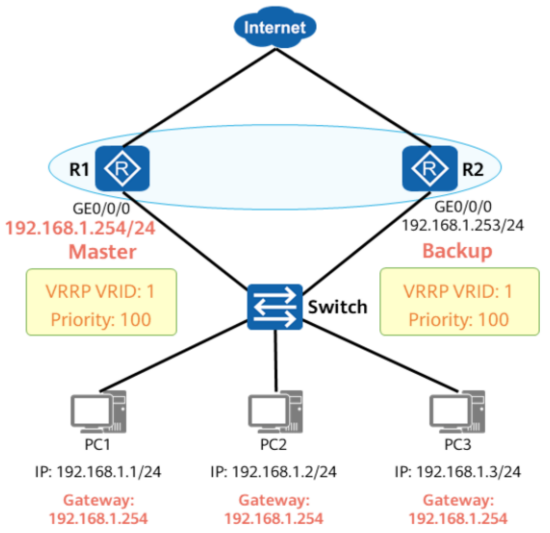
◄- - - Gratuitous ARP packet

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- If a VRRP-enabled device in Initialize state receives an interface Up message and its priority is lower than 255, it switches to the Backup state. The device switches to the Master state when the MASTER_DOWN timer expires.
- If the device with a higher priority and the device with a lower priority start in sequence, the device with a higher priority enters the Master state first. After receiving a VRRP Advertisement packet with a higher priority, the device with a lower priority remains in Backup state.
- If the device with a lower priority and the device with a higher priority start in sequence, the device with a lower priority switches from the Backup state to the Master state first. After receiving the VRRP Advertisement packet with a lower priority, the device with a higher priority switches to the Master state.

VRRP Master/Backup Election (3)



When a router interface is configured as the VRRP IP address owner (the interface IP address is the same as the virtual IP address), the router can directly switch to the Master state without waiting for any timer to expire.

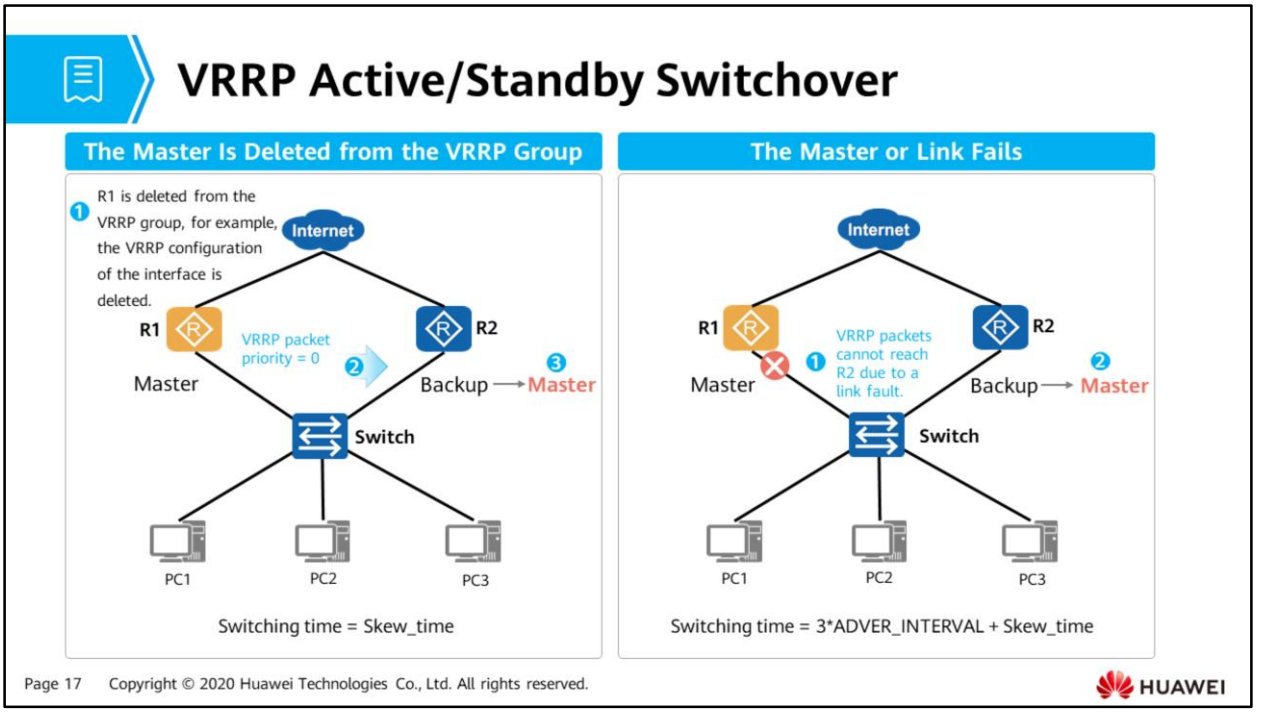
Master/Backup election process when a device interface is configured as the IP address owner

1. The VRRP priorities of GE0/0/0 interfaces on R1 and R2 use the default value — 100. However, the IP address of GE0/0/0 on R1 is the same as the virtual IP address.
2. R1's GE0/0/0 directly switches to the Master state, and R1 becomes the master.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

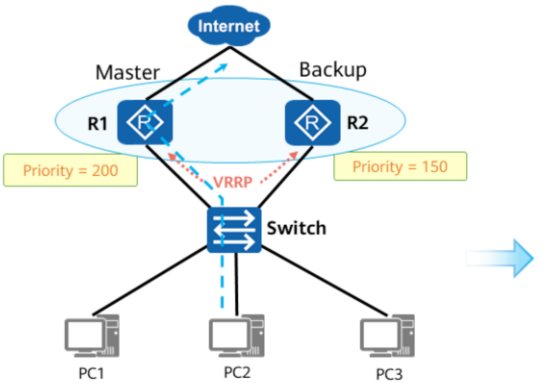
HUAWEI

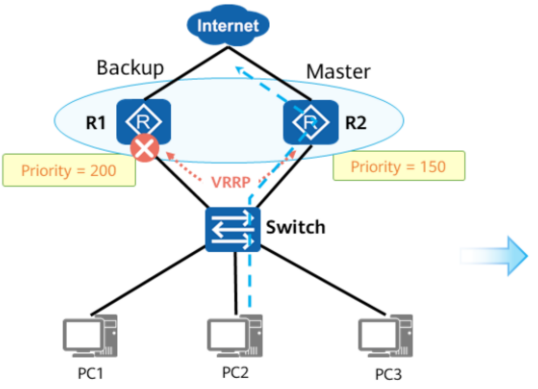
- In most cases, the interface IP address of a VRRP router does not overlap with the IP address of a virtual router. That is, an independent IP address is planned for the virtual router instead of the interface IP address of a router. There is also an exception. For example, if IP addresses are insufficient on some networks, the interface IP address of a router may be used as the IP address of the virtual router. In this case, the router becomes the master.
- The priority of a VRRP-enabled interface cannot be manually set to 255. When the IP address of an interface is configured as the IP address owner, the priority of the interface automatically changes to 255.



- If the master gives up the master role (for example, the master is deleted from the VRRP group), it sends VRRP Advertisement packets carrying a priority of 0 to the backups. Without waiting for the MASTER_DOWN timer to expire, the backup router with the highest priority switches to the Master state after a specified switching time. This switching time is called Skew_Time.
- If the master cannot send VRRP Advertisement packets due to network faults, the backups cannot learn the running status of the master immediately. In this situation, the backup router with the highest priority switches to the Master state after the MASTER_DOWN timer expires.

VRRP Active/Standby Switchback (1)

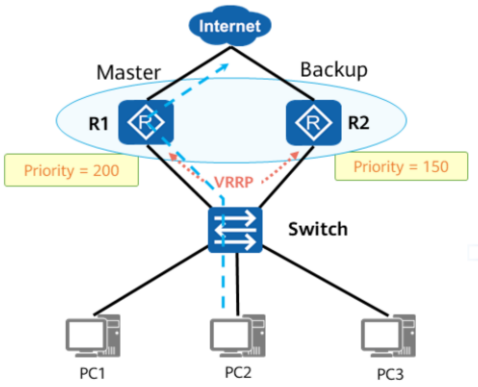




1. In normal situations, the master forwards user packets. As shown in the figure, all user traffic reaches the Internet through R1.
2. When R1 fails, a new VRRP master/backup election is performed. As shown in the figure, R2 becomes the new master to forward user packets.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VRRP Active/Standby Switchback (2)



3. After R1 recovers, a new VRRP master/backup election is performed. R1 becomes the new master to forward user packets because R1 has a higher priority than R2.

VRRP preemption mode:

- **Preemption mode (enabled by default)**: If the backup is enabled with the preemption function, it immediately switches to the Master state when detecting that the priority of the master is lower than its own priority.
- **Non-preemption mode**: If the preemption function is disabled on the backup, the backup remains in Backup state until the master fails even when detecting that the priority of the master is lower than that of the backup.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- When the preemption mode is enabled for a VRRP group and an active/standby switchover is performed, the switching time is as follows:

  Switching time = 3*ADVER_INTERVAL + Skew_time + Delay_time

- In preemption mode, if the master is unstable or the network quality is poor, the VRRP group frequently switches, causing frequent update of ARP entries. To resolve this problem, you can set a preemption delay. After the preemption delay plus the value of the MASTER_INTERVAL timer, if the master becomes stable, a switchback is performed.

Contents

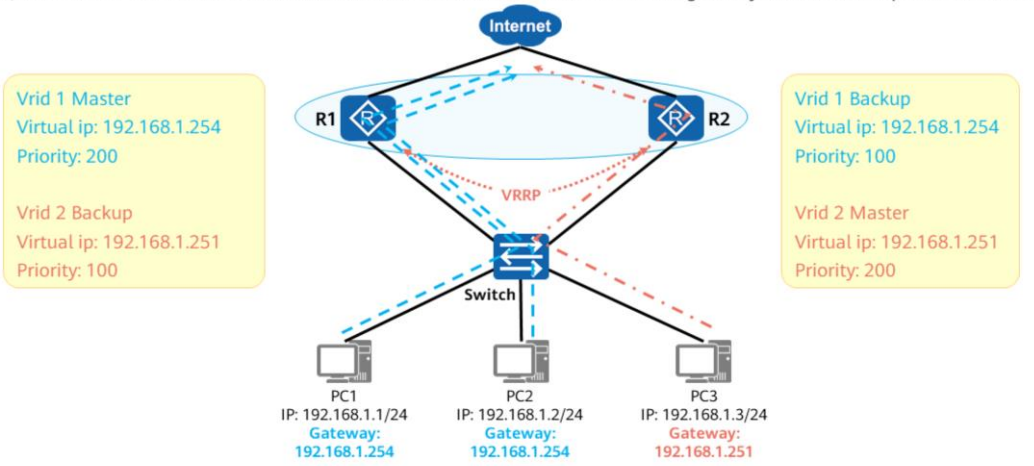
1. Introduction to VRRP
2. VRRP Implementation
3. **Typical Application Scenarios of VRRP**
4. Basic VRRP Configurations

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

VRRP Load Balancing

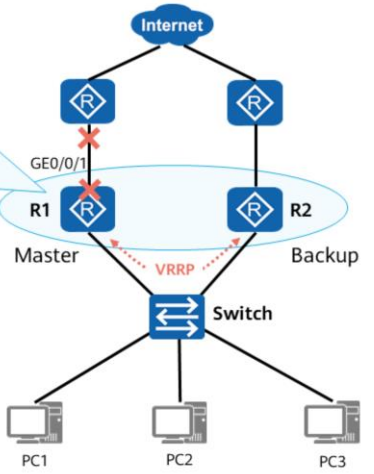
In the scenario where multiple virtual routers (VRRP groups) are created and each physical router plays different roles in different VRRP groups, the virtual IP addresses of different virtual routers function as different intranet gateway addresses to implement load balancing.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VRRP Monitoring the Uplink Interface Status

VRRP can monitor the status of the uplink interface. When a device detects a fault on the uplink interface or link, the device reduces the VRRP priority. This ensures that the backup with a normal uplink can be elected as the Master to forward packets.

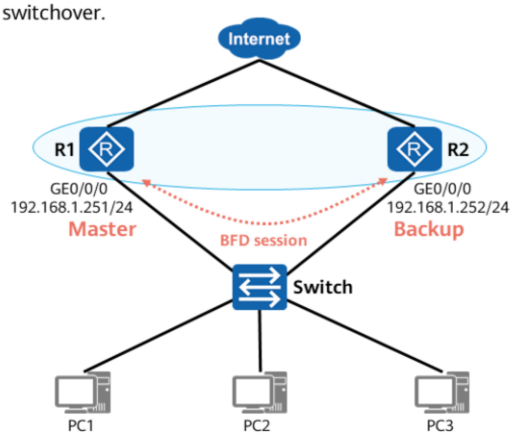


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- If association between VRRP and the uplink interface is not configured and the uplink interface or link of R1 (master) in the VRRP group fails, the VRRP group cannot detect the fault and the master cannot forward traffic. In this case, the active/standby switchover cannot be performed, causing a traffic blackhole.

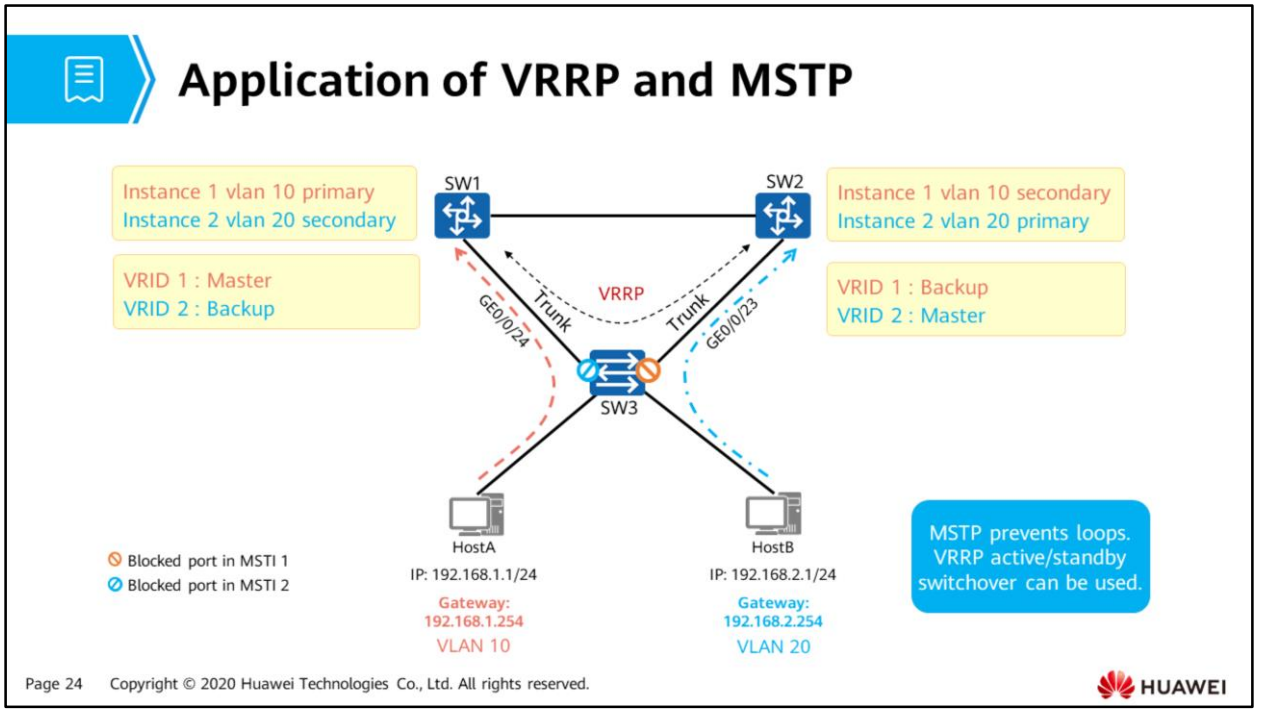
Association Between VRRP and BFD

With association between VRRP and BFD enabled, when the backup detects a fault through BFD, the backup switches to the Master state immediately without waiting for the MASTER_DOWN timer to expire. This implements millisecond-level active/standby switchover.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- If the link between devices in a VRRP group fails, VRRP Advertisement packets cannot be exchanged to negotiate the Master or Backup state. A backup switches to the Master state when the MASTER_DOWN timer expires. During the waiting period, user traffic is still forwarded to the master, resulting in user traffic loss.
- A BFD session is established between the master and backup in a VRRP group and is bound to the VRRP group. BFD immediately detects communication faults in the VRRP group and instructs the VRRP group to perform an active/standby switchover, minimizing service interruptions.
- For association between VRRP and BFD, a VRRP group adjusts priorities according to the BFD session status and determines whether to perform an active/standby switchover according to the adjusted priorities. In practice, delayed preemption is configured on the master and immediate preemption is configured on the backup. When the backup detects that the BFD session goes Down, it increases its priority to be higher than the priority of the master to implement a fast switchover. After the fault is rectified and the BFD session goes Up, the new master reduces its priority and sends a VRRP Advertisement packet. After the delay, the new master becomes the backup again.



- MSTP maps one or more VLANs to an MSTI. Multiple VLANs share a spanning tree, and MSTP implements load balancing.
- The VRRP-enabled gateway can be automatically switched based on network topology changes, improving network reliability.
- VRRP+MSTP can implement load balancing while ensuring network redundancy.

Contents

1. Introduction to VRRP
2. VRRP Implementation
3. Typical VRRP Applications
4. **Basic VRRP Configurations**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Basic VRRP Configuration Commands (1)

1. Create a VRRP group and configure a virtual IP address for the VRRP group.

   [interface-GigabitEthernet0/0/0] **vrrp vrid** *virtual-router-id* **virtual-ip** *virtual-address*

   The virtual IP address of a VRRP group must be unique. Interfaces in the same VRRP group must use the same VRID.

2. Configure a priority for each device in the VRRP group.

   [interface-GigabitEthernet0/0/0] **vrrp vrid** *virtual-router-id* **priority** *priority-value*

   In most cases, the priority of the master is higher than that of the backup.

3. Configure a preemption delay for the VRRP group.

   [interface-GigabitEthernet0/0/0] **vrrp vrid** *virtual-router-id* **preempt-mode timer delay** *delay-value*

4. Configure the VRRP group to work in non-preemption mode.

   [interface-GigabitEthernet0/0/0] **vrrp vrid** *virtual-router-id* **preempt-mode disable**

   By default, the preemption mode is enabled.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Basic VRRP Configuration Commands (2)

5. Associate the VRRP group with an interface.

   [interface-GigabitEthernet0/0/0] **vrrp vrid** *virtual-router-id* **track interface** *interface-type interface-number* **[increased** *value-increased* | **reduced** *value-decreased*]

   You can configure the device to increase or decrease its priority when detecting an uplink interface or link fault. The IP address owner and Eth-Trunk member interfaces cannot be associated with VRRP.

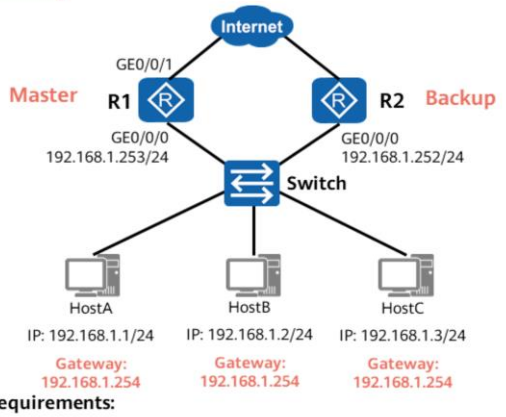
6. Associate the VRRP group with a BFD session.

   [interface-GigabitEthernet0/0/0] **vrrp vrid** *virtual-router-id* **track bfd-session** { *bfd-session-id* | **session-name** *bfd-configure-name* } **[increased** *value-increased* | **reduced** *value-reduced*]

   If **session-name** *bfd-configure-name* is specified, the VRRP group can be associated only with static BFD sessions with automatically negotiated discriminators or static BFD session discriminators.
   If *bfd-session-id* is specified, a VRRP group can be bound to only static BFD sessions.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VRRP Configuration Example



Requirements:

- R1 and R2 form a VRRP group. R1 is the master and R2 is the backup.
- The preemption mode is used when the master recovers. The preemption delay is 10 seconds.
- The master monitors the status of the uplink interface to implement automatic VRRP active/standby switchover.

R1 configuration:

```
[R1] interface GigabitEthernet0/0/0
[R1-GigabitEthernet0/0/0] ip address 192.168.1.253  24
[R1-GigabitEthernet0/0/0] vrrp vrid 1 virtual-ip 192.168.1.254
[R1-GigabitEthernet0/0/0] vrrp vrid 1 priority 120
[R1-GigabitEthernet0/0/0] vrrp vrid 1 preempt-mode timer delay 10
[R1-GigabitEthernet0/0/0] vrrp vrid 1 track interface
GigabitEthernet0/0/1 reduced 30
```

R2 configuration:

```
[R2] interface GigabitEthernet0/0/0
[R2-GigabitEthernet0/0/0] ip address 192.168.1.252 24
[R2-GigabitEthernet0/0/0] vrrp vrid 1 virtual-ip 192.168.1.254
[R2-GigabitEthernet0/0/0] vrrp vrid 1 priority 110
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Verifying Basic VRRP Configurations

```
[R1]display vrrp
  GigabitEthernet0/0/0 | Virtual Router 1                # Set the VRID to 1.
    State : Master                The device is in Master state in the VRRP
    group.
    Virtual IP : 192.168.1.254
    Master IP : 192.168.1.253
    PriorityRun : 120         #Set the priority of the interface in the VRRP
    group to 120.
    PriorityConfig : 120
    MasterPriority : 120
    Preempt : YES   Delay Time : 10 s   # Enable the preemption mode
    and set the preemption delay to 10s.
    TimerRun : 1 s
    TimerConfig : 1 s
    Auth type : NONE
    Virtual MAC : 0000-5e00-0101
    Check TTL : YES
    Config type : normal-vrrp
    Track IF : GigabitEthernet0/0/1    Priority reduced : 30
    IF state : UP
```

```
[R2]display vrrp
  GigabitEthernet0/0/0 | Virtual Router 1
    State : Backup            # The device is in Backup state in the VRRP
    group.
    Virtual IP : 192.168.1.254
    Master IP : 192.168.1.253
    PriorityRun : 110         #The priority of the interface in the VRRP
    group is 110.
    PriorityConfig : 110
    MasterPriority : 120
    Preempt : YES   Delay Time : 0 s    # Enable the preemption mode
    and set the preemption delay to 0.
    TimerRun : 1 s
    TimerConfig : 1 s
    Auth type : NONE
    Virtual MAC : 0000-5e00-0101
    Check TTL : YES
    Config type : normal-vrrp
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Quiz

- (Multiple) Which of the following statements about the IP address in VRRP packets are true?
  - A. The source IP address is the IP address of an interface on the master.
  - B. The source IP address is the virtual IP address of the virtual router.
  - C. The destination IP address is a broadcast IP address.
  - D. The destination IP address is the IP address of the multicast group.
- (Multiple) Which statements about VRRP timers are true?
  - A. The default interval for sending VRRP Advertisement packets is 1s. The interval for sending VRRP Advertisement packets associated with a VRRP router must be the same.
  - B. If the preemption delay is set to 4s, the backup becomes the new master if it does not receive a VRRP Advertisement packet from the master within 4 seconds.
  - C. Set the preemption delay to 4s and the interval for sending VRRP Advertisement packets to 2s. If the backup does not receive a VRRP Advertisement packet from the master within 6s, the backup becomes the new master.
  - D. On a busy network, set the preemption delay to a large value to prevent VRRP flapping.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

1. AD
2. AD

Summary

- As an important reliability technology, VRRP is often used to implement gateway redundancy. VRRP improves network reliability and implements load balancing.
- The VRRP priority is used to control VRRP master/backup election, and active/standby switchover and switchback. To ensure network stability, the preemption delay mechanism is designed for VRRP to reduce network flapping.
- VRRP can monitor the status of uplink interfaces to detect network faults and associate with VRRP to ensure network reliability. VRRP can also be bound to BFD sessions to implement fast convergence. In addition, VRRP can be used together with MSTP, which is a common networking solution in campus networks.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



DHCP Implementation and Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- As the network scale increases continuously and networks become increasingly complex, locations of terminals such as hosts, mobile phones, and tablets on networks change frequently. When a terminal accesses a network, the IP address, gateway address, and DNS server address need to be configured. Manually configuring these parameters for terminals is inefficient and inflexible.
- The Internet Engineering Task Force (IETF) released the Dynamic Host Configuration Protocol (DHCP) in 1993. DHCP implements automatic configuration of network parameters, reducing the configuration and maintenance costs of clients.
- This course describes the working mechanism, application scenarios, and basic configurations of DHCP.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe DHCP implementation.
  - Describe DHCP address allocation rules.
  - Differentiate application scenarios of DHCP and DHCP relay.
  - Perform basic DHCP configurations.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

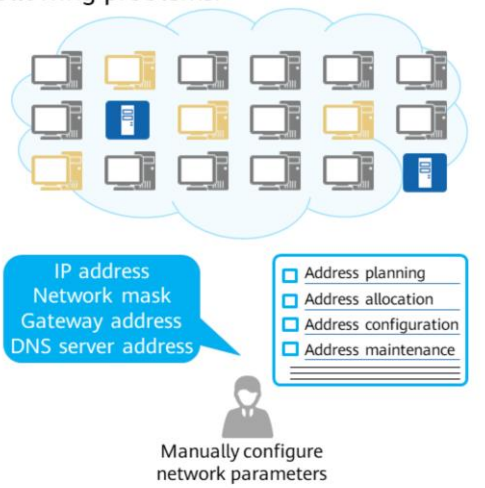
1. **DHCP Background**
2. DHCP Working Mechanism and Configuration
3. DHCP Relay Working Mechanism and Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Problems in Manually Configuring Network Parameters

Manually configuring network parameters has the following problems:

- Inflexibility
- Being vulnerable to errors
- Low IP address usage
- Heavy workload
- High requirements on personnel skills

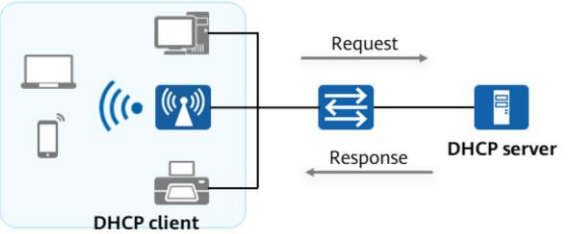


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- When network parameters such as the host IP address, network mask, gateway address, and DNS server address are manually configured, complex operation processes such as address planning, allocation, configuration, and maintenance are required. As a result, address allocation is inflexible, the IP address resource usage is low, the configuration is error-prone due to heavy workload, and there are high requirements on personnel skills.

Basic Concepts of DHCP

- The Dynamic Host Configuration Protocol (DHCP) dynamically assigns IP addresses to hosts and centrally manages host configurations.
- DHCP uses the client/server (C/S) communication mode. Protocol packets are exchanged in UDP mode. Port 67 (DHCP server) and port 68 (DHCP client) are used.
  - In normal cases, a client applies for configurations from the server.
  - The server returns the configuration such as the IP address allocated to a client.
- Compared with manual configuration, DHCP has the following advantages:
  - High efficiency
  - Strong flexibility
  - Easy management



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

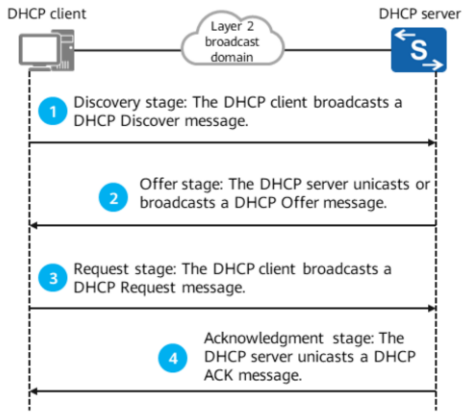
- Network terminals, such as hosts, printers, laptops, mobile phones, and APs, function as DHCP clients to request network parameters from the DHCP server. The DHCP server dynamically allocates network parameters based on the requests from the DHCP clients.

Contents

1. DHCP Background
2. **DHCP Working Mechanism and Configuration**
3. DHCP Relay Working Mechanism and Configuration

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCP Working Mechanism When a DHCP Client Connects to a Network for the First Time

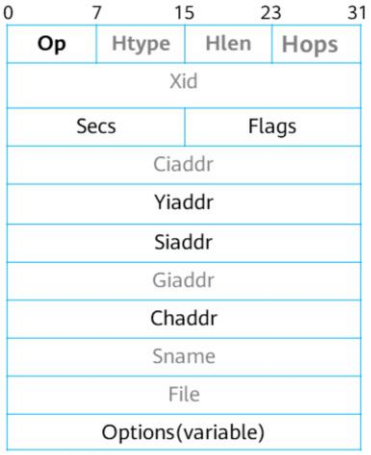


1. Discovery stage: The DHCP client detects DHCP servers.
   - The DHCP client broadcasts a DHCP Discover message to detect DHCP servers. The DHCP Discover message carries the client's MAC address, parameter request list, and broadcast flag.
2. Offer stage: A DHCP server offers network parameters to the DHCP client.
   - A DHCP server selects an address pool on the same network segment as the IP address of the interface receiving the DHCP Discover message, and selects an idle IP address from the address pool. The DHCP server then sends a DHCP Offer message carrying the allocated IP address to the DHCP client.
3. Request stage: The DHCP client selects an IP address.
   - If multiple DHCP servers reply with a DHCP Offer message to the DHCP client, the client accepts only the first received DHCP Offer message. The client then broadcasts a DHCP Request message carrying the selected DHCP server identifier and IP address.
4. Acknowledgment stage: The DHCP server acknowledges the IP address offered to the client.
   - After receiving the DHCP ACK message, the DHCP client broadcasts a gratuitous ARP packet to detect whether other terminals on the network segment use the IP address allocated by the DHCP server.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The DHCP Request message is broadcast so as to notify all the DHCP servers that the DHCP client has selected the IP address offered by a DHCP server. Then the other servers can allocate IP addresses to other clients.
- In the acknowledgement stage, IP address conflicts may occur in the following situations:
  - After receiving the DHCP Discover message, the DHCP server sends a ping packet to the client before assigning an IP address to the client. If the IP address can be pinged, the IP address is unavailable and another IP address is assigned to the client.
  - After the client successfully obtains an IP address, it immediately sends a gratuitous ARP packet. If a response packet is received, the client sends a DHCP Decline message to notify the DHCP server that the allocated IP address conflicts. The DHCP server then sets the IP address status to conflicting. Then, the client sends another DHCP Discover message to request a new IP address.

DHCP Message Format



DHCP message format

Key fields:

- Op (Operation Code): indicates the message type:
  - 1: DHCP Discover message
  - 2: DHCP Offer message
- Secs (seconds): is filled by a client, indicating the number of seconds elapsed since a client obtained or renewed an IP address. The default value is 3600s.
- Flags: indicates the format of the response message sent by a server, which is requested by a client. Only the leftmost bit in this field is used, and the other 15 bits are set to 0. The leftmost bit specifies the mode a DHCP server uses to transmit a DHCP Offer message.
- Yiaddr (your client ip address): indicates the IP address that a DHCP server allocates to a client. The DHCP server fills this field into a DHCP Reply message.
- Siaddr (server ip address): indicates the IP address of the DHCP server.
- Chaddr (client hardware address): indicates the IP address of a DHCP client.
- Options: contains the configuration allocated by the DHCP server to a client.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Htype (hardware type): indicates the type of the hardware address.
- Hlen (hardware length): indicates the length of the hardware address.
- Hops: indicates the number of DHCP relay agents that DHCP messages pass through. This field is set to 0 by a client. The value of this field is increased by 1 each time the DHCP message passes a DHCP relay agent. This field is used to limit the number of DHCP relay agents that DHCP messages pass through.
- Xid: indicates a random number selected by a DHCP client to exchange messages with a DHCP server.
- Sname (server host name): indicates the name of the server from which a client obtains the configuration. This field is optional and is filled in by a DHCP server. This field must be filled in with a character string that ends with 0.
- File (file name): indicates the name of the configuration file for starting DHCP on the client. The DHCP server fills this field and delivers it together with the IP address to the client. This field is optional and must be filled in with a character string that ends with 0.

Options Field

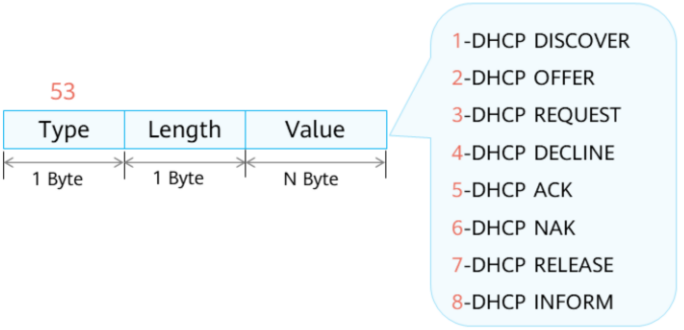
- The options field has variable length and a maximum of 312 bytes. It contains the DHCP message type and configuration parameters allocated by a DHCP server to a client. The configuration parameters include the gateway IP address, DNS server IP address, and IP address lease.
- The options field consists of Type, Length, and Value. The value of the Type sub-field ranges from 1 to 255. The following table lists the common values of Type, Length, and Value.

Type	Length (Byte)	Value	Function
1	4	Subnet Mask	Specifies a subnet mask.
3	4	Router (gateway)	Specifies a gateway address.
50	4	Requested IP Address	Specifies a requested IP address.
51	4	IP Address Lease Time	Specifies an IP address lease.
53	1	Message Type	Specifies a DHCP message type.
54	4	DHCP Server Identifier	Specifies a DHCP server identifier.
55	9	Parameter Request List	Sets the parameter request list. A DHCP client uses this option to request specified configuration parameters.
58	4	Rebinding Time Value	Specifies the lease renewal time (T1), which is 50% of the lease time.
59	4	Renewal Time Value	Specifies the extended period of the lease (T2). Generally, the value is 87.5% of the lease time.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCP Message Types

The Type subfield in the Options field of a DHCP message is set to 53, indicating the DHCP message type. In the figure, when Type is 53 and Length is 1, the Value sub-field ranges from 01 to 08, indicating a DHCP message type.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- DHCP Discover message: A DHCP client broadcasts this message to locate a DHCP server when the client attempts to connect to a network for the first time.
- DHCP Offer message: A DHCP server sends this message in response to a DHCP Discover message. A DHCP Offer message carries configuration information.
- DHCP Request message: A DHCP client broadcasts a DHCP Request message to respond to a DHCP Offer message sent by a DHCP server after the client starts; a DHCP client broadcasts a DHCP Request message to confirm the configuration (including the allocated IP address) after the client restarts; a DHCP client unicasts or broadcasts a DHCP Request message to renew the IP address lease after the client obtains an IP address.
- DHCP Decline message: A DHCP client sends this message to notify the DHCP server when detecting that the IP address assigned by the DHCP server conflicts with another IP address.
- DHCP ACK message: A DHCP server sends this message to acknowledge a DHCP Request message sent from a DHCP client.
- DHCP NAK message: A DHCP server sends this message to reject a DHCP Request message from a DHCP client.
- DHCP Release message: A DHCP client sends this message to release its allocated IP address.
- DHCP Inform message: A DHCP client sends this message to obtain network configuration parameters, such as the gateway address and DNS server address, after

it has obtained an IP address.

User-defined Options

In addition to the options specified in the standard protocol, some options, such as Option 82 and Option 43, are not specified uniformly and are referred to as user-defined options.

- Relay agent information option (Option 82)
  - The Option 82 field contains a maximum of 255 sub-options. If the Option 82 field is defined, at least one sub-option must be defined.
  - The DHCP relay agent or a device enabled with DHCP snooping appends the Option 82 field to the DHCP Request message sent from a DHCP client, and then forwards the DHCP Request message to the DHCP server. An administrator can obtain information about the DHCP client from the Option 82 field, such as the VLAN ID of the switch interface connected to the DHCP client, Layer 2 port number, and MAC address of the relay agent.
- Vendor-specific information option (Option 43)
  - DHCP servers and DHCP clients exchange vendor-specific information through Option 43. When a DHCP server receives a DHCP Discover message with parameter 43 encapsulated in Option 55, it encapsulates Option 43 in a DHCP Offer message and sends the message to the DHCP client.
  - When a device functions as the DHCP server on a WLAN, it can deliver the AC's IP address to connected APs, which facilitates the connection setup between the AC and APs.

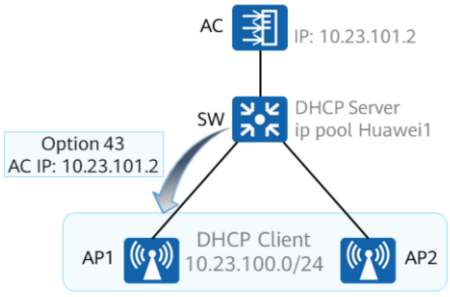
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Commonly used sub-options:
  - Sub-Option1 (Agent Circuit ID Sub-option) The sub-option is usually configured on the DHCP relay agent. It defines the VLAN ID and Layer 2 port number of the switch interface connected to the DHCP client when messages are transmitted. Sub-Option 1 and Sub-Option 2 are used together to identify the DHCP source.
  - Sub-Option 2 (Agent Remote ID Sub-option) This sub-option is usually configured on the DHCP relay agent. It defines that the MAC address of the DHCP relay agent carried in the messages to be transmitted.
  - Sub-Option 5 (Link-selection Suboption): This sub-option contains the IP address added by the DHCP relay agent. In this way, the DHCP server can assign an IP address that is on the same network segment as the IP address to the DHCP client.

Application Example of Option 43

- On a Layer 3 WLAN, when an AP goes online, it needs to obtain the IP address of the AC and establish a CAPWAP tunnel with the AC.
- The IP address of the AP is allocated by the DHCP server. If the IP address of the AC and that of the AP are in different broadcast domains, the AP cannot obtain the IP address of the AC in broadcast mode. As a result, the CAPWAP tunnel cannot be set up.
- The AP obtains the AC's IP address from the Option 43 field in DHCP messages. After obtaining the AC's IP address, the AP establishes a CAPWAP tunnel with the AC and goes online.

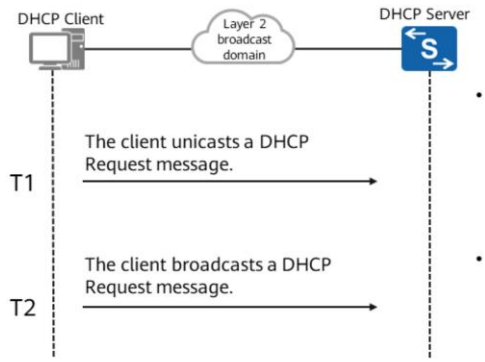


The IP address of the AC is 10.23.101.2, and the gateway address of the network where the AP is located is 10.23.100.1. The AP obtains the IP address from the IP address pool **Huawei1** through DHCP. The DHCP server advertises the IP address of the AC to the AP through the Option 43 field.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCP Address Lease Renewal

The DHCP client generates different renewal requests based on the remaining lease of the IP address.



- When the lease reaches 50% (T1) of its validity period, the DHCP client sends a unicast DHCP Request message to the DHCP server to request lease renewal. If the DHCP client receives a DHCP ACK message from the DHCP server, the lease is successfully renewed.
- If no response is received from the DHCP server when the lease reaches 87.5% (T2) of its validity period, the DHCP client sends a broadcast DHCP Request message to request lease renewal. If the DHCP client receives a DHCP ACK message from the DHCP server, the lease is successfully renewed.
- If no response is received when the lease expires, the DHCP client stops using the IP address and sends a DHCP Discover message to apply for a new IP address.

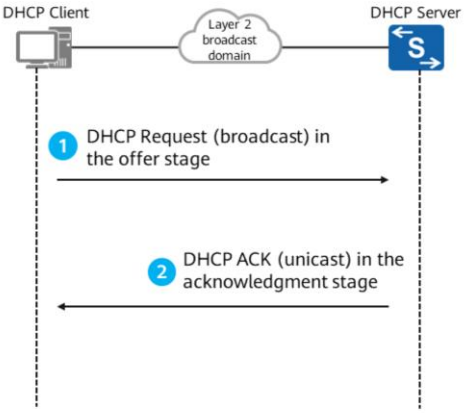
Question: Why is the same IP address assigned to a computer each time?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The DHCP server defines a validity period for each IP address allocated to a DHCP client. The validity period is called the lease. If the DHCP client still needs to use the IP address before the lease expires, the DHCP client can request to extend the lease. If the IP address is not required, the DHCP client can release it. If no idle IP address is available, the DHCP server assigns the IP address released by the client to another client.
- If the DHCP client receives a DHCP NAK message after sending a DHCP Request message at T1 or T2, the DHCP client sends a DHCP Discover message to request a new IP address.
- If a DHCP client does not need to use the allocated IP address before the lease expires, the DHCP client sends a DHCP Release message to the DHCP server to request IP address release. The DHCP server saves the configuration of this DHCP client and records the IP address in the allocated IP address list. The IP address can then be allocated to this DHCP client or other clients. A DHCP client can send a DHCP Inform message to the DHCP server to request configuration update.

DHCP Client Reuse of an IP Address

If a DHCP client reconnects to the network, it can reuse an IP address that has been allocated to it. For example, a host on a network functions as a DHCP client. When the host is powered off and then powered on, it needs to obtain related network parameters again. In this case, the host can request to allocate an IP address that has been used.

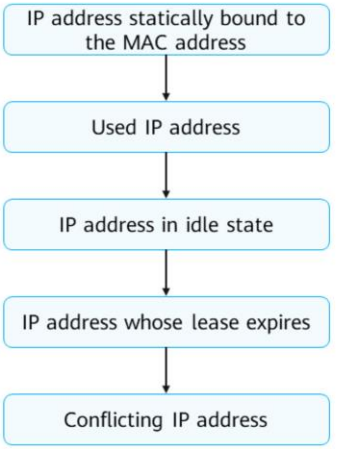


1. Offer stage
   - The DHCP client broadcasts a DHCP Request message carrying the IP address that the client has used. The requested IP address is added in the Option 50 field.
2. Acknowledgment stage
   - After receiving the DHCP Request message, the DHCP server checks whether there is a lease record based on the MAC address in the message. If there is a lease record matching the MAC address, the DHCP server replies with a DHCP ACK message to notify the DHCP client that the requested IP address can be used. If there is no lease record, the DHCP server does not respond to the message.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Not all clients can reuse IP addresses that have been allocated to them.

IP Address Allocation Sequence



A DHCP server assigns IP addresses to a client in the following sequence:

- IP address that is in the database of the DHCP server and is statically bound to the MAC address of the client
- IP address that has previously been assigned to the client, that is, IP address in the requested IP Addr Option of the DHCP Discover message sent by the client
- IP address that is first found when the DHCP server searches the DHCP address pool for available IP addresses
- If the DHCP address pool has no available IP address, the DHCP server searches the expired IP addresses and conflicting IP addresses, and then assigns a valid IP address to the client. If all the IP addresses are in use, an error message is reported.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCP Configuration Commands (1)

1. Create a global address pool.

   [Huawei]**ip pool** *ip-pool-name*

2. Configure the gateway address of the DHCP client.

   [Huawei-ip-pool-HW]**gateway-list** *ip-address*

3. Configure the range of IP addresses that can be allocated dynamically in the global address pool.

   [Huawei-ip-pool-HW]**network** *ip-address* [**mask** { *mask* | *mask-length* }]

4. Configure the IP addresses that are not automatically allocated in the address pool.

   [Huawei-ip-pool-HW]**excluded-ip-address** *start-ip-address* [*end-ip-address*]

5. Set the lease for IP addresses in the address pool.

   [Huawei-ip-pool-HW] **lease** { **day** *day* [**hour** *hour* [**minute** *minute*]] | **unlimited** }

6. Configure the DHCP server to allocate a fixed IP address to a specified DHCP client.

   [Huawei-ip-pool-HW] **static-bind ip-address** *ip-address* **mac-address** *mac-address* **[option-template** *template-name* **| description** *description* **]**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In this example, the name of the IP address pool is **HW**.
- By default, a DHCP server does not allocate fixed IP addresses to specified clients.

DHCP Configuration Commands (2)

1. Configure an interface address pool.

   [Huawei]**interface** *interface-type interface-number [subinterface-number]*

   [Huawei-GigabitEthernet0/0/1]**ip address** *ip-address { mask | mask-length }*

   The IP address segment of the interface is the interface address pool. The interface address mask cannot be set to 31; otherwise, the interface address pool may fail to be configured.

2. Configure a gateway IP address in the interface address pool.

   [Huawei-GigabitEthernet0/0/1]**dhcp server gateway-list** *ip-address*

3. Configure the DHCP server to allocate a fixed IP address to a specified DHCP client.

   [Huawei-GigabitEthernet0/0/1]**dhcp server static-bind ip-address** *ip-address* mac-address *mac-address* [description *description*]

4. Configure the IP addresses that are not automatically allocated in the address pool.

   [Huawei-GigabitEthernet0/0/1]**dhcp server excluded-ip-address** *start-ip-address [end-ip-address]*

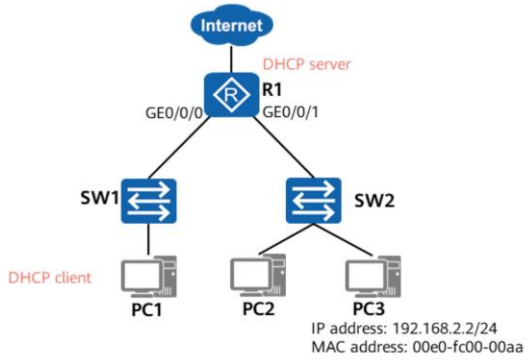
5. Set the lease for IP addresses in the address pool.

   [Huawei-GigabitEthernet0/0/1]**dhcp server lease** { **day** *day* [**hour** *hour* [**minute** *minute*]] | **unlimited** }

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- GigabitEthernet0/0/1 is used as an example.

DHCP Configuration Example



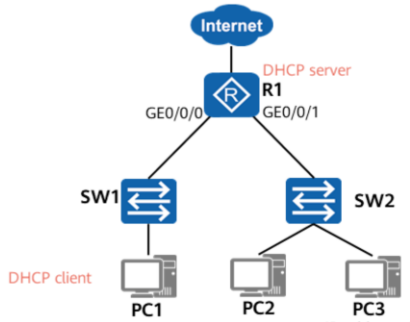

```
[R1]dhcp enable
[R1]ip pool HW
[R1-ip-pool-HW]gateway-list 192.168.1.1
[R1-ip-pool-HW]network 192.168.1.0 mask 24
[R1-ip-pool-HW]excluded-ip-address 192.168.1.200 192.168.1.254
[R1]interface GigabitEthernet 0/0/0
[R1-GigabitEthernet0/0/0]dhcp select global # Select a global address pool.
```

Requirements:

- Assign an IP address to PC1 based on the global address pool.
- Assign IP addresses to PC2 and PC3 based on the interface address pool, and assign fixed IP addresses to PC3.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCP Configuration Example




```
[R1]interface  GigabitEthernet 0/0/1
[R1-GigabitEthernet0/0/1]ip  address 192.168.2.1 24
[R1-GigabitEthernet0/0/1]dhcp select interface     # Select an interface address pool.
[R1-GigabitEthernet0/0/1]dhcp  server  excluded-ip-address  192.168.2.254
[R1-GigabitEthernet0/0/1]dhcp server static-bind ip-address 192.168.2.2 mac-address 00e0-fc00-00aa  #Assign a fixed IP address to PC3.
```

Requirements:

- Assign an IP address to PC1 based on the global address pool.
- Assign IP addresses to PC2 and PC3 based on the interface address pool, and assign fixed IP addresses to PC3.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCP Configuration Result

```
[R1]display ip pool
Pool-name      :          HW
  Gateway-0    :          192.168.1.1
  Mask         :          255.255.255.0
  IP address Statistic
    Total:          253
    Used:             2     Idle:       198
    Expired:          0     Conflict:   0       Disable:   55
```

```
PC3>ipconfig
IPv4 address.....: 192.168.2.2
Subnet mask.....: 255.255.255.0
Gateway.....: 192.168.2.1
Physical address.....: 54-89-98-86-2B-F4
```

PC3 obtains a statically bound IP address.

IP addresses in an address pool can be in one of the following states:

- Used: indicates that the IP address is used.
- Idle: indicates that the IP address is idle.
- Expired: indicates that the lease of the IP address expires and the IP address is idle.
- Conflict: indicates that the IP address conflicts with another IP address on the network.
- Disable: indicates that the IP address cannot be used.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

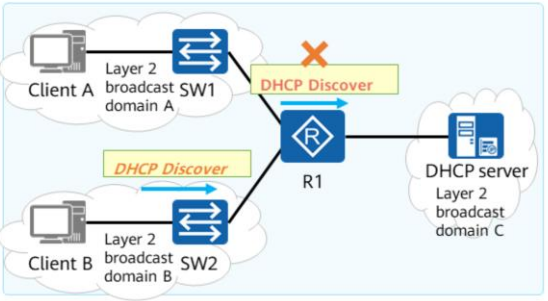
1. DHCP Background
2. DHCP Working Mechanism and Configuration
3. **DHCP Relay Working Mechanism and Configuration**

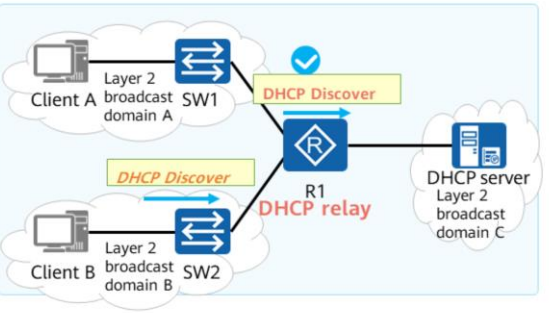
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Introduction to DHCP Relay

- As the network scale expands and the number of network devices increases, different users in an enterprise may be distributed on different network segments. In normal cases, a DHCP server cannot meet address allocation requirements of multiple network segments. If the DHCP server is required to allocate IP addresses, DHCP messages need to be sent across network segments.
- A DHCP relay agent transparently transmits DHCP messages between a DHCP client and a DHCP server that reside in different broadcast domains. The DHCP relay function allows DHCP clients and DHCP server that are in different broadcast domains to communicate.





Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCP Relay Message Format

A DHCP relay agent forwards DHCP messages between a DHCP client and a DHCP server, so the DHCP relay agent modifies only some fields in DHCP messages. The message format remains unchanged, as shown in the following figure.

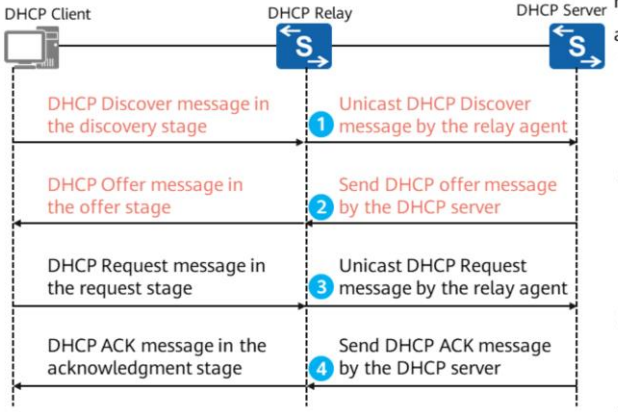


Hops: indicates the number of DHCP relay agents that a DHCP message passes through. This field is set to 0 by a DHCP client or server. Its value increases by 1 each time the message passes through a DHCP relay agent.
Giaddr (gateway ip address): indicates the IP address of the first DHCP relay agent. When a client sends a DHCP Request message, the first DHCP relay agent fills its IP address in this field when forwarding the message to the DHCP server.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The Hops field limits the number of DHCP relay agents that a DHCP message can pass through. A maximum of 16 DHCP relay agents are allowed between a DHCP server and a DHCP client. If the value of this field is larger than 16, DHCP messages are discarded.
- The DHCP server determines the network segment address of a client based on the Giaddr field, so the DHCP server can select an appropriate address pool and assign an IP address on the network segment to the client. The DHCP server returns a DHCP Offer message to the DHCP relay agent.The DHCP relay agent then forwards the DHCP Offer message to the client. If the DHCP Discover message passes through multiple DHCP relay agents before reaching the DHCP server, the value of this field is the IP address of the first DHCP relay agent and remains unchanged. However, the value of the Hops field increases by 1 each time the DHCP Discover message passes through a DHCP relay agent.

Working Mechanism of DHCP Relay



Working mechanism when a DHCP client connects to a network for the first time in the scenario where a DHCP relay agent is deployed

1. Discovery stage: After receiving the DHCP Discover message broadcast by a DHCP client, the DHCP relay agent unicasts the DHCP Discover message to the DHCP server or the next-hop relay agent through routing.
2. Offer stage: The DHCP server selects an address pool based on the Giaddr field in the DHCP Discover message to allocate network parameters to the DHCP client. After receiving the DHCP Offer message, the DHCP relay agent unicasts or multicasts the message to the DHCP client.
3. Request stage: The DHCP relay agent processes the DHCP Request message from the client using the same method described in the Discovery stage.
4. Acknowledgment stage: The DHCP relay agent processes the DHCP ACK message from the server using the same method described in the Offer stage.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

1. After receiving a DHCP Discover message, the DHCP relay agent processes the message as follows:
   - Checks the value of the Hops field. If this value exceeds 16, the DHCP relay agent discards the message. Otherwise, the DHCP relay agent increases this value by 1 and proceeds to the next step.
   - Checks the value of the Giaddr field. If this value is 0, the DHCP relay agent sets the Giaddr field to the IP address of the interface receiving the DHCP Discover message. If not, the DHCP relay agent does not change the field and proceeds to the next step.
   - Changes the destination IP address of the DHCP Discover message to the IP address of the DHCP server or the next-hop DHCP relay agent, and changes the source IP address to the IP address of the interface connecting the DHCP relay agent to the client. The message is then unicast to the DHCP server or the next-hop DHCP relay agent.
2. After receiving the DHCP Discover message, the DHCP server selects an address pool on the same network segment as the value of the Giaddr field in the message, allocates parameters such as an IP address to the client, and unicasts a DHCP Offer message to the DHCP relay agent identified by the Giaddr field. After receiving the DHCP Offer message, the DHCP relay agent performs the following operations:
   - Checks the value of the Giaddr field. If this value is not the IP address of the interface receiving the DHCP Offer message, the DHCP relay agent discards the

message. Otherwise, the DHCP relay agent proceeds to the next step.

- Checks the value of the Flags field. If this value is 1, the DHCP relay agent sends a broadcast DHCP Offer message to the DHCP client. Otherwise, the DHCP relay agent sends a unicast DHCP Offer message.

DHCP Relay Configuration Commands

1. Enable the DHCP relay function on an interface.

   [Huawei-GigabitEthernet0/0/0]**dhcp select relay**

2. Specify an IP address for the DHCP server in the interface view.

   [Huawei-GigabitEthernet0/0/0]**dhcp relay server-ip** *ip-address*

3. Create a DHCP server group.

   [Huawei]**dhcp server group** *group-name*

4. Add DHCP servers to the DHCP server group.

   [Huawei-dhcp-server-group-HW]**dhcp-server** *ip-address* [*ip-address-index*]

5. Configure a DHCP server group for the interface.

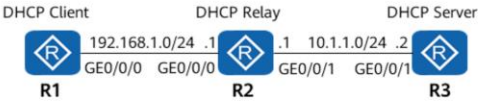
   [Huawei-GigabitEthernet0/0/0]**dhcp relay server-select** *group-name*

6. Enable the DHCP client function on the interface.

   [Huawei-GigabitEthernet0/0/0]**ip address dhcp-alloc**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCP Relay Configuration Example (1)



The configuration requirements are as follows:

- R1 obtains an IP address through DHCP.
- The DHCP relay function is enabled on GE0/0/0 of R2 and the IP address of the DHCP server is set to 10.1.1.2.
- An address pool named **HW-1** is created on R3, with the IP address range of 192.168.10/24 and gateway address of 192.168.1.1.

Configuration of R1 and R2:

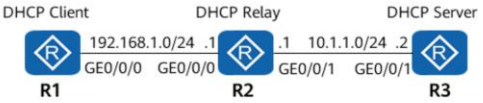
```
[R1]interface GigabitEthernet0/0/0
[R1-GigabitEthernet0/0/0]ip address dhcp-alloc
[R1-GigabitEthernet0/0/0]quit
```

```
[R2]dhcp server group HW
[R2-dhcp-server-group-HW]dhcp-server 10.1.1.2
[R2-dhcp-server-group-HW]quit
[R2]interface GigabitEthernet 0/0/1
[R2-GigabitEthernet0/0/1]ip address 10.1.1.1 24
[R2-GigabitEthernet0/0/1]quit
[R2]interface GigabitEthernet 0/0/0
[R2-GigabitEthernet0/0/0]ip address 192.168.1.1 24
[R2-GigabitEthernet0/0/0]dhcp select relay
[R2-GigabitEthernet0/0/0]dhcp relay server-select HW
[R2-GigabitEthernet0/0/0]quit
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Before configuring DHCP on each device, run the **dhcp enable** command in the system view to enable DHCP.

DHCP Relay Configuration Example (2)



The configuration requirements are as follows:

- GE0/0/0 on R1 obtains an IP address through DHCP.
- The DHCP relay function is enabled on GE0/0/0 of R2 and the IP address of the DHCP server is set to 10.1.1.2.
- An address pool named **HW-1** is created on R3, with the IP address range of 192.168.10/24 and gateway address of 192.168.1.1.

Configuration of R3:

```
[R3]ip pool HW-1
[R3-ip-pool-HW-1]network 192.168.1.0 mask 24
[R3-ip-pool-HW-1]gateway-list 192.168.1.1
[R3-ip-pool-HW-1]quit
[R3]interface GigabitEthernet 0/0/1
[R3-GigabitEthernet0/0/1]ip address 10.1.1.2 24
[R3-GigabitEthernet0/0/1]dhcp select global
[R3-GigabitEthernet0/0/1]quit
[R3]ip route-static 192.168.1.0 255.255.255.0 10.1.1.1
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

DHCP Relay Configuration Verification

Check the IP address obtained by GE0/0/0 on R1.

```
<R1>display dhcp client
DHCP client lease information on interface GigabitEthernet0/0/0 :
  Current machine state          : Bound
  Internet address assigned via  : DHCP
  Physical address               : 00e0-fce6-4691
  IP address                     : 192.168.1.254
  Subnet mask                    : 255.255.255.0
  Gateway ip address             : 192.168.1.1
  DHCP server                    : 10.1.1.2
  ......
```

The command output shows that GE0/0/0 on R1 has obtained an IP address and a gateway address.

Check the DHCP relay information on R2.

```
<R2>display dhcp relay all
 DHCP relay agent running information of interfaceGigabitEthernet0/0/0 :
 Server group name                 : HW
 Gateway address in use            : 192.168.1.1
<R2>display dhcp relay statistics
 The statistics of DHCP RELAY:
DHCP packets received from clients  : 2
DHCP packets sent to clients        : 2
DHCP packets received from servers  : 2
DHCP packets sent to servers        : 2
......
```

The command output shows that the DHCP relay function is enabled on GE0/0/0 of R2, and R2 functions as the DHCP relay agent to exchange messages with the DHCP client and DHCP server for four times.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



Quiz

1. (Single) Which of the following messages is sent by the DHCP client to the DHCP server to renew the lease?
   A. DHCP DISCOVER
   B. DHCP OFFER
   C. DHCP REQUEST
   D. DHCP ACK
2. (Single) Which of the following commands can be used to enable the DHCP relay function on a router interface?
   A. dhcp select server
   B. dhcp select global
   C. dhcp select interface
   D. dhcp select relay

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

1. C
2. D

Summary

- This course describes the format of DHCP messages and functions of key fields. It describes how a DHCP server assigns network parameters, such as IP addresses, to DHCP clients through message exchange. In addition, this course describes how a DHCP client renews its IP address lease and obtains an IP address after restart.
- When a DHCP client and a DHCP server are in different broadcast domains, the DHCP relay function enables DHCP messages to be transmitted across network segments. This function implements message negotiation between the DHCP client and the DHCP server.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Introduction to Network Management Protocols

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- With the rapid development of new technologies such as artificial intelligence, big data, and cloud computing, various industries will be marching towards digital transformation in the near future. In addition, enterprise services will become diversified along with digital transformation, making network management one of the most challenging tasks for us. We cannot manage the entire network manually. Instead, we are in desperate need of an automated, smart tool to help us manage devices and resources on the network.
- This course starts with a brief introduction about the development history of network management and the related concepts, followed by a detailed description of the protocols used in network management and their respective application scenarios. Finally, typical application cases of network management are provided.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to understand:
  - The development trend and functions of network management.
  - The functions of network management protocols.
  - The similarities and differences between the implementation mechanisms and application scenarios of different network management protocols.
  - The application scenarios of network management protocols.

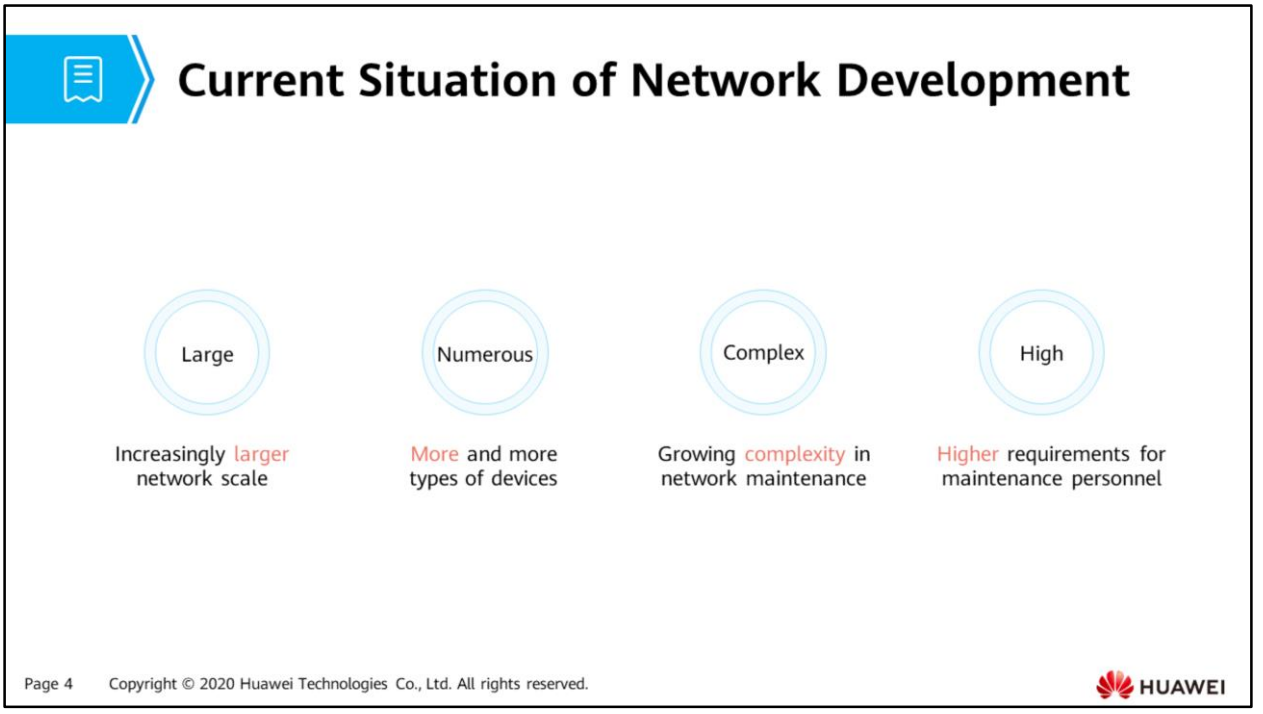
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

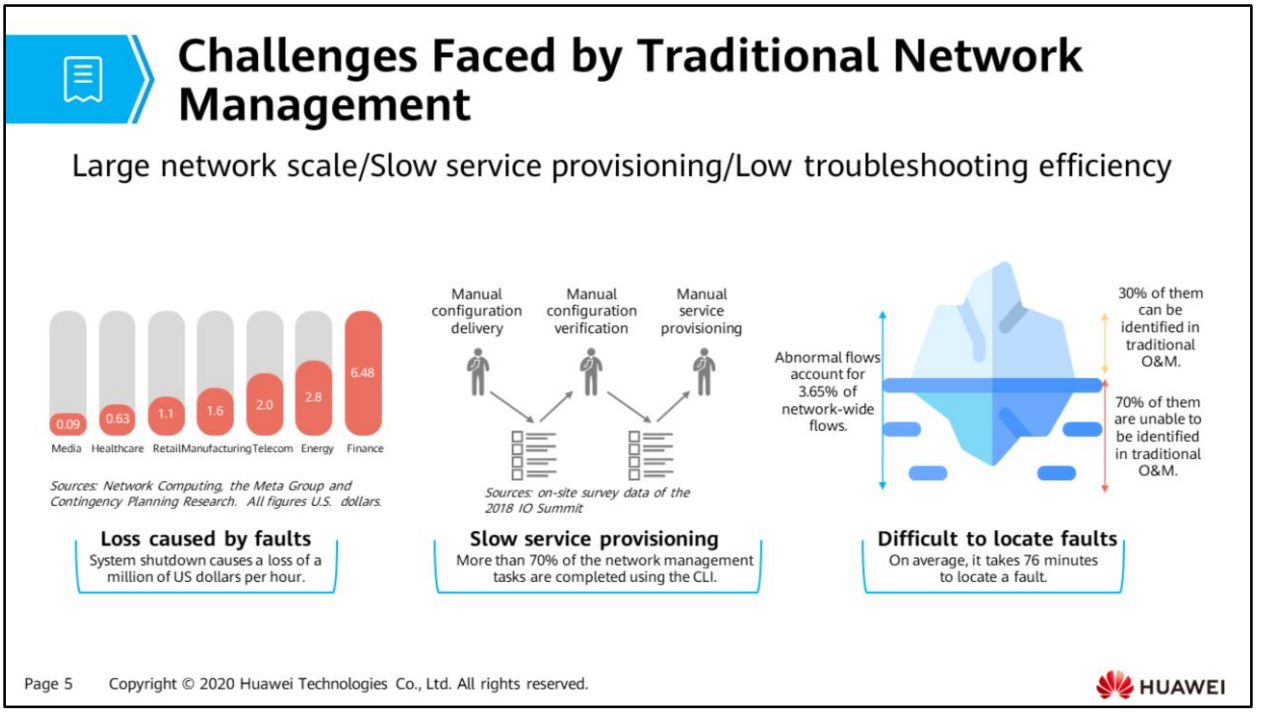
1. **Development History of Network Management**
2. Functions of Network Management
3. Categories of Network Management Protocols
4. Application Scenarios of Network Management

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

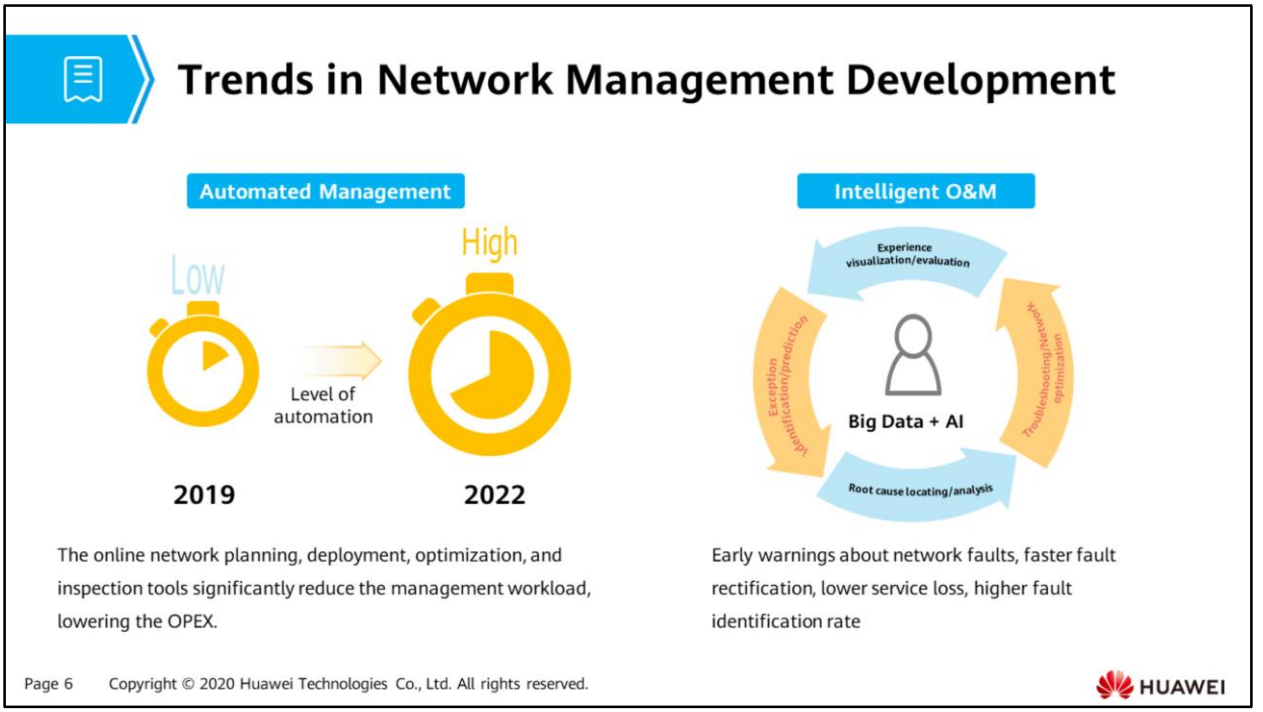
HUAWEI



- The rapid development of the Internet industry brings about great changes to networks, with multi-service convergence as the major trend in future network development. Network convergence requires management convergence, that is, the unified network management system is required to centrally manage multiple services and devices.



- As the new technologies such as artificial intelligence, big data, and cloud computing are developing rapidly, industries will undergo digital transformation in the next decade and enterprise services will become diversified during the implementation of digital transformation. Digitalization brings changes to network models, and the traditional network management mode can no longer meet the new requirements of digital services. To be specific, traditional network construction, management, and O&M methods cannot meet new network requirements that arise during digitalization.



- OPEX means operating expense, which is the sum of the maintenance cost, marketing expense, labor cost, and depreciation expense during the enterprise operations.
- In April 2019, a well-known consulting firm in the industry released a report about using AI and automation to improve network reliability. According to this report, 65% of the enterprises will have network automation technologies deployed on their campus networks by 2022. The proportion, however, is only 17% today.
- Automated management: Network management is just like domestic washing machines, which evolve from manual to semi-automated, then to fully automated and even intelligent washing today, making it possible for everyone to operate a complex machine and complete complex tasks. This is also true for network management. It starts with commands-based per-device configuration and management, then evolves to the graphical user interface-based management and control system, and finally to today's service language-based automatic network configuration. Among all the time an enterprise spends in network management, almost one third is invested in network planning and deployment. In the future, network automation will be implemented in two aspects:
  - Full-lifecycle automation: means whether tools can be used to implement automation in the full lifecycle covering network planning, deployment, policy provisioning, network status monitoring, maintenance, and management.
  - Network-wide automation: means whether enterprise LAN, WLAN, and WAN networks can be centrally managed and policies can be configured in a unified manner, and whether service policies can be defined globally based on user

identities and application types.

Contents

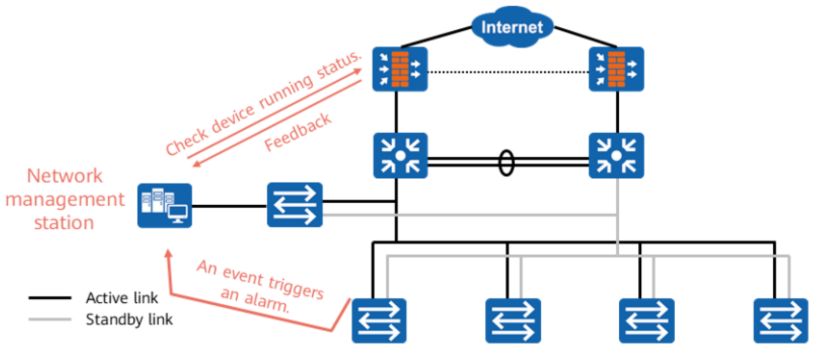
1. Development History of Network Management
2. **Functions of Network Management**
3. Categories of Network Management Protocols
4. Application Scenarios of Network Management

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

What Is Network Management?

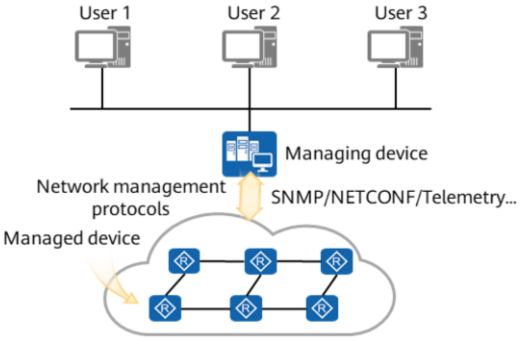
- Network management is a basic requirement of a network. Currently, users can manage a network using various protocols, tools, applications, or devices. The managed objects include hardware and software on the network.
- The purpose of network management is to ensure that a network runs properly by means of monitoring, testing, configuring, analyzing, evaluating, and controlling the software and hardware on the network.

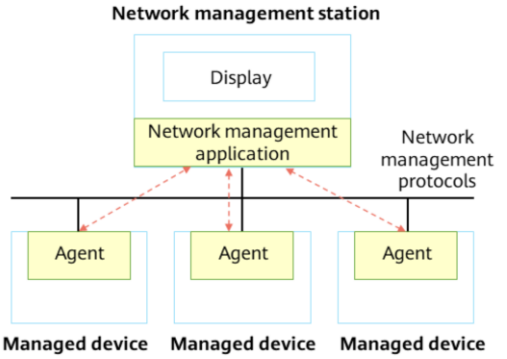


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

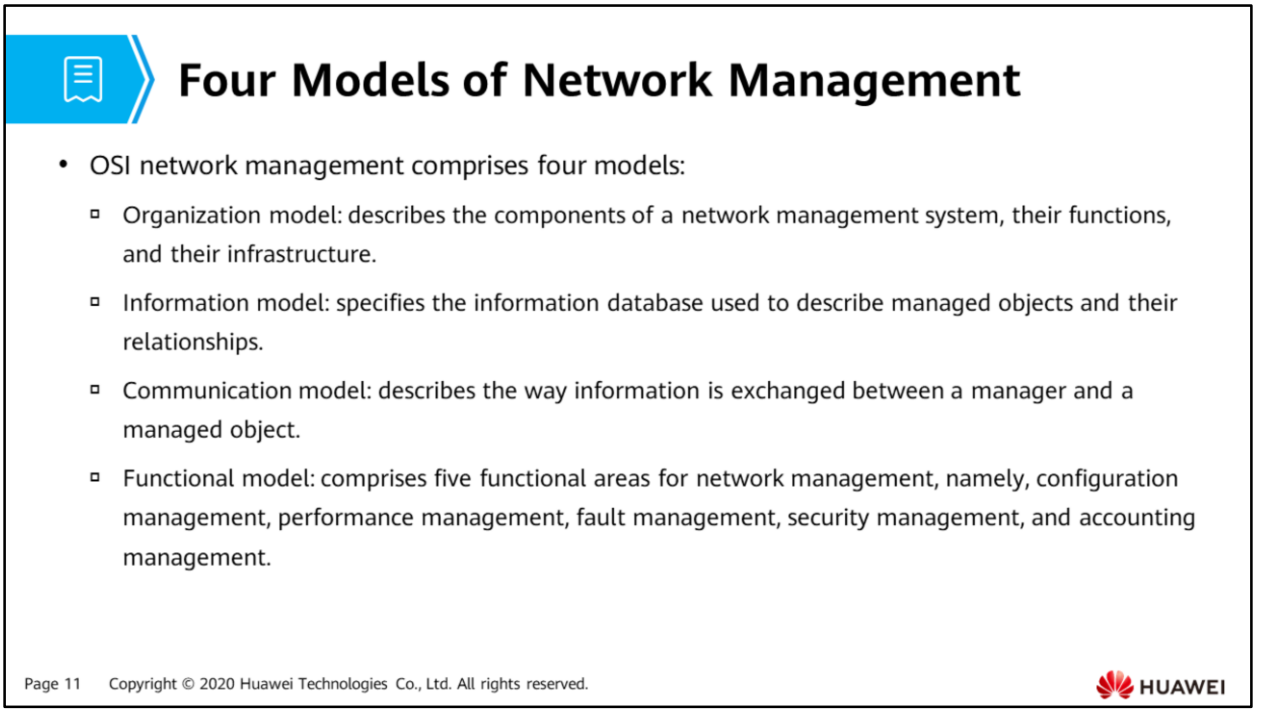
Typical Architecture of Network Management Systems

- Network management systems usually have the same basic architecture, which consists of two key elements:
  - Managing device, also called a network management station
  - Managed device, also called an agent

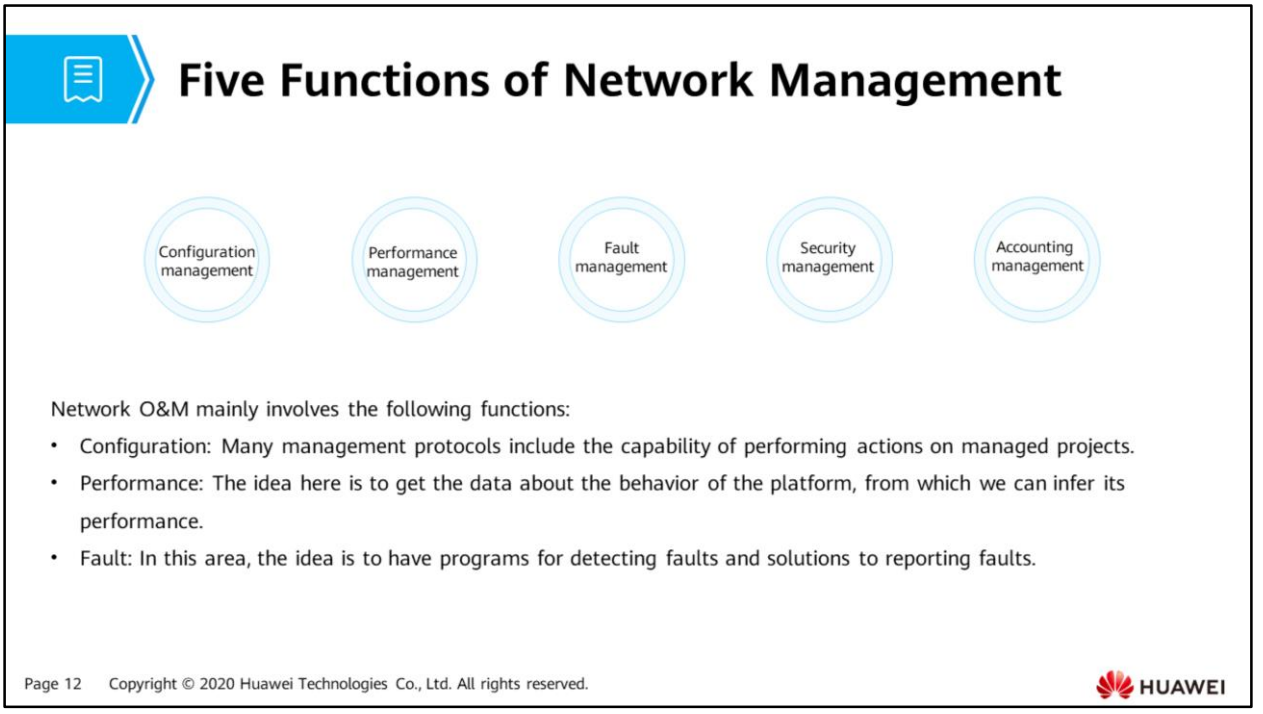




Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- The organization model defines the terms manager, agent, and managed object. It describes the components of a network management system, their functions, and their basic architecture.
- The information model is related to the relationship and storage of management information. It specifies the information database that describes the managed objects and their relationships. The structure of management information (SMI) defines the syntax and semantics of the management information stored in the Management Information Base (MIB). Both the agent process and manager process use the MIB to exchange and store management information.
- The communication model deals with the way information is exchanged between agents and managers and between managers. The communication model contains three key elements: transport protocol, application protocol, and the actual message to be transmitted.
- The functional model defines five functional areas for network management: configuration management, performance management, fault management, security management, and accounting management.



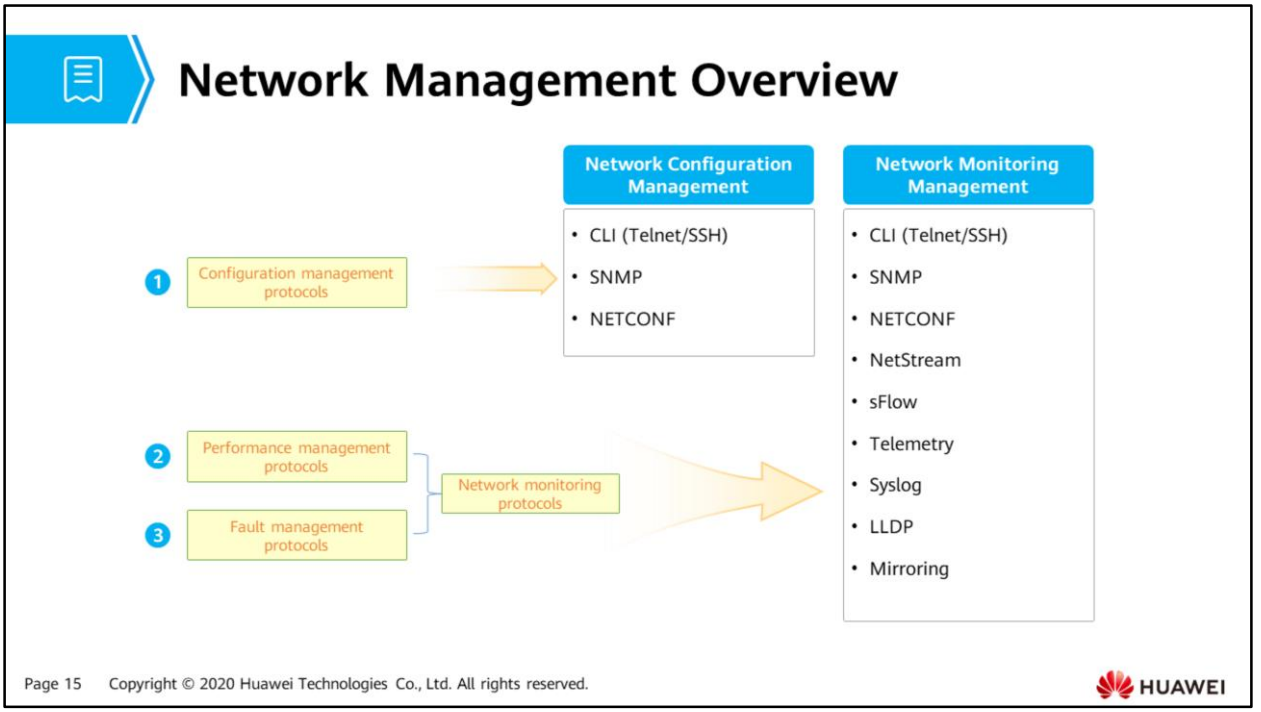
- OSI defines five functional models for network management:
- Configuration management:
  - Configuration management is concerned with initializing a network, provisioning the network resources and services, and monitoring and controlling the network. More specifically, the responsibilities of configuration management include setting, maintaining, adding, and updating the relationship among components and the status of the components during network operation.
  - Configuration management consists of both device configuration and network configuration. Device configuration can be performed either locally or remotely. Automated network configuration, such as Dynamic Host Configuration Protocol (DHCP) and Domain Name System (DNS), plays a key role in network management.
- Performance management:
  - Performance management is concerned with evaluating and reporting the behavior and the effectiveness of the managed network objects. A network monitoring system can measure and display the status of the network, such as collecting statistics about the traffic volume, network availability, response time, and throughput.

Contents

1. Development History of Network Management
2. Functions of Network Management
3. **Categories of Network Management Protocols**
4. Application Scenarios of Network Management

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

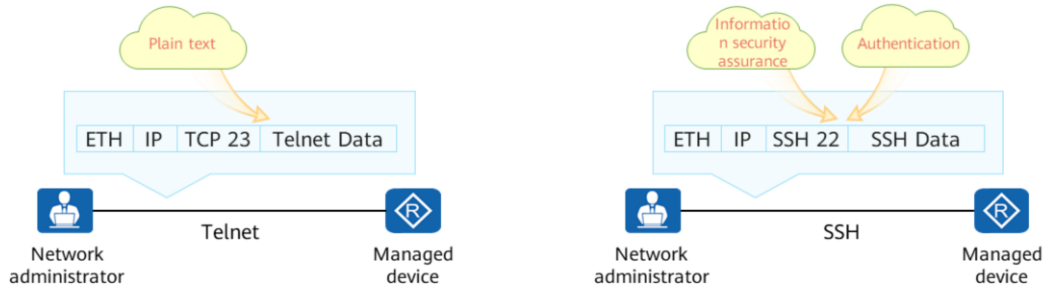


- The command-line interface (CLI) supports both network configuration management and network monitoring management.
- The Set function of the Simple Network Management Protocol (SNMP) supports network configuration management, and its Trap function supports network monitoring management.
- The Edit function of the Network Configuration Protocol (NETCONF) supports network configuration management, and its Get function supports network monitoring management.

CLI (Telnet/SSH)

Supports network configuration management and network monitoring management.

The CLI is the most widely used user interface before the graphical user interface is popularized. When using the CLI, users input instructions using a keyboard, rather than using a mouse. A device will then execute the instructions they receive.



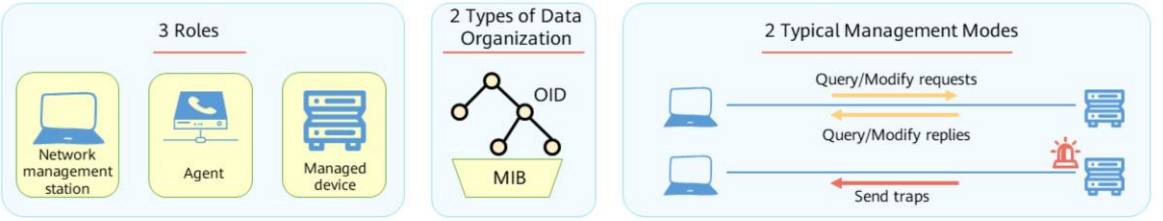
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A network administrator can use the CLI to configure devices and monitor networks, which are simple and convenient. However, automation tools must be used to perform batch configuration, once large-scale deployment is needed.
- Telnet is an abbreviation of the words “telecom (Telecommunications) networks".
  - Telnet uses the dedicated TCP port 23. Telnet is not a secure communications protocol and it transmits data, including passwords, in plain text over the network or Internet.
  - Telnet does not use any authentication policies or data encryption techniques.
- SSH (Secure Shell)
  - SSH uses the dedicated TCP port 22. It is a secure protocol that transmits encrypted data over the network or Internet. Once encrypted, it is extremely difficult to extract and read the data.
  - SSH uses public keys to authenticate access users, which provides higher security.
- Telnet and SSH are two methods for remotely managing devices, among which SSH is more secure. Therefore, SSH is usually a required protocol on the networks.

SNMP

Supports network configuration management and network monitoring management.

SNMP is a standardized network management protocol widely used on TCP/IP networks. It uses a central computer — known as a network management station (NMS) — that runs network management software to manage network elements. Three SNMP versions are available: SNMPv1, SNMPv2c, and SNMPv3. Users can select one or multiple versions if needed.

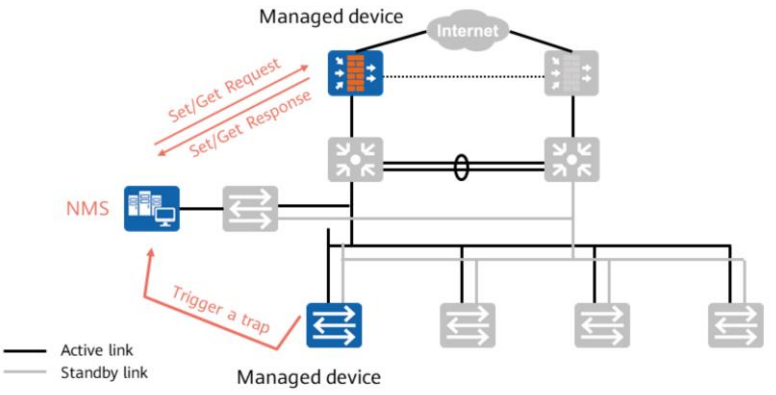


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- NMS: The NMS sends various query packets to and receives traps from managed devices.
- Managed devices refer to the devices that are managed by the NMS.
- An agent is a process residing on a managed device. An agent provides the following functions:
  - Receives and parses query packets from the NMS.
  - Reads or writes management variables based on the packet type, generates response packets, and sends the response packets to the NMS.
  - Proactively generates a trap when an event occurs (for example, when a port goes up or down, the STP topology changes, or the OSPF neighbor relationship is down) based on the trap triggering conditions defined by each protocol module, and reports the event to the NMS.
- The Management Information Base (MIB) is a database that specifies the variables maintained by managed devices, that is, the information that can be queried and set by the agents. The MIB defines a series of attributes for managed devices, including the name, status, access permission, and data type of the managed objects.
- Object identifier (OID): A MIB uses a tree structure, with each node in the tree indicating a managed object. An object can be uniquely identified by a path, known as the OID, that starts from the root of the tree.

Application Scenarios of SNMP



Configure the SNMP management program in the NMS, enable the agent program on the managed device, and configure the SNMP protocol on the network.

After SNMP is configured:

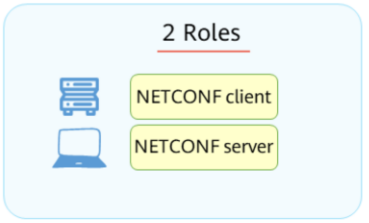
- The NMS can obtain or change device information through the agent to implement remote monitoring and management.
- The agent can report device status to the NMS in a timely manner.

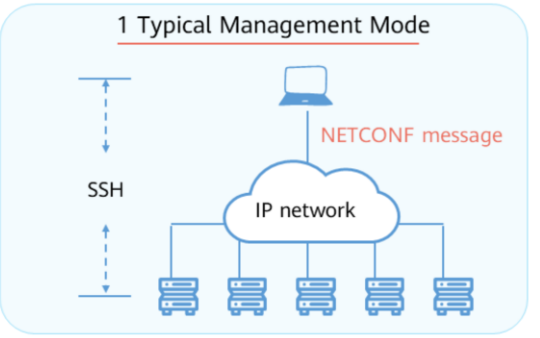
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

NETCONF

Supports network configuration management and network monitoring management.

As an XML-based network configuration protocol, NETCONF is designed to automate network configuration in a programmable manner, so as to simplify and accelerate network service deployment.

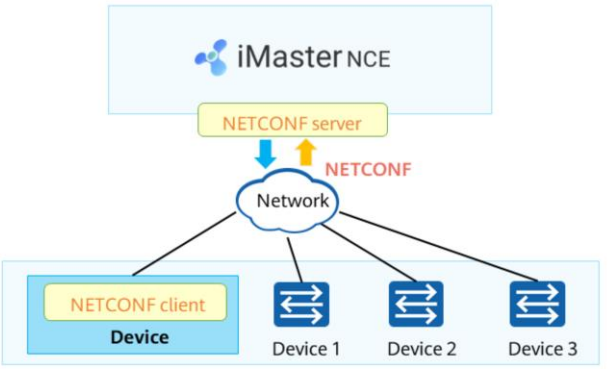




Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- NETCONF uses SSH to secure transmission and uses Remote Procedure Calls (RPCs) to implement communication between the client and server.
- NETCONF messages are presented as XML documents.

Application Scenarios of NETCONF



Advantages of NETCONF:

- Is a standardized protocol that features structured data and is highly scalable.
- Uses a layered framework and supports a wide range of operations.
- Supports multiple configuration databases, through which operation verification and rollback can be implemented based on a transaction mechanism.
- Allows configuration data and status data to be obtained separately, and allows the data to be compared between devices.
- Provides network-level service configuration capabilities and supports network-level configuration transactions.
- Supports configuration backup and restoration.

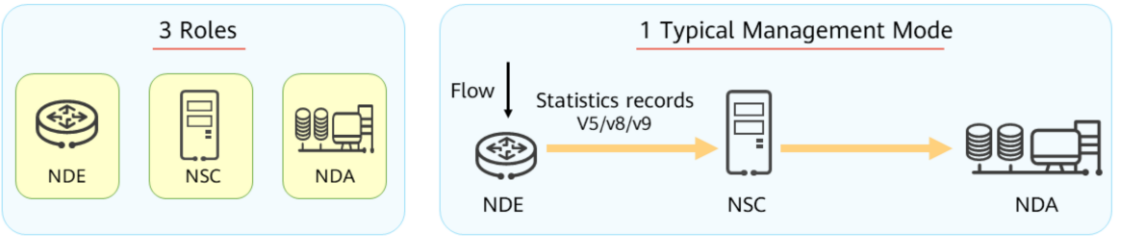
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- NETCONF provides a set of mechanism for managing network devices. With this mechanism, users can add, modify, delete, back up, restore, lock, and unlock network device configurations. In addition, NETCONF provides transaction and session operation functions to obtain network device configuration and status information.

NetStream

Supports network monitoring management.

- NetStream is a technology that collects statistics about and analyzes service traffic on a network based on network flow information. NetStream can be deployed at the access, aggregation, and core layers of a network.
- NetStream can collect statistics for IP packets (UDP, TCP, ICMP packets) and MPLS packets.

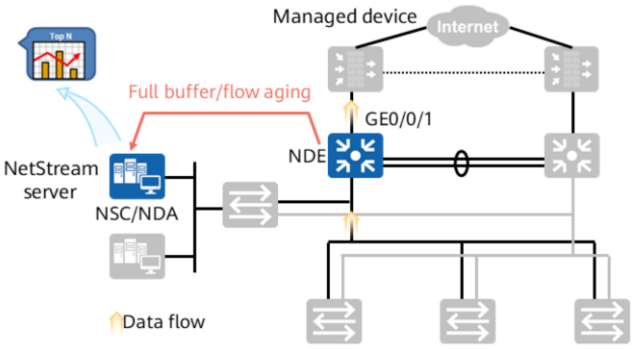


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A typical NetStream system has three components: NetStream data exporter (NDE), NetStream collector (NSC), and NetStream data analyzer (NDA).
  - NDE: An NDE is a device configured with NetStream functions. It analyzes and processes network flows, extracts flows that meet conditions for statistics collection, and exports the statistics to the NDA. The NDE can perform operations (such as aggregation) on the statistics before exporting them to the NDA.
  - NSC: An NSC is a program running in Windows or UNIX that parses packets from NDEs and saves the statistics to a database for the NDA to parse. It can collect, filter, and aggregate data exported from multiple NDEs.
  - NDA: An NDA is a network traffic analysis tool that extracts statistics from the NSC, processes the statistics, and generates reports. The reports provide reference for various services, such as traffic-based charging, network planning, and attack monitoring. Typically, the NDA provides a graphical user interface for users to easily obtain, display, and analyze collected data.
- Flow statistics can be exported in two modes:
  - Original flow statistics export: After the aging timer expires, the statistics of each flow are exported to the NSC. The advantage of this mode is that the NSC can obtain the detailed statistics of all flows.
  - Aggregation flow statistics export: The device summarizes the original flows with the same aggregation keywords to obtain statistics on the aggregation flow. In this way, originals flows are aggregated before they are exported, significantly saving

network bandwidth.

Application Scenarios of NetStream



How NetStream works:

- The device configured with NetStream (that is, NDE) periodically sends collected flow statistics to the NSC.
- The NSC processes the flow statistics and sends them to the NDA.
- The NDA analyzes the traffic statistics and provides them for use in scenarios such as accounting and network planning.

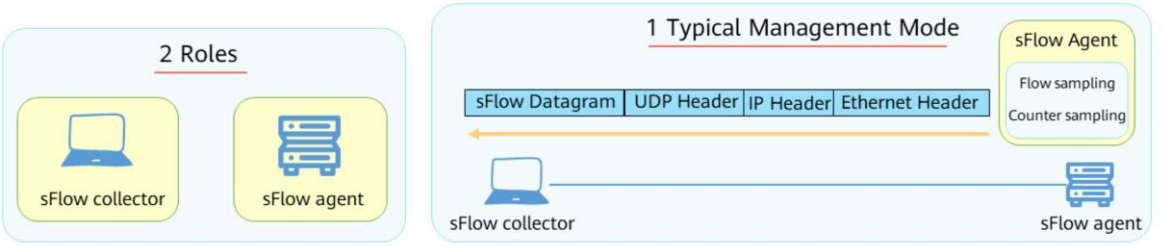
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- In real networking, the NSC and NDA are typically integrated on one NetStream server. The NDE samples packets to obtain outbound traffic information on GE0/0/1 and creates NetStream flows based on certain conditions. When the NetStream buffer is full or a NetStream flow is aged out, the NDE encapsulates statistics in NetStream packets and sends the packets to the NetStream server. The NetStream server analyzes and processes the NetStream packets, and then displays the analysis result.
- Implementation and limitations of traditional traffic statistics collection methods:
  - IP packet-based statistics collection: The collected statistics are simple and include only limited types of information.
  - ACLs: A large number of ACLs are required and statistics about mismatching packets cannot be collected.
  - SNMP: The protocol has limited functions. It collects statistics through continuous polling, wasting CPU and network resources.
  - Port mirroring: This function has high cost and occupies one port of the device. Statistics cannot be collected on ports that do not support mirroring.
  - Physical-layer replication: The cost is high, and dedicated hardware devices need to be purchased.

sFlow

Supports network monitoring management.

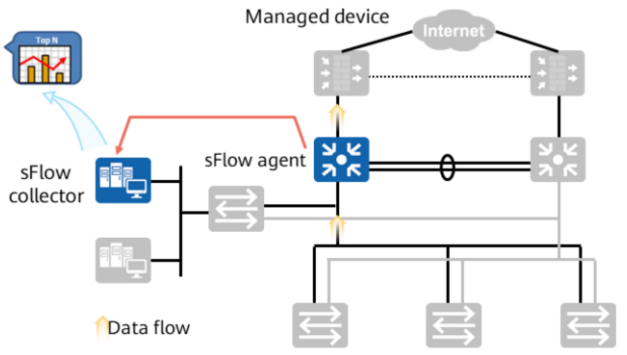
- Sampled flow (sFlow) is a network traffic monitoring technology based on packet sampling.
- The sFlow system involves an sFlow agent embedded in a device and a remote sFlow collector. The sFlow agent obtains traffic statistics on an interface using **sampling**, encapsulates them into **sFlow packets**, and sends the packets to the designated sFlow collector. The sFlow collector analyzes the sFlow packets and displays the traffic statistics.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- With flow sampling, an sFlow agent samples packets in the specified direction on the specified interface based on a sampling rate, and analyzes the packets to obtain information about packet data content. Flow sampling focuses on traffic details, facilitating monitoring and analysis of traffic behaviors on the network.
  - With flow sampling, an sFlow agent can obtain the entire packet or part of the packet header.
- With counter sampling, an sFlow agent periodically obtains traffic statistics on an interface. In contrast with flow sampling, counter sampling focuses on traffic statistics on an interface rather than traffic details.

Application Scenarios of sFlow



Enterprise network users often have explicit requirements for traffic on device interfaces and the overall running status of devices. Enterprises require a traffic monitoring technique that samples packets on device interfaces to promptly find abnormal traffic and the source of attack traffic, so that they can quickly rectify faults to ensure networks can run properly. sFlow focuses on traffic on interfaces, traffic forwarding, and the overall device status, so it can be used to monitor and locate network exceptions, especially on enterprise networks.

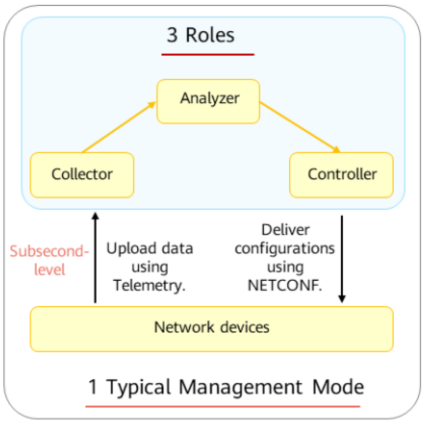
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- As shown in the figure, an sFlow agent is connected to a remote sFlow collector so that traffic statistics can be collected and analyzed based on interfaces.
- NetStream is also a technology that collects and analyzes traffic statistics. In NetStream, a network device preliminarily collects and analyzes traffic statistics and then saves the statistics to a cache. The network device exports the statistics when they expire or when the cache overflows. Different from NetStream, sFlow does not require a cache, because a network device only samples packets and a remote collector will collect and analyze traffic statistics.
- Therefore, sFlow has the following advantages over NetStream:
  - Fewer resources and lower costs: sFlow does not require a cache, so it uses only a small number of resources on network devices, lowering costs.
  - Flexible collector deployment: The collector can be deployed flexibly, enabling traffic statistics to be collected and analyzed based on various traffic characteristics.

Telemetry

Supports network monitoring management.

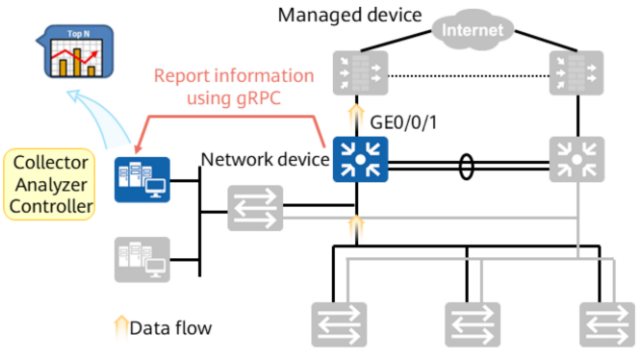
- Telemetry is also called network Telemetry. It is used to monitor networks, including packet check and analysis, intrusion and attack detection, intelligent data collection, and application performance management.
- Advantages:
  - Supports multiple implementation modes, meeting diversified user requirements.
  - Collects a wide variety of data with high precision to fully reflect network status.
  - Continuously reports data with only one-time data subscription.
  - Locates faults rapidly and accurately.



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- With the popularization of networks and emergence of new technologies, the network scale is growing, network deployment is increasingly complex, and users have higher requirements on service quality. To meet user requirements, network O&M must be more refined and intelligent. Network O&M are, however, faced with the following challenges:
  - Ultra-large scale: A large number of devices need to be managed and massive amount of information is monitored.
  - Quick fault locating: Users want faults to be located within seconds or even subseconds on complex networks.
  - Refined monitoring: Various types of data needs to be monitored at a finer granularity to reflect the network status completely and accurately. With the monitoring information, possible faults can be predicted, providing a sound foundation for network optimization. Network O&M involves monitoring not only traffic statistics on interfaces, packet loss on each flow, CPU usage, and memory usage, but also the latency and jitter of each flow, latency of each packet on its transmission path, and buffer usage on each device.
- The collector, analyzer, and controller are components of the network management system.
  - The collector receives and stores monitoring data reported by network devices.
  - The analyzer analyzes the monitoring data received by the collector and processes the data, for example, displays the data on the graphical user interface.
  - The controller uses NETCONF to deliver configurations to devices, so as to manage these devices. To be specific, the controller can deliver configurations to network devices and adjust the forwarding behavior of the network devices based on the data provided by the analyzer. It can also control which data network devices need to sample and report.

Application Scenarios of Telemetry



The gRPC-based Telemetry technology can collect traffic statistics on interfaces, CPU usage, and alarm data of devices, encode the data using Protocol Buffers, and report the data to the collector in real time for storage.

	Telemetry	SNMP Get	SNMP Trap	CLI	Syslog
Working mode	Push	Pull	Push	Pull	Push
Precision	Subsecond-level	Minute-level	Second-level	Minute-level	Second-level

Page 26 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

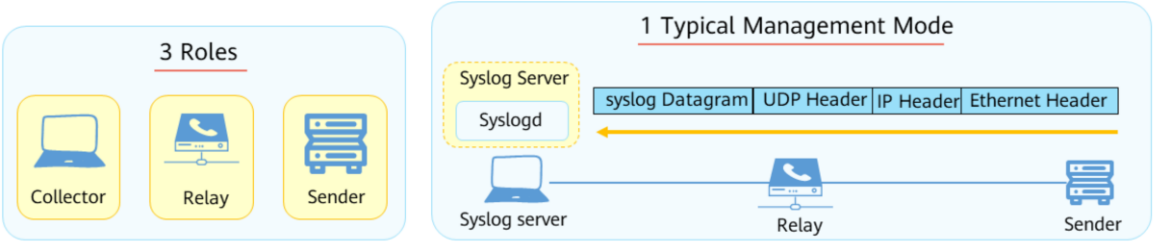
HUAWEI

- Google Remote Procedure Call (gRPC) is a Google-developed open-source high performance RPC framework that uses HTTP/2 as its underlying transport protocol. It provides multiple methods for configuring and managing network devices that are available in multiple programming languages.
- Traditional network monitoring methods (such as SNMP, CLI, and Syslog) cannot meet network O&M requirements.
  - SNMP and CLI obtain data in pull mode. That is, data is obtained from devices using requests. This method limits the number of network devices that can be monitored, and data cannot be quickly obtained using this method.
  - SNMP Trap and Syslog obtain data in push mode. That is, devices proactively report data to the monitoring device. However, they only report events and alarms. The monitoring data is limited and cannot accurately reflect the actual network status.
- Telemetry is a remote data collection technology that monitors device performance and faults. It obtains abundant monitoring data in push mode in a timely manner. The data helps quickly locate network faults and resolve the preceding network O&M problems.

Syslog

Supports network monitoring management.

- System log (Syslog) is an industry standards-compliant protocol for recording device logs. It logs events that occur at any time on UNIX systems, routers, switches, and other network devices. These logs provide administrators with insights into the system status.
- The Syslog protocol offers a mechanism that allows a host to transmit event messages to a receiver host, known as the Syslog server, over an IP network.



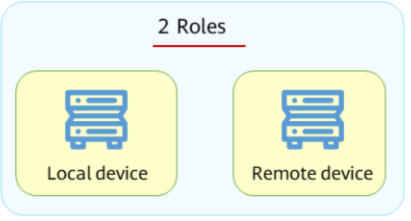
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

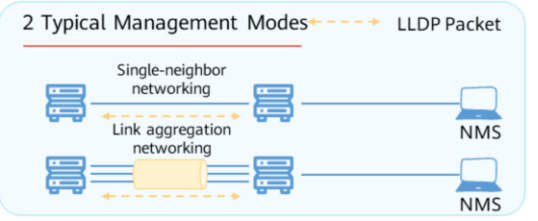
- While this protocol was originally developed on the University of California Berkeley Software Distribution (BSD) TCP/IP system implementations, its value to operations and management has led it to be ported to many other operating systems as well as being embedded into many other networked devices. RFC 3164 and RFC 3195 provide general-purpose definitions for this protocol. The former describes Syslog messages transmitted over UDP, whereas the latter defines Syslog messages transmitted over TCP.
- Almost all network devices can use the Syslog protocol to transport logs to a remote Syslog server over UDP. The remote Syslog server must use syslogd to listen on UDP port 514, process local logs and logs received from external systems based on the configuration in the **syslog.conf** file, and write specified events to specific files.
- There are three roles in the Syslog system:
  - Sender: refers to the network element that generates Syslog messages.
  - Relay: refers to the network element or another device that forwards Syslog messages it receives.
  - Collector: refers to the Syslog server that does not forward Syslog messages it receives.

Others — LLDP

Supports network monitoring management.

- The Link Layer Discovery Protocol (LLDP) is a Layer 2 protocol defined in the IEEE 802.1ab standard. LLDP allows a device to send local management information such as its management IP address, device ID, and port ID to its neighbors, which then save the received information in their management information bases (MIBs), so that the network management system (NMS) can search required information in the MIBs to determine link status.



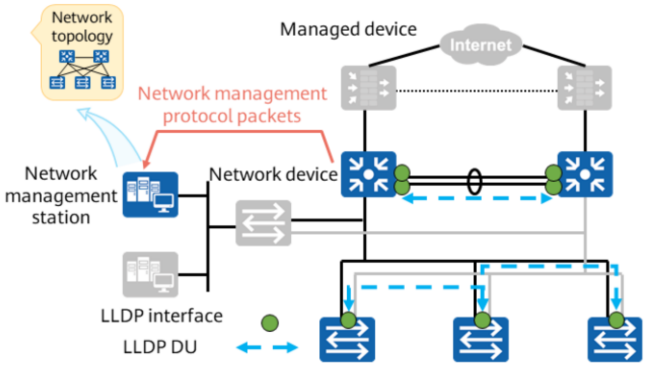


Page 28 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- LLDP is a neighbor discovery protocol. It defines a standard method for Ethernet network devices, such as switches, routers, and Wireless Local Area Network (WLAN) access points, to advertise their presence to neighboring devices and save discovery information about neighboring devices. Detailed device information, including device configurations and identification, can all be advertised using LLDP.
- LLDP data units (DUs) are transmitted periodically and reserved only for a certain period. IEEE has defined a recommended transmission interval of 30 seconds. After receiving an LLDP DU from a neighboring network device, an LLDP-enabled device stores the LLDP DU in an SNMP MIB defined by IEEE and keeps the LLDP DU valid within a certain period defined by the TTL carried in the LLDP DU.
- The protocol enables the NMS to accurately discover and simulate the physical network topology. LLDP-enabled devices transmit and receive advertisements, and they store the information advertised by their neighboring devices. The advertised information of a neighboring device includes its management address, device type, and port number, and this information helps determine the type of the neighboring device and the ports through which they connect to each other.
- Single-neighbor networking:
  - In single-neighbor networking mode, interfaces of two switches are directly connected and each interface has only one neighbor.
- Link aggregation networking:
  - In link aggregation networking, interfaces between switches are directly connected and bundled into a link aggregation group. Each interface in a link aggregation group has only one neighbor.

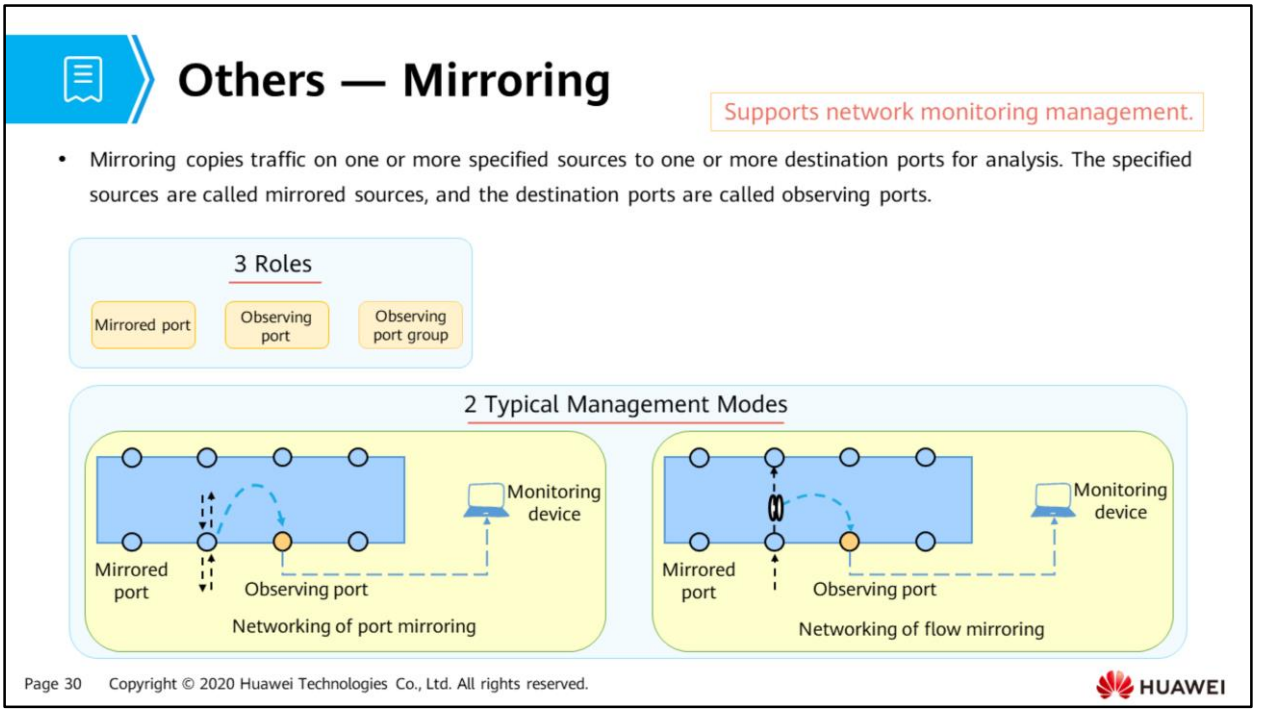
Application Scenarios of LLDP



- As shown in the figure on the left, switches are directly connected through a single link or an Eth-Trunk. The network management station is routable with the switches, and the network management protocol has been configured.
- The switches use LLDP to discover link-layer neighbor information and report the information to the NMS through a network management protocol, so that the NMS can display the network topology on its graphical user interface.

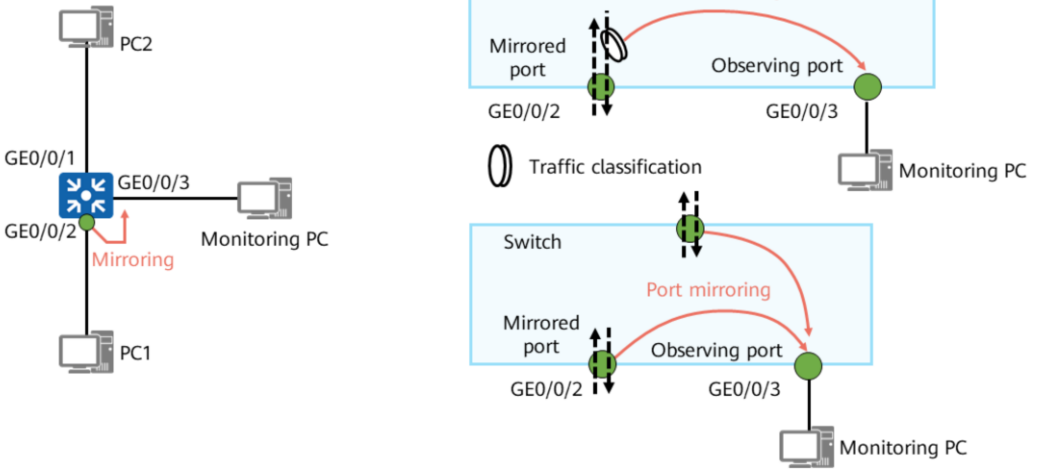
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Ethernet link aggregation, also called Eth-Trunk in short, bundles multiple physical links into a logical link to increase available link bandwidth.



- During network maintenance, you may need to obtain and analyze packets in some circumstances. For example, if you detect suspected attack packets, you need to obtain and analyze the packets without affecting packet forwarding. The mirroring function copies packets on a mirrored port to an observing port for analysis by a monitoring device, without affecting packet processing on the mirrored port. This function facilitates network monitoring and troubleshooting.
- Basic concepts:
  - A mirrored port is a monitored port, on which all the packets or packets matching traffic classification rules are copied to an observing port.
  - An observing port is connected to a monitoring device and transmits the packets copied from a mirrored port.
  - An observing port group is a group of ports connected to multiple monitoring devices. Packets mirrored to an observing port group are copied to all the member ports in the observing port group.
- Port mirroring: enables a device to copy the packets passing through a mirrored port and send them to a specified observing port for analysis and monitoring.
- Flow mirroring: enables a device to copy the specified packets passing through a mirrored port to an observing port for analysis and monitoring. In flow mirroring, a traffic policy containing the mirroring behavior is applied to a mirrored port. If the packets passing through the mirrored port match the traffic classification rule, they are copied to the observing port.

Application Scenarios of Mirroring



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

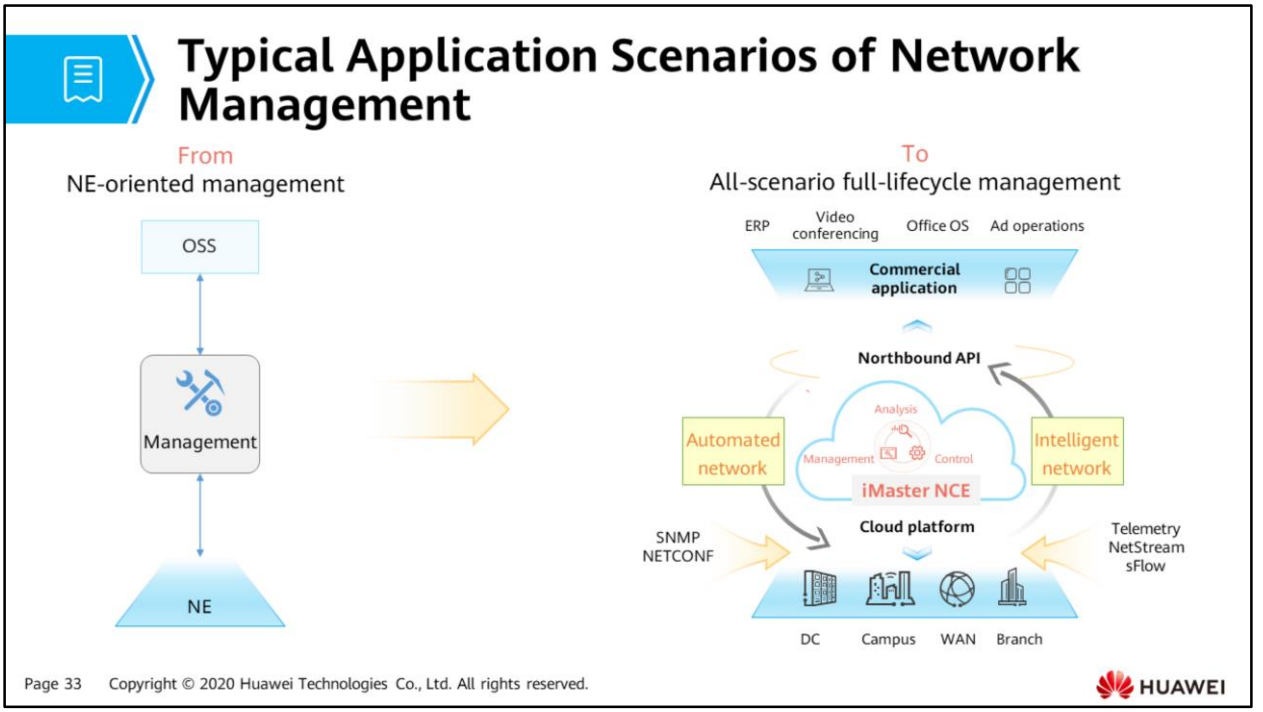
- In some scenarios, we may need to monitor incoming or outgoing packets on a specific interface of a switch or analyze specific traffic. For example, in the figure, the interface GE0/0/2 carries a large amount of traffic, and when a network fault occurs, we need to analyze the packets sent and received by this interface so as to locate the fault. To do so, we can connect a PC to interface GE0/0/3, install protocol analysis software on the PC, and deploy port mirroring to mirror incoming and outgoing traffic of GE0/0/2 to GE0/0/3. Then, all we need to do is using the protocol analysis software on the PC to view packets.
- It should be noted that without port mirroring, packets will not be sent to GE0/0/3 unless the destination of the packets is this interface. Therefore, port mirroring is by essence copying the traffic of a specific port to a monitoring port.

Contents

1. Development History of Network Management
2. Functions of Network Management
3. Categories of Network Management Protocols
4. **Application Scenarios of Network Management**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI



- With the rapid development of networks, digital transformation is gaining unprecedented importance, and network management is gradually shifting from NE-oriented management to scenario-oriented automation.
- The network management and control system is the evolution direction of autonomous networks and consists of the controller, analyzer, and manager.
- SNMP and NETCONF are used to deliver configurations, whereas Telemetry, NetStream, and sFlow are used to report data.

What Is iMaster NCE?

SDN-based automatic service configuration/deployment
AI-powered intelligent analysis/prediction/troubleshooting

Automated + Intelligent



Unified data base
Centralized detection/locating/processing

Manager + Controller + Analyzer



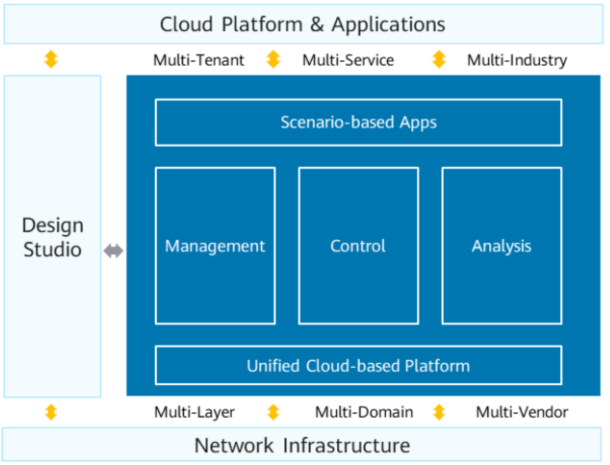
Full lifecycle management
Simulation/Verification/Monitoring/Optimization

Plan + Construct + Maintain + Optimize



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

iMaster NCE Product Architecture and Customer Benefits



Automated network
Intent-driven, model-based
Multi-domain, multi-layer, and multi-vendor

Intelligent network
Big data- and AI-powered predictive maintenance

Agile system integration
Programmable design studio
Open REST APIs

Cloud-based elastic expansion
Public and private cloud support
Elastic expansion & high reliability

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Quiz

1. (Multiple) What are the two key elements in the network management system? ( )
   A. Managing device
   B. Managed device
   C. Agent
   D. Network management protocol
2. (Multiple) Which of the following statements are true about basic network management functions?
   A. Configuration management
   B. Performance management
   C. Fault management
   D. Security management
   E. Accounting management

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

1. AB
2. ABCDE

Summary

- With the rapid development of digital technologies, technologies such as big data, AI, and cloud are intertwined with business development, promoting network management transformation. Specifically, network management will gradually develop towards automated management and intelligent O&M.
- Network management consists of four models: organization model, information model, communication model, and functional model.
- This part describes the network management protocols in terms of configuration, performance, and fault management. Each protocol is illustrated from three dimensions: role, data organization, and typical management mode.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

More Information

- RFC 6241 — Network Configuration Protocol
  - https://tools.ietf.org/html/rfc6241
- Network Telemetry Framework — IETF Tools
  - https://tools.ietf.org/id/draft-song-opsawg-ntf-02.html

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Recommendations

- https://e.huawei.com/en/talent/#/home

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Large-Scale WLAN Deployment

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- Currently, most enterprises utilize both wired and wireless networks to support office services. In office areas, in addition to the network ports used for wired connections, full Wi-Fi access is also available, making the office environment more open and intelligent. In the future, high-bandwidth services, including enterprise cloud desktop office, telepresence conference, and 4K video, will be migrated from wired to wireless networks. Likewise, new technologies such as VR/AR, virtual assistant, and automation factory, will be directly deployed on wireless networks. All these new application scenarios pose higher requirements on enterprise WLAN design and planning.
- This course describes typical applications, key technologies, and configurations for large-scale WLAN networking.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Cloud desktops, also known as desktop virtualization or cloud computers, are replacing traditional computers. With cloud desktops, users do not need to purchase computer hosts. After installing a client, users can access VM hosts on the backend server through a specific communication protocol to implement interactive operations, with the same experience as traditional computers. In addition, the cloud desktop mode is the latest mobile office solution that allows users to access the Internet with smart devices such as smartphones and tablets.
- A telepresence conference system has HD cameras and audio devices deployed to hold life-size face-to-face and eye-to-eye video conferences.
- Virtual reality (VR): uses a computer to simulate a 3D environment and enables users to interact by means of gloves and glasses.
- Augmented reality (AR): is also known as mixed reality (MR), a new technology developed on the basis of VR. Based on information provided by a computer system, AR enhances users' perception of the real world, applies virtual information to the real world, and adds virtual objects, scenarios, or system prompt information generated on the computer to the real scenario, thereby augmenting the real world. AR is typically implemented through a transparent head-mounted display system and a registration system (positioning for the user observation point and virtual objects generated by a computer in the AR system).

Objectives

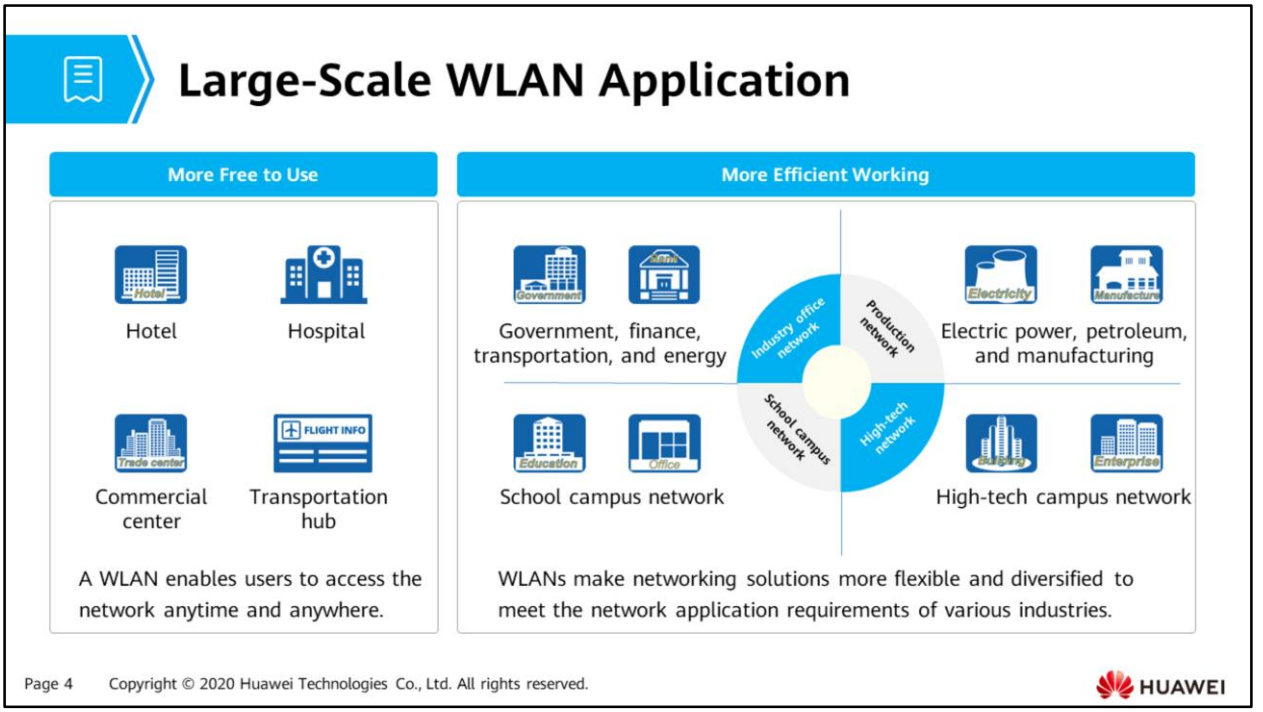
- Upon completion of this course, you will be able to:
  - Describe the typical network architecture and characteristics of a large-scale WLAN.
  - Describe key technologies for constructing a large-scale WLAN.
  - Configure large-scale WLAN networking.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Large-Scale WLAN Networking Overview**
2. VLAN Pool
3. DHCP
4. Roaming
5. Reliability
6. NAC

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- In high-tech parks, a large number of emerging technologies will be deployed, such as IoT, 5G convergence, and autonomous driving.

Large-Scale WLAN Characteristics



Large network scale

There are a large number of widely distributed devices with various models, resulting in high O&M costs.



A large number of widely distributed users

A large number of users are widely distributed, which pose high requirements on network experience.



High access security requirements

Guests, employees, and partners access the internal network irregularly. "One password for all" is no longer applicable to such a scenario.



High reliability requirements

WLAN ACs control the entire WLAN, and a fault may lead to great economic losses.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Huawei's Large-Scale WLAN Solution



Unified device management

Unified management of devices on the entire network, automatic configuration backup, and real-time alarm reporting greatly simplify network management



Roaming and free mobility

When a STA moves on the campus, seamless roaming ensures that the network access rights and experience of the STA are consistent as long as its identity remains unchanged.



Access and STA security protection

Access control and STA security protection ensure all-around security.

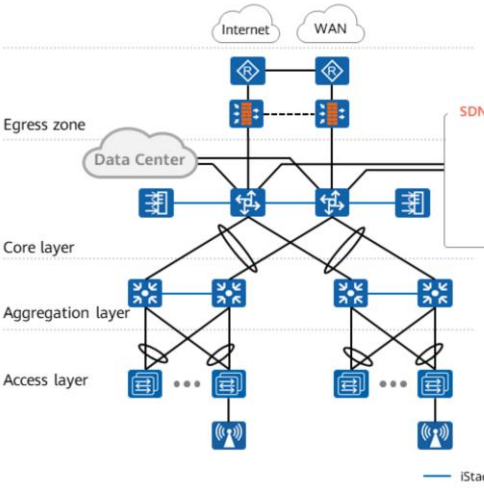


High reliability

Multiple reliability technologies, such as dual-link cold backup, dual-link hot standby (HSB) backup, and N+1 backup, ensure stable WLAN running.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

WLAN Networking Solution



An SDN controller can configure and manage the WLAN in a unified manner, achieving automatic service provisioning and full-lifecycle network management. Big data and AI technologies are also used to make the campus network intelligent, simplified, and secure. Such a solution enables the campus network with wired and wireless convergence.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Key Technologies on a Large-Scale WLAN

Technology	Function
VLAN pool	A VLAN pool can be used to assign access users to different VLANs, reducing the number of broadcast domains and broadcast packets on the network, and improving network performance.
DHCP options 43 and 52	When ACs and APs are connected through a Layer 3 network, the APs cannot discover an AC by sending broadcast request packets. In this case, the DHCP server needs to carry the Option 43 field (IPv4) or Option 52 field (IPv6) in the response packets to notify APs of the AC's IP address.
Roaming	WLAN roaming allows a STA to move between the coverage areas of different APs with nonstop service transmission.
Reliability	To ensure stable running of WLAN services, reliability technologies enable services to be smoothly switched to a backup device upon a failure of the master device.
NAC	Network Access Control (NAC) technology is an end-to-end security technology that ensures network security by authenticating access clients and users.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

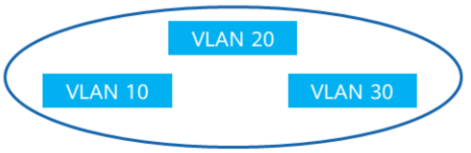
Contents

1. Large-Scale WLAN Networking Overview
2. **VLAN Pool**
3. DHCP
4. Roaming
5. Reliability
6. NAC

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

What Is a VLAN Pool?

- Challenges to WLANs
  - Mobility of WLAN STAs may lead to a large number of IP address requests in specific areas.
  - In most cases, an SSID maps to only one service VLAN. If IP addresses are increased by expanding the subnet range, the broadcast domain will be expanded, causing network congestion upon transmission for a large number of broadcast packets.
- VLAN pooling technology adds VLAN resources to a VLAN pool and provides VLAN assignment algorithms.



VLAN pool

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- A VLAN pool can be used to assign access users to different VLANs, reducing the number of broadcast domains and broadcast packets on the network, and improving network performance.
- Because of high mobility, a large number of users may access a WLAN from an area and then roam to another area. As a result, the number of users in the area becomes large, therefore requiring a large number of IP addresses, such as the entrance of a stadium or the lobby of a hotel. Currently, an SSID maps to only one service VLAN that covers only one subnet. If a large number of users access the network from an area, they can obtain IP addresses only by expanding the subnet range. However, this expands the broadcast domain, causing severe network congestion brought by a large number of broadcast packets (such as ARP and DHCP).
- To solve this problem, one SSID needs to map multiple VLANs so that STAs are distributed to different VLANs to reduce the broadcast domain. The VLAN pool can manage and allocate multiple VLANs, therefore achieving mapping from one SSID to multiple VLANs.

VLAN Assignment Algorithms

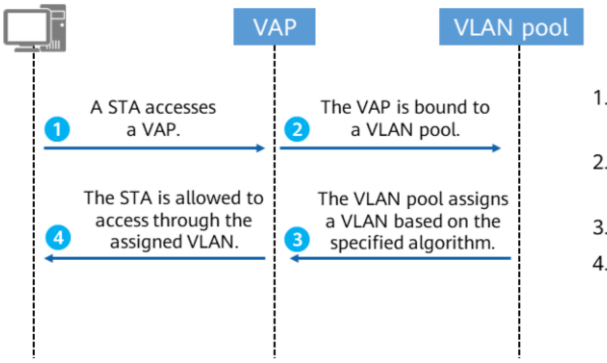
- Even assignment algorithm: assigns STAs to different VLANs according to the order in which STAs go online.
- Hash assignment algorithm: assigns STAs to VLANs based on the hash result of their MAC addresses.
- Comparison between the two VLAN assignment algorithms

Assignment Algorithm	Advantage	Disadvantage
Even	The number of STAs in each VLAN is even.	VLANs and IP addresses may easily change when STAs go online again.
Hash	VLANs and IP addresses remain unchanged for STAs who go online for multiple times.	The number of STAs in each VLAN is uneven.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- Even assignment algorithm: assigns STAs to different VLANs according to the order in which STAs go online. When STAs go offline and online again, their VLANs and IP addresses may easily change.
- Hash assignment algorithm: assigns STAs to VLANs based on the hash result of their MAC addresses. VLANs and IP addresses remain unchanged for STAs. However, the number of users in each VLAN is uneven.

VLAN Assignment Process



1. When a STA accesses the network through a VAP, the STA checks whether the VAP has a VLAN pool.
2. If the VAP profile is bound to a VLAN pool, the even or hash assignment algorithm of the VLAN pool is used.
3. A VLAN is assigned to the STA.
4. The STA goes online from the assigned VLAN.

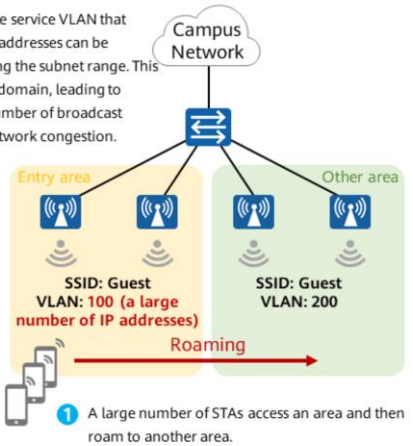
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Virtual access point (VAP): A physical AP can be virtualized into multiple VAPs, each of which provides the same functions as the physical AP. You can create different VAPs on an AP to provide the wireless access service for different user groups.

VLAN Pool Application Example

Problem: Entry Effect

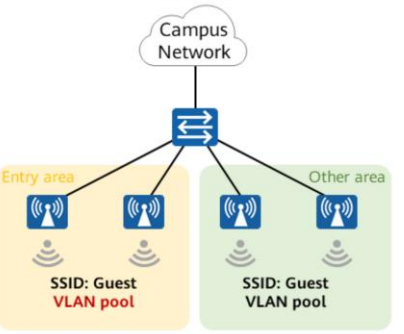
3 If an SSID maps to only one service VLAN that covers only one subnet, IP addresses can be increased only by expanding the subnet range. This will expand the broadcast domain, leading to transmission for a large number of broadcast packets that may cause network congestion.



2 A large number of STAs access the network in this area, therefore requiring a large number of IP addresses.

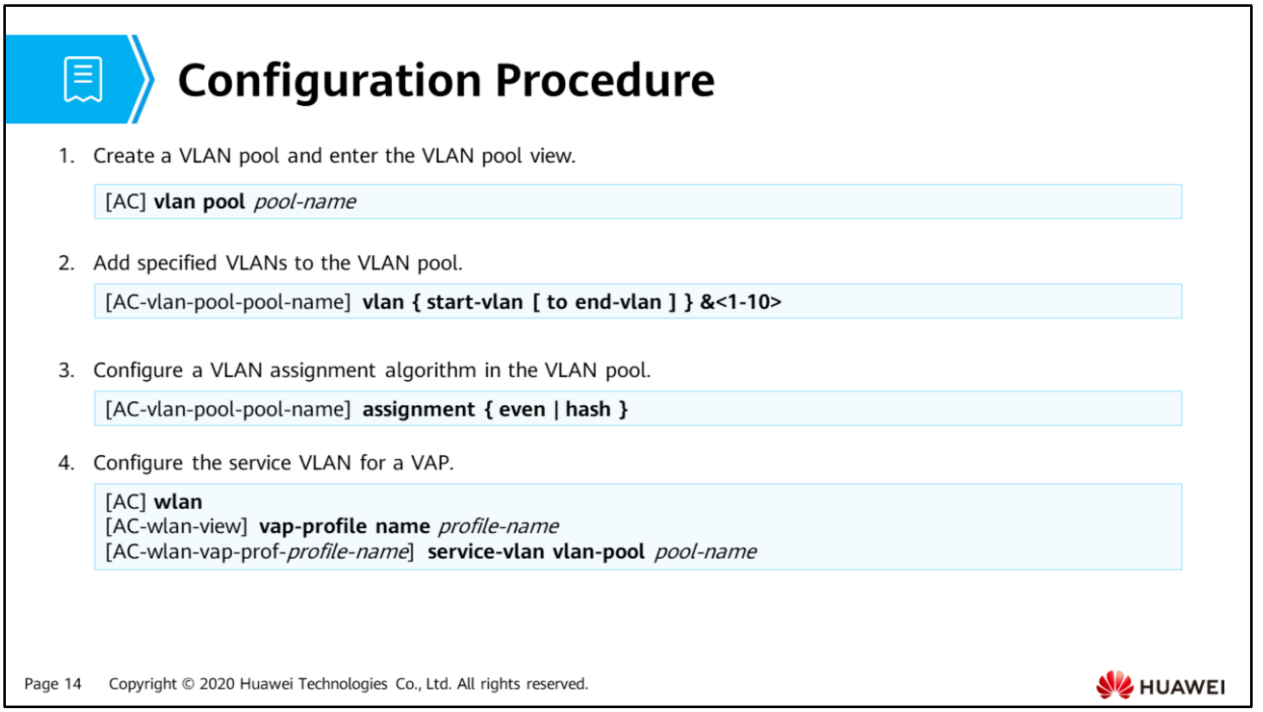
1 A large number of STAs access an area and then roam to another area.

Solution



- In this scenario, the VLAN pool can be used as the service VLAN.
- A VLAN pool provides the VLAN management and assignment algorithms. In this way, an SSID can map to multiple VLANs so that a large number of STAs can be distributed to different VLANs, reducing the number of broadcast domains.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- The VLAN assignment algorithm can be configured for a VLAN pool.
- assignment { even | hash }
  - When the VLAN assignment algorithm is set to even, service VLANs are assigned to STAs from the VLAN pool based on the order in which STAs go online. Address pools mapping the service VLANs evenly assign IP addresses to STAs. If a STA goes online many times, it obtains different IP addresses.
  - When the VLAN assignment algorithm is set to hash, VLANs are assigned to STAs from the VLAN pool based on the harsh result of their MAC addresses. As long as the VLANs in the VLAN pool do not change, the STAs obtain fixed service VLANs. A STA is preferentially assigned the same IP address when going online at different times.

Configuration Example (1/2)



- The AC functions as the DHCP server for STAs, which has the DHCP function enabled.
- The DHCP server supports two network segments: 10.1.2.0/24 and 10.1.3.0/24.
- DHCP clients can dynamically obtain IP addresses on the 10.1.2.0 and 10.1.3.0 network segments from the DHCP server, with the gateway addresses of 10.1.2.254 and 10.1.3.254, respectively.

VLAN pool configuration on the AC

```
[AC] vlan pool STA
[AC-vlan-pool-STA] vlan 20 30
[AC-vlan-pool-STA] assignment hash
[AC-vlan-pool-STA] quit
```

```
[AC] wlan
[AC-wlan-view] vap-profile name huawei
[AC-wlan-vap-prof-huawei] service-vlan vlan-pool STA
Info: This operation may take a few seconds, please wait. Done.
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (2/2)

1. Check brief configuration information about all VLAN pools on the AC.

```
<AC> display vlan pool all
--------------------------------------------------------------------------------
Name                    Assignment              VLAN total
--------------------------------------------------------------------------------
STA                     hash                    2
--------------------------------------------------------------------------------
Total: 2
```

2. Check detailed configuration about VLAN pool **STA** on the AC.

```
<AC> display vlan pool name STA
--------------------------------------------------------------------------------
Name                    : STA
Total                   : 2
Assignment              : hash
VLAN ID                 : 20 30
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

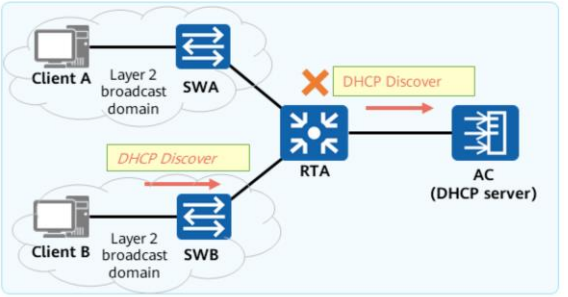
1. Large-Scale WLAN Networking Overview
2. VLAN Pool
3. **DHCP**
4. Roaming
5. Reliability
6. NAC

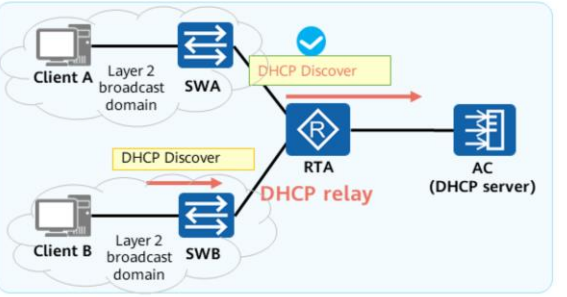
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

DHCP Relay

- A DHCP client broadcasts IP packets to search for DHCP servers on the local network segment. When a DHCP client and server are located in different network segments, the router between them does not forward such broadcast IP packets.
- A DHCP relay agent is introduced to transparently transmit DHCP packets across network segments, allowing a DHCP server to serve clients on multiple network segments at the same time.





Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- Along with the expanding network scale, an increasing number of network devices are deployed on the network. However, users in an enterprise may be distributed on different network segments. In normal cases, one DHCP server cannot meet such IP address allocation requirements. In most cases, DHCP clients on the network segments of the enterprise are not in the same Layer 2 broadcast domain as the DHCP server. To obtain IP addresses from the DHCP server, DHCP clients have to transmit DHCP packets across network segments.

Configuration Procedure

1. Enable the DHCP relay function on an interface.

   [Huawei-GigabitEthernet0/0/0]**dhcp select relay**

2. Specify the IP address of the DHCP server in the interface view.

   [Huawei-GigabitEthernet0/0/0]**dhcp relay server-ip** *ip-address*

3. Create a DHCP server group.

   [Huawei]**dhcp server group** *group-name*

4. Add DHCP servers to the DHCP server group.

   [Huawei-dhcp-server-group-HW]**dhcp-server** *ip-address* [*ip-address-index*]

5. Apply the DHCP server group to the interface.

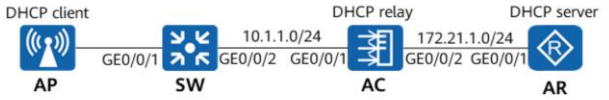
   [Huawei-GigabitEthernet0/0/0]**dhcp relay server-select** *group-name*

6. Enable the DHCP client function on the interface.

   [Huawei-GigabitEthernet0/0/0]**ip address dhcp-alloc**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (1/4)



- Configure VLAN 10 as the management VLAN, and enable the AP to obtain an IP address through DHCP.
- Configure basic interworking parameters on the switch, AC, and AR.
- Configure the AP, AC, and AR as the DHCP client, DHCP relay agent, and DHCP server, respectively, and enable the DHCP function.
- Enable the DHCP relay function on the AC and specify the IP address of the DHCP server as 172.21.1.2.
- Create an IP address pool named **AP** on the AR, and set the IP segment to 10.1.1.0/24 and the gateway address to 10.1.1.2.

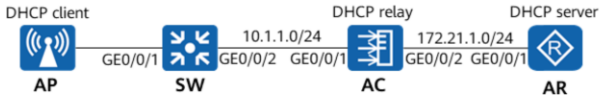
Configurations on the switch and AC

```
[SW] vlan 10
[SW-vlan10] quit
[SW] interface GigabitEthernet 0/0/1
[SW-GigabitEthernet0/0/1] port link-type access
[SW-GigabitEthernet0/0/1] port default vlan 10
[SW-GigabitEthernet0/0/1] quit
[SW] interface GigabitEthernet 0/0/2
[SW-GigabitEthernet0/0/2] port link-type trunk
[SW-GigabitEthernet0/0/2] port trunk allow-pass vlan 10
[SW-GigabitEthernet0/0/2] quit
[SW] interface Vlanif 10
[SW-Vlanif10] ip address 10.1.1.1 24
```

```
[AC] vlan batch 10 20
[AC] interface GigabitEthernet 0/0/1
[AC-GigabitEthernet0/0/1] port link-type trunk
[AC-GigabitEthernet0/0/1] port trunk allow-pass vlan 10
[AC-GigabitEthernet0/0/1] quit
[AC] interface GigabitEthernet 0/0/2
[AC-GigabitEthernet0/0/1] port link-type access
[AC-GigabitEthernet0/0/1] port default vlan 20
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (2/4)



- Configure VLAN 10 as the management VLAN, and enable the AP to obtain an IP address through DHCP.
- Configure basic interworking parameters on the switch, AC, and AR.
- Configure the AP, AC, and AR as the DHCP client, DHCP relay agent, and DHCP server, respectively, and enable the DHCP function.
- Enable the DHCP relay function on the AC and specify the IP address of the DHCP server as 172.21.1.2.
- Create an IP address pool named **AP** on the AR, and set the IP segment to 10.1.1.0/24 and the gateway address to 10.1.1.2.

Configurations on the AC and AR

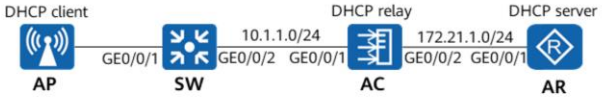
```
[AC] interface Vlanif 10
[AC-Vlanif10] ip address 10.1.1.2 24
[AC-Vlanif10] quit
[AC] interface Vlanif 20
[AC-Vlanif20] ip address 172.21.1.1 24
[AC-Vlanif20] quit
```

```
[AR] interface GigabitEthernet 0/0/1
[AR-GigabitEthernet0/0/1] ip address 172.21.1.2 24
[AR-GigabitEthernet0/0/1] quit
```

```
[AC] dhcp server group AP
[AC-dhcp-server-group-AP] dhcp-server 172.21.1.2
[AC-dhcp-server-group-AP] quit
[AC] interface Vlanif 10
[AC-Vlanif10] dhcp select relay
[AC-Vlanif10] dhcp relay server-select AP
[AC-Vlanif10] quit
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (3/4)



- Configure VLAN 10 as the management VLAN, and enable the AP to obtain an IP address through DHCP.
- Configure basic interworking parameters on the switch, AC, and AR.
- Configure the AP, AC, and AR as the DHCP client, DHCP relay agent, and DHCP server, respectively, and enable the DHCP function.
- Enable the DHCP relay function on the AC and specify the IP address of the DHCP server as 172.21.1.2.
- Create an IP address pool named **AP** on the AR, set the IP segment to 10.1.1.0/24 and the gateway address to 10.1.1.2, and add a static route so that the AR can access the 10.1.1.0 network segment.

Configuration on the AR

```
[AR] ip pool AP
[AR-ip-pool-AP] network 10.1.1.0 mask 24
[AR-ip-pool-AP] gateway-list 10.1.1.2
[AR-ip-pool-AP] excluded-ip-address 10.1.1.1
[AR-ip-pool-AP] quit
[AR] interface GigabitEthernet 0/0/1
[AR-GigabitEthernet0/0/1] dhcp select global
[AR-GigabitEthernet0/0/1] quit
[AR] ip route-static 10.1.1.0 255.255.255.0 172.21.1.1
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (4/4)

- Check the IP address allocation of the DHCP pool on the AR.

```
[AR] display ip pool name AP used
...
Network section :
 ------------------------------------------------------------------------------
 Index    IP              MAC                Lease   Status
 ------------------------------------------------------------------------------
   253    10.1.1.254      00e0-fcca-1150     2181    Used
 ------------------------------------------------------------------------------
[AR]
```

The command output shows that the DHCP server has allocated an IP address to the AP.

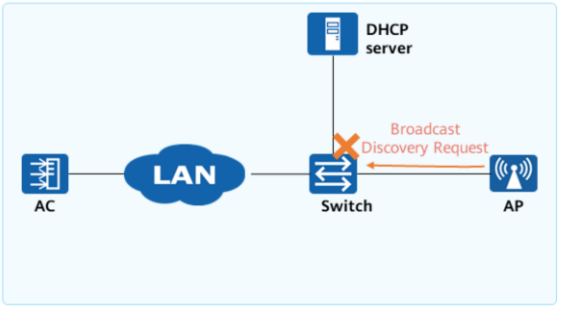
- Check DHCP relay information on the AC.

```
<AC> display dhcp relay all
 DHCP relay agent running information of interface Vlanif10 :
 Server group name       : AP
 Gateway address in use  : 10.1.1.2
```

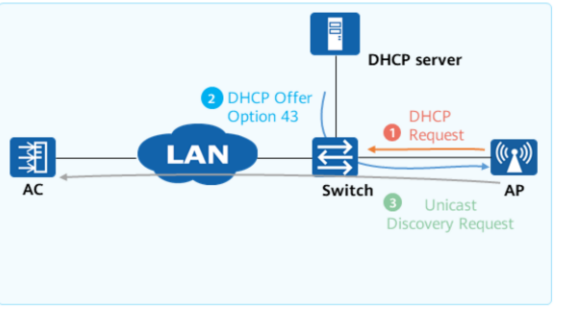

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

AC Discovery Mechanism in the WLAN Layer 3 Networking

When ACs and APs are connected through a Layer 3 network, the APs cannot discover an AC by sending broadcast DHCP Request packets. In this case, the DHCP server needs to carry the Option 43 field (IPv4) or Option 52 field (IPv6) in the response packets to notify APs of the AC's IP address.



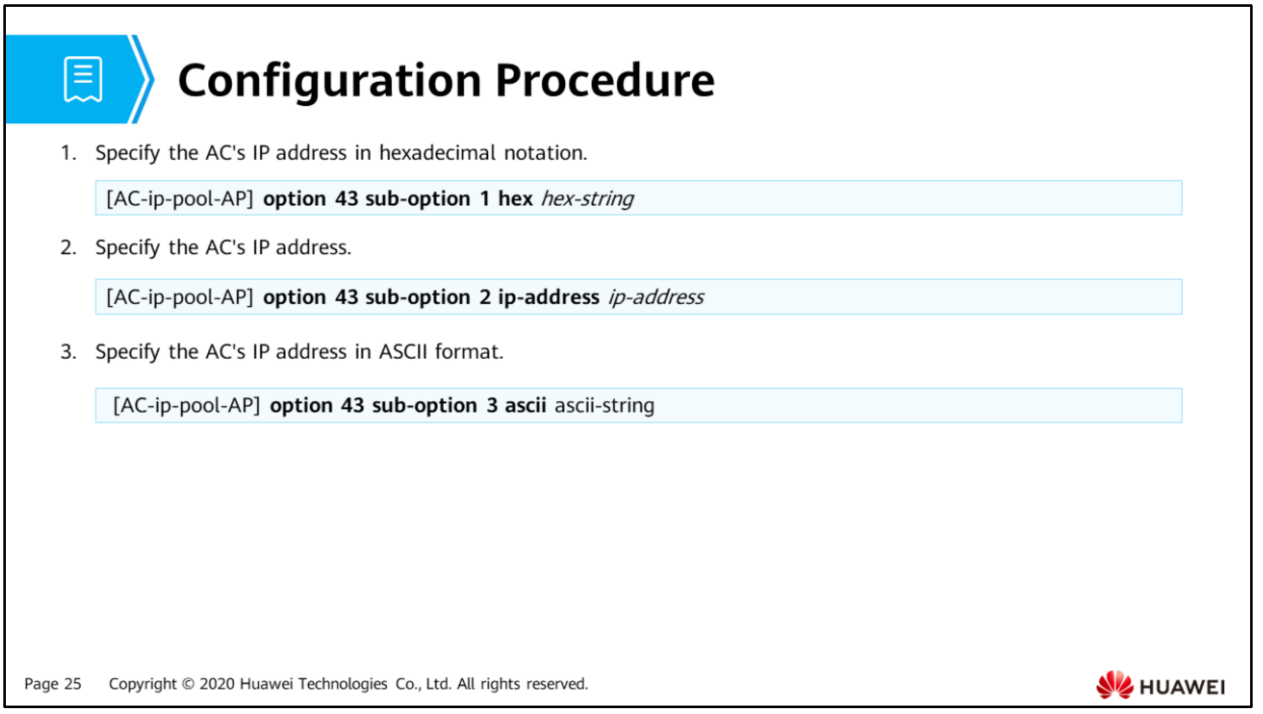
In Layer 3 WLAN networking, an AP cannot discover an AC by broadcasting Discovery Request packets, leading to a failure to establish a CAPWAP tunnel.



In the WLAN Layer 3 networking, after DHCP Option 43 is configured, the AP can obtain the AC's IP address while obtaining its own IP address, and then establish a unicast connection with the AC.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- If ACs and APs are connected through a Layer 2 network, you can configure Option 43 to carry the IP address of a specified AC in the unicast request packets sent from an AP. If the AP does not receive any response after sending the unicast packet for 10 times, the AP will attempt to discover ACs in the same network segment in broadcast mode. Therefore, Option 43 is optional in Layer 2 networking but mandatory in Layer 3 networking.
- The Type value of Option 43 is 43 (0x2B). Option 43 is the vendor-specific information option, through which a DHCP server and clients exchange vendor information. When a DHCP server receives a DHCP Request message asking for Option 43, it encapsulates Option 43 in the DHCP Response message and sends it to the DHCP client. (In this course, Option 43 contains the AC's IP address.)



- In this example, a switch or an AC functions as the DHCP server. A large-scale network typically deploys an independent DHCP server. In practice, however, we can also deploy a switch or AC as a DHCP server. Then, run one of the following commands to configure Option 43:
  - option 43 sub-option 1 hex C0A80001C0A80002: configures the device to specify AC IP addresses 192.168.0.1 and 192.168.0.2 in hexadecimal notation for APs. In the command, **C0A80001** indicates the hexadecimal format of 192.168.0.1, and **C0A80002** indicates the hexadecimal format of 192.168.0.2.
  - option 43 sub-option 2 ip-address 192.168.0.1 192.168.0.2: configures the device to specify AC IP addresses 192.168.0.1 and 192.168.0.2 for APs.
  - option 43 sub-option 3 ascii 192.168.0.1,192.168.0.2: configures the device to specify AC IP addresses 192.168.0.1 and 192.168.0.2 in ASCII format for APs, with multiple IP addresses separated by commas (,).

Configuration Example (1/2)



- Configure VLAN 10 as the management VLAN, and enable the AP to obtain an IP address through DHCP.
- The basic configurations of the switch, AC, and AR, and the DHCP relay configuration have been completed. The AP can obtain the IP address 10.1.1.254 and the AC's IP address 100.100.100.100.
- Create an IP address pool named **AP** on the AR, set the IP segment to 10.1.1.0/24 and the gateway address to 10.1.1.2, and add a static route so that the AR can access the 10.1.1.0 network segment.

Configurations on the AR and AC

```
[AR] ip pool AP
[AR-ip-pool-ap] option 43 sub-option 3 ascii 100.100.100.100
[AR-ip-pool-ap] quit
```

```
[AC] interface LoopBack 0
[AC-LoopBack0] ip address 100.100.100.100 32
[AC-LoopBack0] quit
[AC] capwap source interface LoopBack 0
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (2/2)

- Check the configuration of the DHCP address pool on the AR.

```
[AR] display ip pool name AP
Pool-name         : AP
  Pool-No         : 0
  Lease           : 1 Days 0 Hours 0 Minutes
  Option-code     : 43
  Option-subcode  : 3
  Option-type     : ascii
  Option-value    : 100.100.100.100
...
  Position        : Local        Status        : Unlocked
  Gateway-0       : 10.1.1.2
  Mask            : 255.255.255.0
...
```

The command output shows that the option field has been configured successfully.

- Check whether the AP can discover the AC.

```
[AC] display ap unauthorized record
Unauthorized AP record:
Total number: 1
------------------------------------------------------------------------------
AP type: AP4030TN
AP SN: 210235448310C92A877C
AP MAC address: 00e0-fcca-1150
AP IP address: 10.1.1.254
Record time: 2020-06-18 11:51:34
------------------------------------------------------------------------------
[AC]
```

The command output shows that the AP has successfully discovered the AC. You can enable the AP to join the AC at any time.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

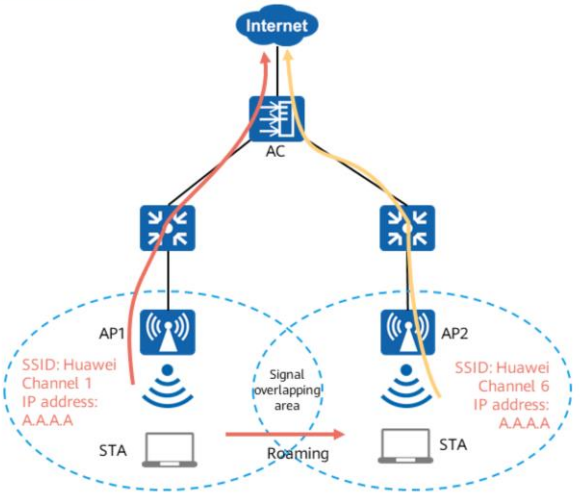
Contents

1. Large-Scale WLAN Networking Overview
2. VLAN Pool
3. DHCP
4. **Roaming**
5. Reliability
6. NAC

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

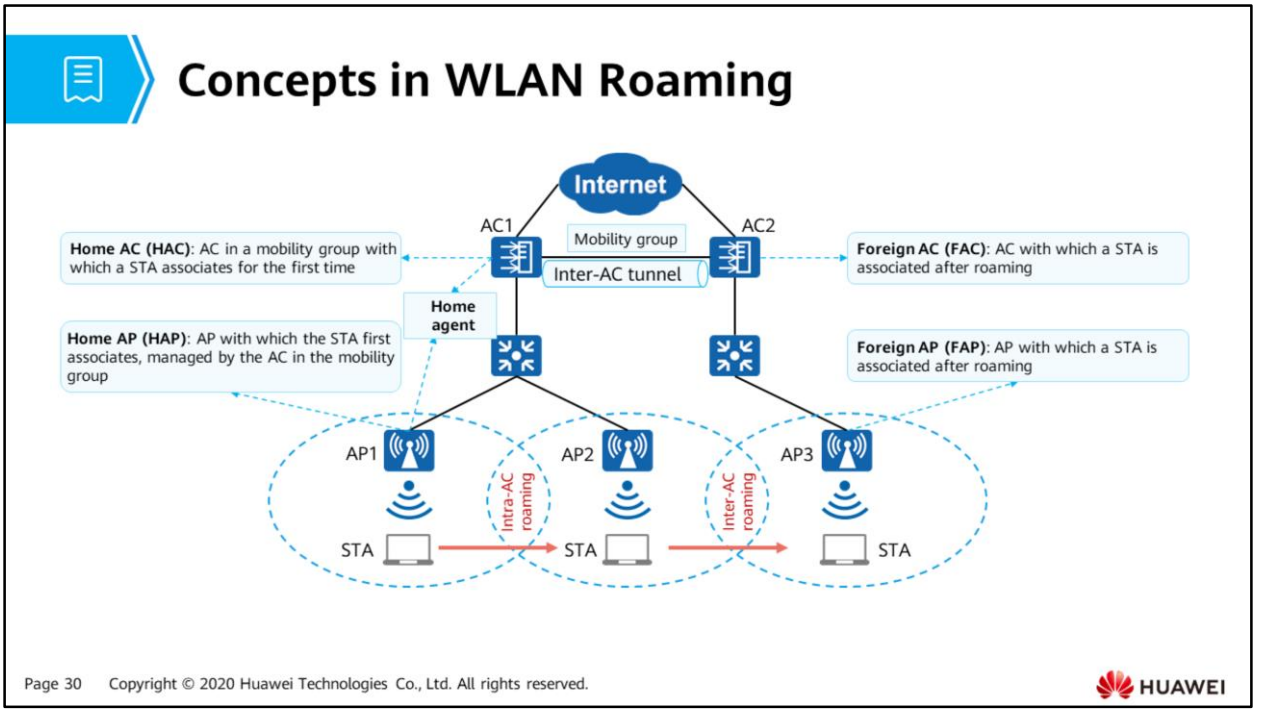
WLAN Roaming Overview



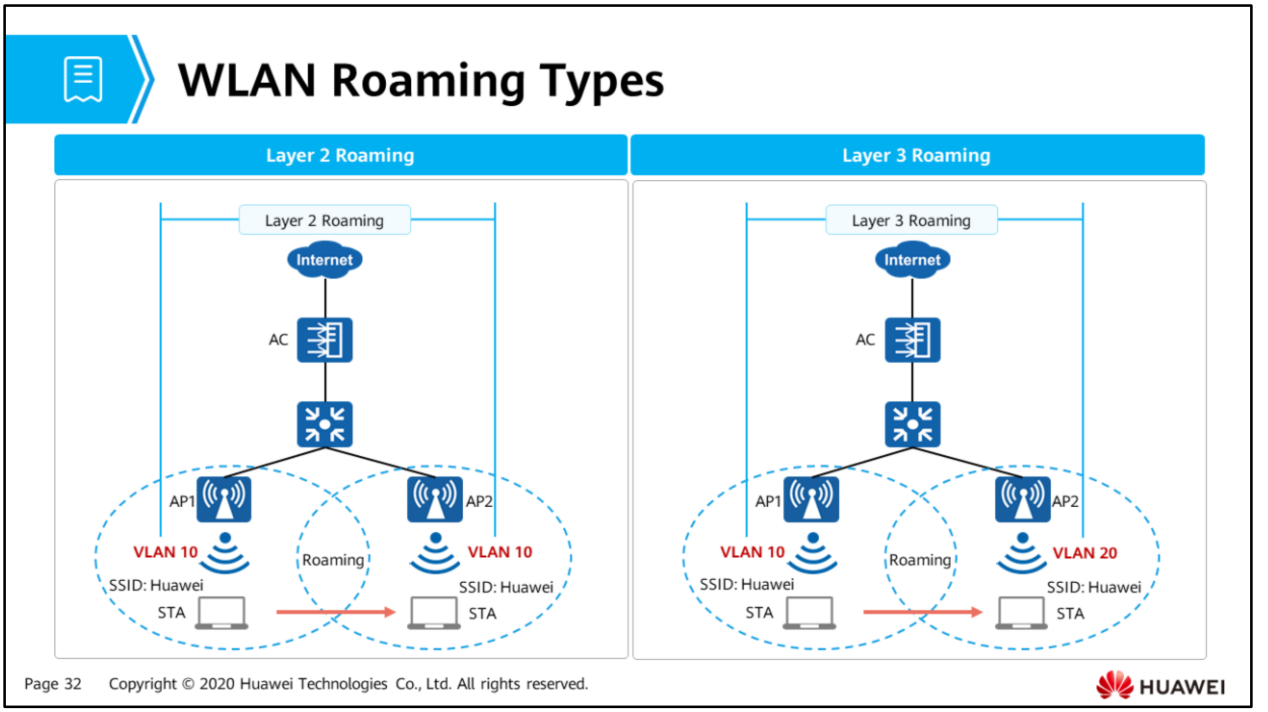
- WLAN roaming allows a STA to move between the coverage areas of different APs with nonstop service transmission.
- The APs involved in WLAN roaming must have the same SSID, same configurations in the security profile (different profile names allowed), and the same authentication mode and parameter settings in the authentication profile.
- WLAN roaming aims to achieve the following goals:
  - Avoid packet loss or service interruption caused by a long authentication duration.
  - Ensure that user authorization information does not change.
  - Ensure that the STA's IP address does not change.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- When a STA moves away from an AP, the link signal quality decreases gradually. If the signal quality falls below the roaming threshold, the STA proactively roams to a nearby AP to achieve better signal quality.
- As shown in the figure, roaming is completed through the following steps:
  - The STA has set up a link with AP1 and sends Probe Request packets on various channels. After AP2 receives the Probe Request frame on channel 6, it sends a Probe Response frame to the STA on channel 6. After receiving the Probe Response frame, the STA determines to associate with AP2.
  - The STA sends an Association Request frame to AP2 over channel 6, AP2 replies with an Association Response to the STA, and association between the STA and AP2. During the entire process, the association relationship between the STA and AP1 is maintained.
  - The STA is disassociated from AP1. The STA sends a Disassociation frame to AP1 over channel 1 (channel used by AP1).



- Intra-AC roaming: A STA is associated with the same AC during roaming.
- Inter-AC roaming: A STA is associated with different ACs during roaming.
- Inter-AC tunnel: To support inter-AC roaming, ACs in a mobility group need to synchronize STA and AP information with each other. Therefore, the ACs set up a CAPWAP tunnel to synchronize data and forward packets. As shown in the figure, AC1 and AC2 set up a CAPWAP tunnel for data synchronization and packet forwarding.
- Mobility server
    - When a STA roams between ACs, an AC is selected as the mobility server to maintain the membership table of the mobility group and deliver member information to ACs in the group. In this way, ACs in the same mobility group can identify each other and set up inter-AC tunnels.
    - A mobility server can be an AC outside or inside a mobility group.
    - An AC can function as the mobility server of multiple mobility groups, and can be added only to one mobility group.
    - A mobility server managing other ACs in a mobility group cannot be managed by another mobility server. That is, if an AC functions as a mobility server to synchronize roaming configurations to other ACs, it cannot be managed by another mobility server or synchronize roaming configurations from other ACs. (An AC with a mobility group configured cannot be configured as a mobility server.)



- Layer 2 roaming: A STA switches between two APs (or multiple APs) that are bound to the same SSID and have the same service VLAN (within the same IP address segment). During roaming, the access attributes (such as the service VLAN and obtained IP address) of the STA do not change. During roaming, packet loss and reconnection do not occur.
- Layer 3 roaming: The service VLANs of the SSIDs are different, and APs provide different Layer 3 service networks with different gateways before and after roaming. In this case, to ensure that the IP address of a roaming STA remains unchanged, the STA's traffic needs to be sent back to the AP on the initial access network segment to implement inter-VLAN roaming.
- Sometimes, two subnets may have the same service VLAN ID but belong to different subnets. Based on the VLAN ID, the system may incorrectly consider that STAs roam between the two subnets at Layer 2. To prevent this situation, configure a roaming domain to determine whether the STAs roam within the same subnet. The system determines Layer 2 roaming only when STAs roam within the same VLAN and same roaming domain; otherwise, the system determines Layer 3 roaming.

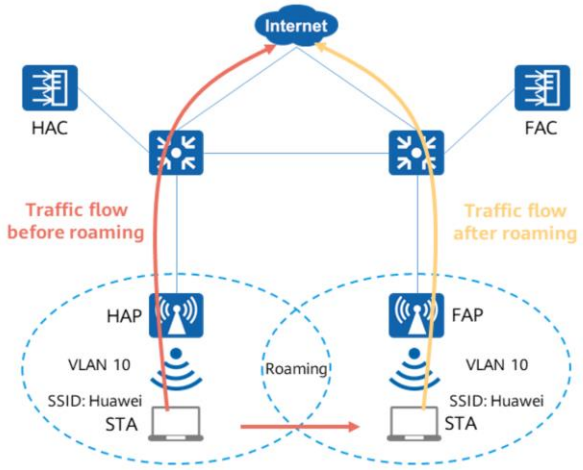
Traffic Forwarding Models in WLAN Roaming

- Depending on the WLAN data forwarding type and whether data is forwarded across Layer 3, four traffic forwarding models in WLAN roaming are classified, as described in the following table.

Forwarding Model	Characteristics
Direct forwarding in Layer 2 roaming	STAs stay on the same subnet before and after Layer 2 roaming. Similar to packet forwarding for new STAs, the FAP or FAC forwards packets of Layer 2 roaming STAs on the local network but does not send the packets back to the home agent over a tunnel.
Tunnel forwarding in Layer 2 roaming	
Direct forwarding in Layer 3 roaming	Service packets between the HAP and HAC are not encapsulated with the CAPWAP header. Therefore, whether the HAP and HAC reside on the same subnet cannot be determined. In this case, packets are forwarded back to the HAP by default.
Tunnel forwarding in Layer 3 roaming	Service packets between the HAP and HAC are encapsulated with the CAPWAP header. In this case, the HAP and HAC can be considered on the same subnet. Instead of forwarding the packets back to the HAP, the HAC directly forwards the packets to the upper-layer network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Inter-AC Layer 2 Roaming — Direct Forwarding

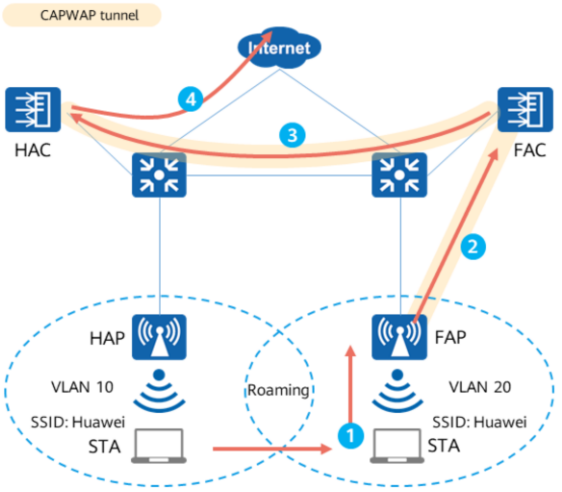


- Before roaming:
  - The STA sends service packets to the HAP.
  - After receiving the service packets, the HAP forwards them to the upper-layer network through the gateway (switch).
- After roaming:
  - The STA sends service packets to the FAP.
  - After receiving the service packets, the FAP forwards them to the upper-layer network through the gateway (switch).

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- The traffic flow for inter-AC Layer 2 roaming is the same for tunnel and direct forwarding modes, and is not mentioned here.

Inter-AC Layer 3 Roaming — Tunnel Forwarding

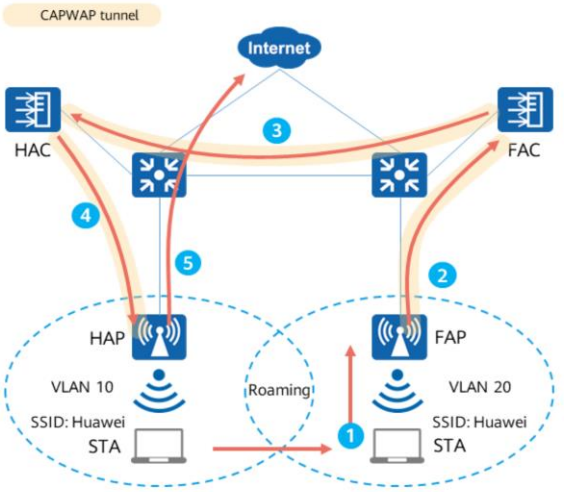


- Before roaming:
  1. The STA sends service packets to the HAP.
  2. After receiving the service packets, the HAP sends them to the HAC through a CAPWAP tunnel.
  3. The HAC forwards the service packets to the upper-layer network through the switch.
- After roaming:
  1. The STA sends service packets to the FAP.
  2. After receiving the service packets, the FAP sends them to the FAC through a CAPWAP tunnel.
  3. The FAC forwards the service packets to the HAC through a CAPWAP tunnel between them.
  4. The HAC forwards the service packets to the upper-layer network through the switch.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- STAs stay in different subnets before and after Layer 3 roaming. To enable the STAs to access the original network after roaming, ensure that user traffic is forwarded to the original subnet over CAPWAP tunnels.
- In tunnel forwarding mode, service packets between the HAP and HAC are encapsulated with the CAPWAP header. In this case, the HAP and HAC can be considered on the same subnet. Instead of forwarding the packets back to the HAP, the HAC directly forwards the packets to the upper-layer network.

Inter-AC Layer 3 Roaming — Direct Forwarding (HAP as the Home Agent)

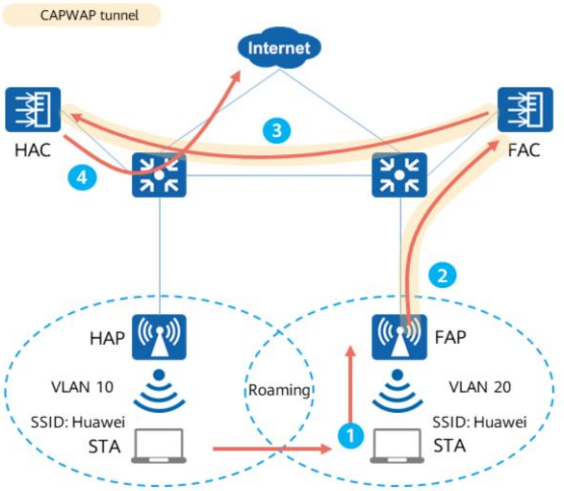


- Before roaming:
  1. The STA sends service packets to the HAP.
  2. After receiving the service packets, the HAP sends them to the upper-layer network through the switch.
- After roaming:
  1. The STA sends service packets to the FAP.
  2. After receiving the service packets, the FAP sends them to the FAC through a CAPWAP tunnel.
  3. The FAC forwards the service packets to the HAC through a CAPWAP tunnel between them.
  4. The HAC sends the service packets to the HAP through a CAPWAP tunnel.
  5. The HAP forwards the service packets to the upper-layer network.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- In direct forwarding mode, service packets between the HAP and HAC are not encapsulated with the CAPWAP header. Therefore, whether the HAP and HAC reside on the same subnet cannot be determined. In this case, packets are forwarded back to the HAP by default. If the HAP and HAC reside on the same subnet, you can configure a higher-performance HAC as the home agent. This reduces the load on the HAP and improves the forwarding efficiency.

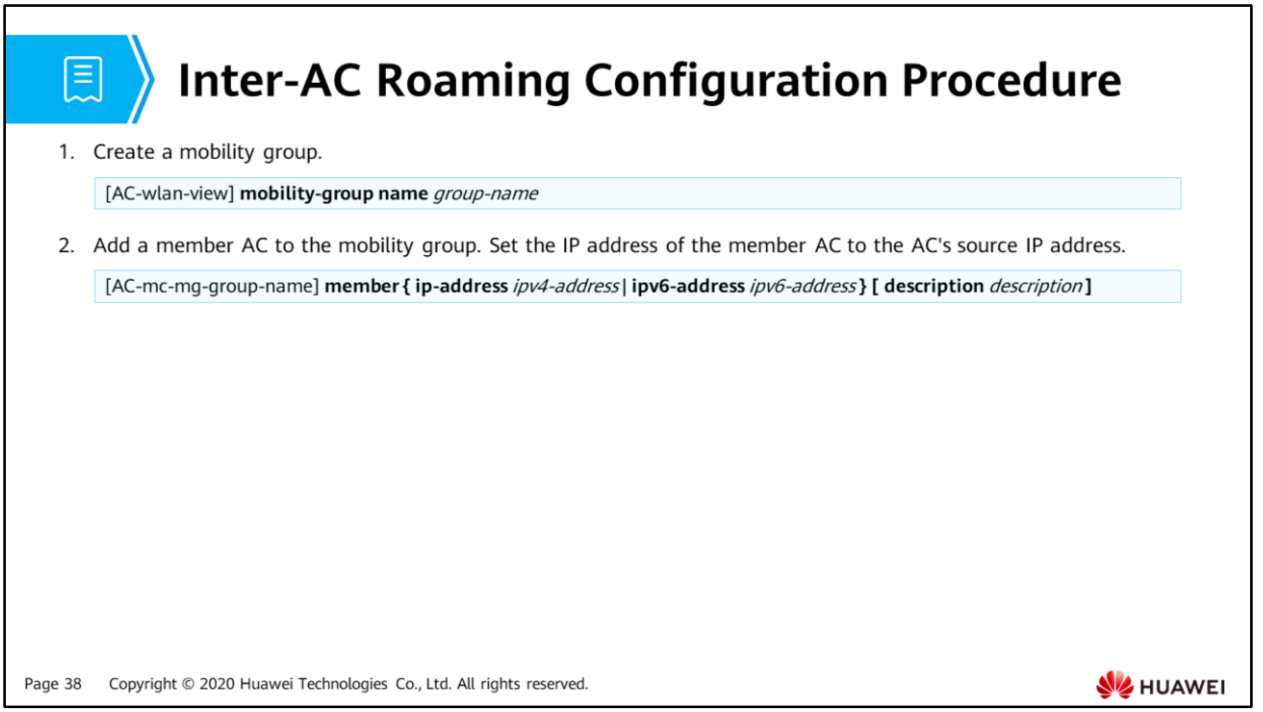
Inter-AC Layer 3 Roaming — Direct Forwarding (HAC as the Home Agent)



- Before roaming:
  1. The STA sends service packets to the HAP.
  2. After receiving the service packets, the HAP sends them to the upper-layer network through the switch.
- After roaming:
  1. The STA sends service packets to the FAP.
  2. After receiving the service packets, the FAP sends them to the FAC through a CAPWAP tunnel.
  3. The FAC forwards the service packets to the HAC through a CAPWAP tunnel between them.
  4. The HAC forwards the service packets to the upper-layer network.

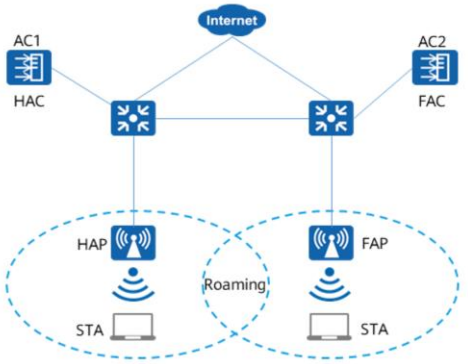
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- In direct forwarding mode, service packets between the HAP and HAC are not encapsulated with the CAPWAP header. Therefore, whether the HAP and HAC reside on the same subnet cannot be determined. In this case, packets are forwarded back to the HAP by default. If the HAP and HAC reside on the same subnet, you can configure a higher-performance HAC as the home agent. This reduces the load on the HAP and improves the forwarding efficiency.



- Configure a mobility group.
  - If a mobility server is specified, configure the mobility group on the mobility server.
  - If no mobility server is specified, configure a mobility group for each member AC.

Configuration Example (1/2)



- Deploy Layer 3 networking between the HAP and HAC and between the FAP and FAC.
- Add the HAC and FAC to a mobility group to ensure normal service traffic for STAs.

WLAN roaming configurations on AC1 and AC2

```
[AC1-wlan-view] mobility-group name mobility
[AC1-mc-mg-mobility] member ip-address 10.1.201.100
[AC1-mc-mg-mobility] member ip-address 10.1.201.200
[AC1-mc-mg-mobility] quit
```

```
[AC2-wlan-view] mobility-group name mobility
[AC2-mc-mg-mobility] member ip-address 10.1.201.100
[AC2-mc-mg-mobility] member ip-address 10.1.201.200
[AC2-mc-mg-mobility] quit
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (2/2)

- After STA roaming is completed, check the STA roaming track on the AC.

```
<AC> display station roam-track sta-mac 28b2-bd35-4af3
Access SSID:huawei-guest1
Rx/Tx: Rx-Rate/Tx-Rate Mbps
--------------------------------------------------------------------------------------------------------------------------
L2/L3     AC IP          AP name   Radio ID  BSSID          TIME                In Rx/Tx  RSSI  Out Rx/Tx  RSSI
--------------------------------------------------------------------------------------------------------------------------
--        10.1.201.100   ap1       1         cccc-8110-2250 2020/06/18 14:09:06 130/130   -44   130/130    -44
L3        10.1.201.200   ap2       1         cccc-8110-22b0 2020/06/18 14:12:24 130/6     -42   -/-
--------------------------------------------------------------------------------------------------------------------------
Number of roam track: 1
```

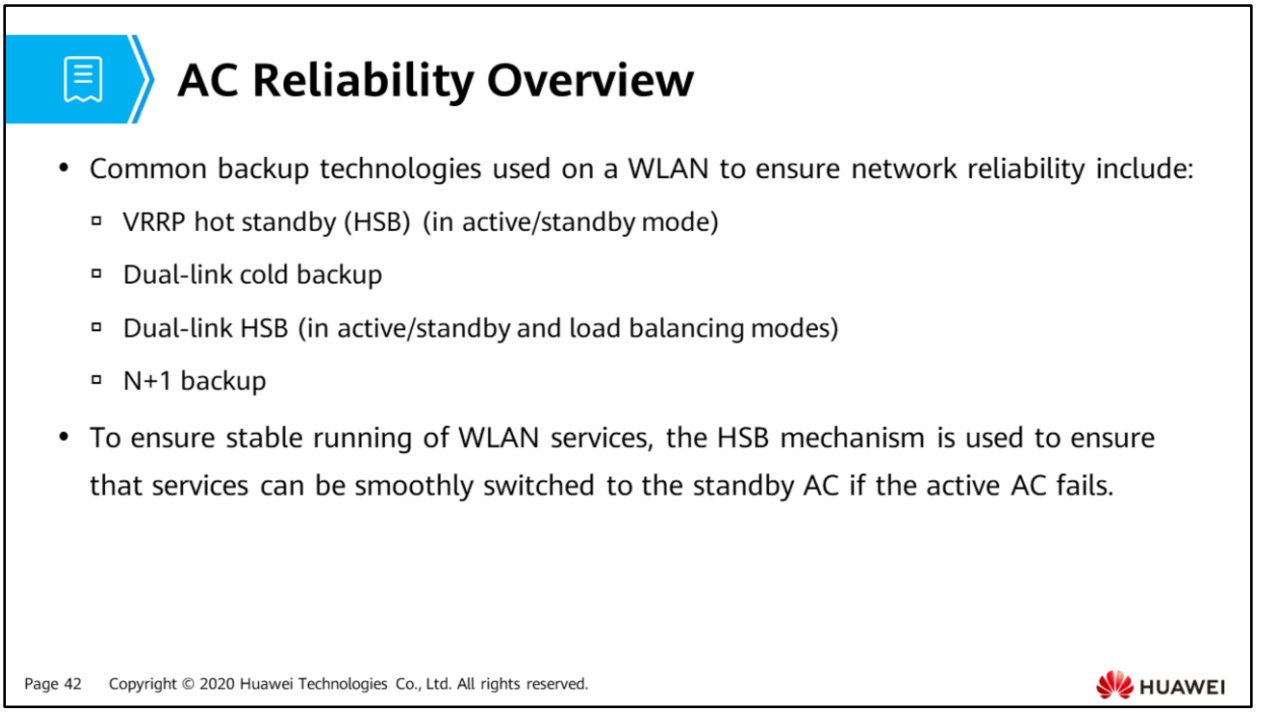

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

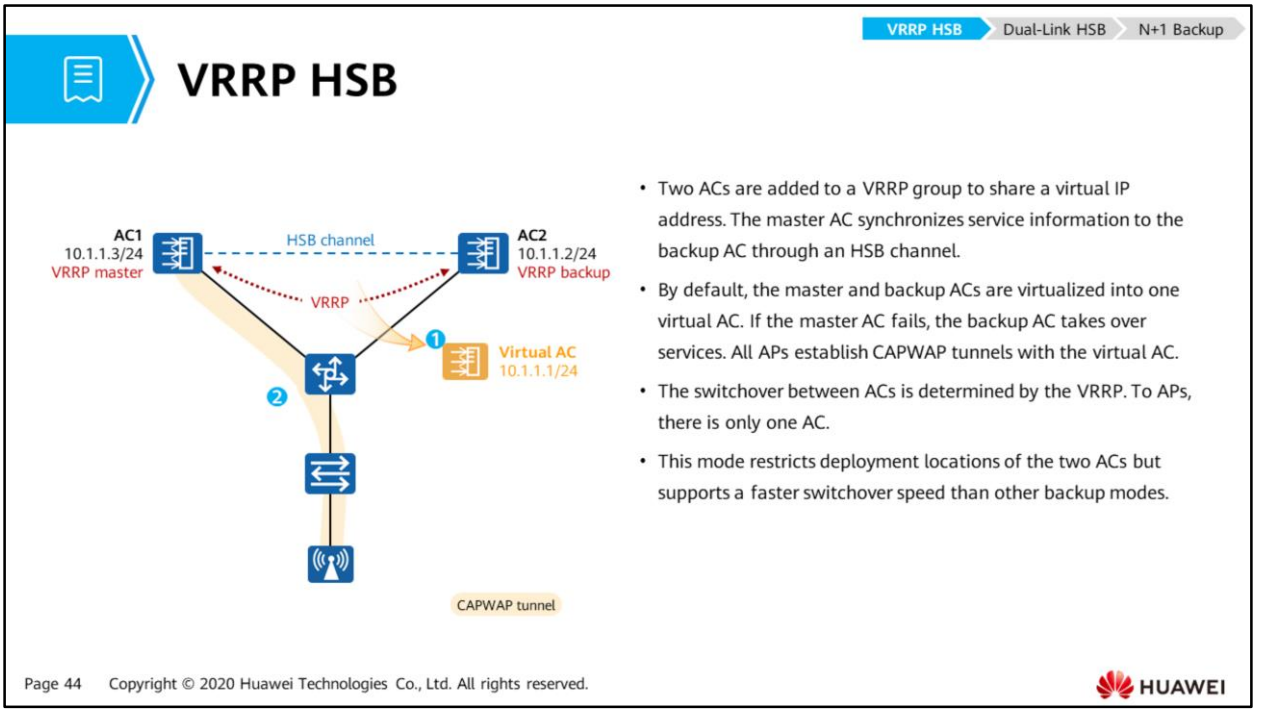
1. Large-Scale WLAN Networking Overview
2. VLAN Pool
3. DHCP
4. Roaming
5. **Reliability**
6. NAC

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI



- In HSB mode, there are two devices, one acting as the active and the other the standby. The active device forwards services and the standby device monitors the forwarding. In addition, the active device sends the standby device the status information and information that needs to be backed up in real time. In the case that the active device becomes faulty, the standby device takes over services.
- VRRP HSB
  - The active and standby ACs have independent IP addresses, which are virtualized into one using VRRP. APs set up CAPWAP links with this virtual IP address.
  - The active AC backs up information about APs, STAs, and CAPWAP links, and synchronizes such information to the standby AC through the HSB service. If the active AC fails, the standby AC takes over services.
- Dual-link HSB
  - An AP sets up an active and a standby CAPWAP link with the active and standby ACs, respectively.
  - The active AC backs up only STA information and synchronizes such information to the standby AC through the HSB service. If the active AC fails, APs connected to it switch to the standby links and the standby AC takes over services.



- Currently, the AC supports HSB of a single VRRP instance, but does not support load balancing. HSB has the following characteristics:
  - Uplinks can back up each other. The active and standby devices can track the status of uplink interfaces. The active/standby status of an AC may be different from its downlink status.
  - MSTP is used to prevent loops on multiple downlinks (including wired and wireless links). When the MSTP status changes, the MAC/ARP entries on the links are automatically deleted.

HSB Concepts

- HSB is Huawei's public active/standby mechanism.
- HSB service: establishes and maintains an HSB channel, and notifies the active and standby service modules of channel connect/disconnect events.
- HSB group: has an HSB service bound and provides a data backup channel for each of active and standby service modules. An HSB group is bound to a VRRP instance, and the active and standby roles are negotiated using the VRRP mechanism. Additionally, the HSB group instructs service modules to process events such as batch backup, real-time backup, and active/standby switchover.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- In VRRP HSB, HSB services are registered with the same HSB group, which is bound to the HSB service and a VRRP instance. In this way, services can obtain the active/standby status of the current user and active/standby switchover events through the HSB group. Additionally, backup data is sent and received through the interface of the HSB group.

HSB Service

- An HSB service establishes an HSB channel between two devices that back up each other, and maintains the link status of the HSB channel by notifying the HSB group of any link failure.
- An HSB service provides the following functions:
  - Establishes an HSB channel.
  - Maintains the link status of the HSB channel.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- The HSB service involves the following two aspects:
  - Establishing an HSB channel: A TCP channel is established for sending HSB packets by specifying the IP addresses and port numbers of the local and peer devices. The HSB service notifies the HSB group of any link failure.
  - Maintaining the link status of the HSB channel: HSB packets are sent and retransmitted to prevent long TCP interruption that is not detected by the protocol stack. If a device does not receive an HSB packet from the peer device within the period (retransmission interval x retransmission times), the local device receives an exception notification and then re-establishes a channel to the peer.

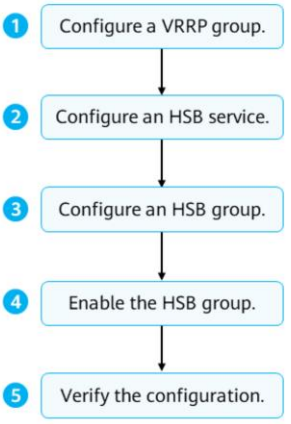
Data Synchronization

- In VRRP HSB mode, information including user entries, CAPWAP link information, and AP entries can be backed up in real time, in batches, or periodically.
  - Batch backup: The master AC synchronizes all existing session entries to the backup AC at a time to ensure information consistency between the ACs. Batch backup is triggered when the master and backup ACs are determined.
  - Real-time backup: The master AC backs up new entries or entry changes to the backup device in a timely manner.
  - Periodic backup: The backup device checks whether its existing session entries are consistent with those on the master device every 30 minutes. If they are inconsistent, the session entries on the master device are synchronized to the backup device.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- When the active AC fails, service traffic can be switched to the standby AC only if the standby AC has the same session entries as the active AC. Otherwise, the session may be interrupted. Therefore, a mechanism is required to synchronize session information to the standby device when session entries are created or modified on the active device. The HSB module provides the data backup function. It establishes an HSB channel between two devices that back up each other, maintains the link status of the HSB channel, and sends and receives packets.
- HSB service backup in real time involves backup for the following information:
  - User data information
  - CAPWAP tunnel information
  - AP entries
  - DHCP address information
- The HSB channel can be carried by the direct physical link between two ACs or by a switch. For example, the HSB channel can reuse the physical channel where VRRP packets are exchanged.

VRRP HSB Configuration Process



1. Create a VRRP group and configure a virtual IP address for the VRRP group.
2. Create an HSB service and specify the IP addresses and port numbers for establishing an HSB channel.
3. Create an HSB group, and bind the HSB service, VRRP group, WLAN service, and DHCP to the HSB group.
4. Enable the HSB group so that the HSB group configurations take effect.
5. Verify the VRRP HSB configuration.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

Configuration Procedure (1/3)

1. Create a VRRP group and configure a virtual IP address for the VRRP group on an interface.

   [AC-GigabitEthernet0/0/1] **vrrp vrid** *virtual-router-id* **virtual-ip** *virtual-address*

2. Set the AC priority in the VRRP group. The default value is 100. The master AC must have a higher priority.

   [AC-GigabitEthernet0/0/1] **vrrp vrid** *virtual-router-id* **priority** *priority-value*

3. Create an HSB service and enter the HSB service view.

   [AC] **hsb-service** *service-index*

4. Configure IP addresses and port numbers for establishing an HSB channel.

   [AC-hsb-service-0] **service-ip-port local-ip** { *local-ipv4-address*| *local-ipv6-address*} **peer-ip** { *peer-ipv4-address*| *peer-ipv6-address*} **local-data-port** *local-port* **peer-data-port** *peer-port*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Procedure (2/3)

1. Create an HSB group and enter the HSB group view.

   [AC] **hsb-group** *group-index*

2. Bind an HSB service to the HSB group.

   [AC-hsb-group-0] **bind-service** *service-index*

3. Bind a VRRP group to the HSB group.

   [AC-hsb-group-0] **track vrrp vrid** *virtual-router-id* **interface** *interface-type interface-number*

4. Bind the WLAN service to the HSB group.

   [AC] **hsb-service-type ap hsb-group** *group-index*

5. Bind the DHCP service to the HSB group.

   [AC] **hsb-service-type dhcp hsb-group** *group-index*

6. Bind the NAC service to the HSB group.

   [AC] **hsb-service-type access-user hsb-group** *group-index*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Procedure (3/3)

1. Enable the HSB group.

   [AC-hsb-group-0] **hsb enable**

2. Display information about the HSB group.

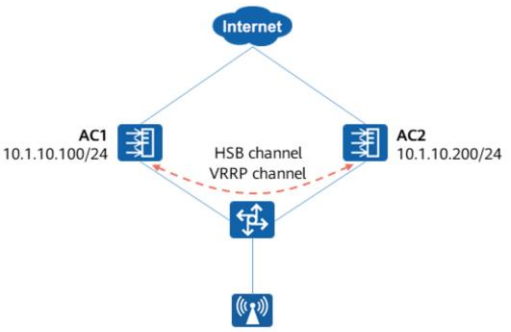
   [AC] **display hsb-group** *group-index*

3. Display information about the HSB service.

   [AC] **display hsb-service** *service-index*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (1/3)



- Add AC1 and AC2 to a VRRP group through VLANIF 10, set the virtual IP address of the VRRP group to 10.1.10.1, and configure AC1 as the master device with a priority of 120.
- Implement the HSB function.

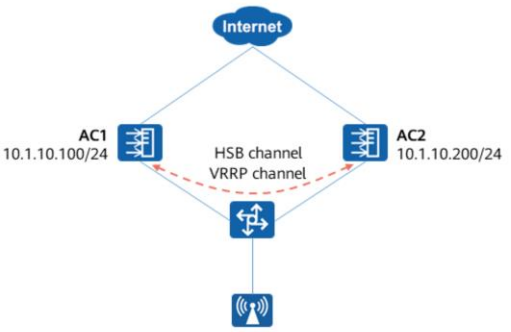
VRRP configurations on AC1 and AC2

```
[AC1]interface Vlanif10
[AC1-Vlanif10]ip address 10.1.10.100 255.255.255.0
[AC1-Vlanif10]vrrp vrid 1 virtual-ip 10.1.10.1
[AC1-Vlanif10]vrrp vrid 1 priority 120
```

```
[AC2]interface Vlanif10
[AC2-Vlanif10]ip address 10.1.10.200 255.255.255.0
[AC2-Vlanif10]vrrp vrid 1 virtual-ip 10.1.10.1
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (2/3)



- Add AC1 and AC2 to a VRRP group through VLANIF 10, set the virtual IP address of the VRRP group to 10.1.10.1, and configure AC1 as the master device with a priority of 120.
- Implement the HSB function.

HSB configuration on AC1

```
[AC1]hsb-service 0
[AC1-hsb-service-0]service-ip-port local-ip 10.1.10.100 peer-ip 10.1.10.200
local-data-port 10241 peer-data-port 10241
[AC1-hsb-service-0]quit
```

```
[AC1]hsb-group 0
[AC1-hsb-group-0]bind-service 0
[AC1-hsb-group-0]track vrrp vrid 1 interface Vlanif10
[AC1-hsb-group-0]quit
```

```
[AC1]hsb-service-type access-user hsb-group 0
[AC1]hsb-service-type dhcp hsb-group 0
[AC1]hsb-service-type ap hsb-group 0
```

```
[AC1]hsb-group 0
[AC1-hsb-group-0]hsb enable
```

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The configuration on AC2 is the same as that on AC2, and is not mentioned here.

Configuration Example (3/3)

- Check the HSB service on AC1.

```
[AC1] display hsb-service 0
Hot Standby Service Information:
----------------------------------------------------------------
 Local IP Address      : 10.1.10.100
 Peer IP Address       : 10.1.10.200
 Source Port           : 10241
 Destination Port      : 10241
 Keep Alive Times      : 2
 Keep Alive Interval   : 1
 Service State         : Connected
 Service Batch Modules :
 Shared-key            : -
----------------------------------------------------------------
```

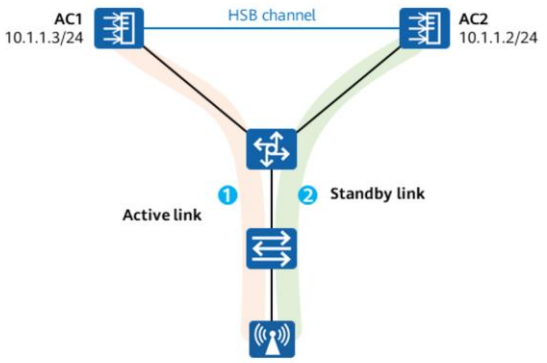
- Check the running status of the HSB group on AC1.

```
[AC1] display hsb-group 0
Hot Standby Group Information:
----------------------------------------------------------------
 HSB-group ID                : 0
 Vrrp Group ID               : 1
 Vrrp Interface              : Vlanif10
 Service Index               : 0
 Group Vrrp Status           : Master
 Group Status                : Active
 Group Backup Process        : Realtime
 Peer Group Device Name      : AirEngine 9700-M
 Peer Group Software Version : V200R019C00
 Group Backup Modules        : Access-user
                               DHCP
                               AP
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VRRP HSB | Dual-Link HSB | N+1 Backup

Dual-Link HSB

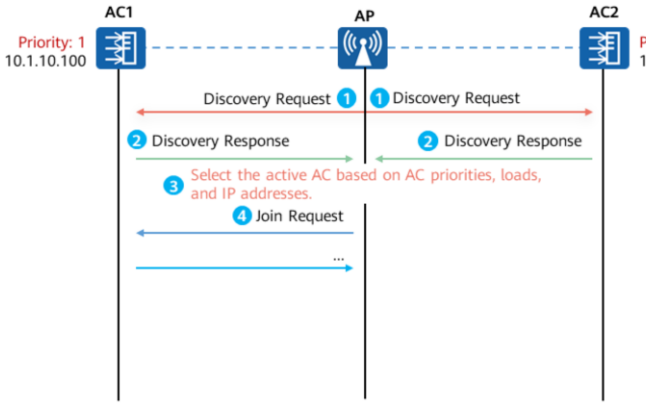


- In a dual-link HSB scenario, services are directly bound to the HSB service. In this way, ACs back up service using HSB and maintain the active/standby status based on the dual-link mechanism.
- An AP sets up CAPWAP tunnels with both the active and standby ACs. ACs synchronize service information through an HSB channel.
- When the link between the AP and active AC fails, the AP instructs the standby AC to take over services from the active AC.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

- In addition to the active/standby HSB mode, the load balancing mode is supported. In load balancing mode, you can specify AC1 as the active AC for some APs and AC2 as the active AC for other APs, so that the APs set up primary CAPWAP links with their own active ACs.
- Dual-link HSB frees active and standby ACs from location restrictions and allows for flexible deployment. The two ACs can implement load balancing to make efficient use of resources. However, service switching takes a relatively long time.
- As shown in the figure, dual-link HSB is deployed between AC1 and AC2. Only the HSB service is bound to the ACs to set up an HSB channel. An AP establishes CAPWAP tunnels with two ACs in sequence and determines the active and standby ACs based on the AC priorities in the CAPWAP packets sent from the ACs.

Active/Standby Negotiation and Active Link Establishment



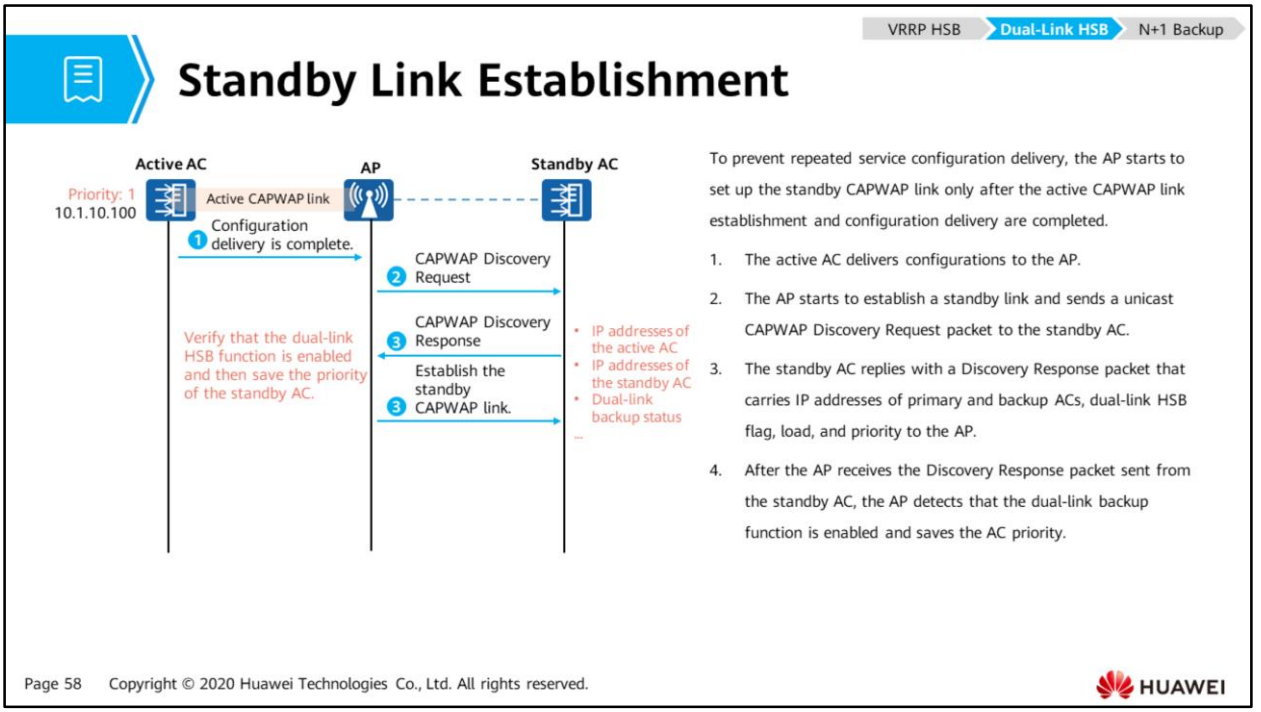
The AP sets up an active link with the AC and selects the active AC in the Discovery phase.

1. After dual-link HSB is enabled, the AP sends Discovery Request packets.
2. The ACs that receive the Discovery Request packets reply with Discovery Response packets.
3. After receiving the Discovery Response packets, the AP selects an active AC based on AC priorities, loads, and IP addresses.
4. The AP sets up an active link with the active AC.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- The procedure for establishing the active link is the same as that for establishing a normal CAPWAP tunnel, except that the active AC needs to be selected in the Discovery phase.
- After the dual-link backup function is enabled in the Discovery phase, the AP sends a Discovery Request packet in unicast or broadcast mode:
  - If the IP addresses of active and standby ACs have been allocated in static, DHCP, or DNS mode, the AP sends the Discovery Request packet in unicast mode to request connections with the ACs.
  - If no IP addresses are allocated to ACs or there is no response to the unicast packet, the AP sends another Discovery Request packet in broadcast mode to discover available ACs in the same network segment.
- Regardless of the unicast or broadcast Discovery Request packets, if the ACs function normally, they reply with the Discovery Response packets to the Discovery Request packets. A Discovery Response packet contains the dual-link backup flag, priority, load, and IP address of the ACs.



- If the priority of this AC is higher than the priority of the other AC, the AP performs an active/standby switchover only after the tunnel is set up.
- The AP sends a Join Request packet, notifying the AC that the configurations have been delivered. After receiving the Join Request packet, the AC sets up a CAPWAP tunnel with the AP but does not deliver configurations to the AP.
- After the backup tunnel is set up, the AP selects the active and standby ACs again based on the tunnel priorities.
- By default, the CAPWAP heartbeat interval is 25s and the number of CAPWAP heartbeat detections is 6. If the dual-link backup function is enabled, the CAPWAP heartbeat interval is set to 25s, and the number of CAPWAP heartbeat detections is set to 3.
- Note:
    - To configure dual-link backup on a WDS or Mesh network, set the CAPWAP heartbeat interval to 25 seconds and set the number of heartbeat packet transmissions to at least 6. If this configuration is not performed, the AC sends heartbeat packets 3 times at an interval of 25 seconds by default. This may cause unstable WDS or Mesh link status and result in user access failures.
    - If you set the CAPWAP heartbeat detection interval and the number of CAPWAP heartbeat detections smaller than the default values, CAPWAP link reliability is degraded. Exercise caution when you set the values. The default values are recommended.

Configuration Procedure (1/2)

1. Configure an IP address for the standby AC.

   [AC-wlan-view] **ac protect protect-ac** { **ip-address** *ip-address*}

2. Configure the priority for the local AC. The default value is 0.

   [AC-wlan-view] **ac protect priority** *priority*

3. Enable global revertive switching.

   [AC-wlan-view] **undo ac protect restore disable**

4. Enable the dual-link HSB function.

   [AC-wlan-view] **ac protect enable**

5. Restart APs to make the dual-link HSB function take effect.

   [AC-wlan-view] **ap-reset** { **all** | **ap-name** *ap-name* | **ap-mac** *ap-mac* | **ap-id** *ap-id* | **ap-group** *ap-group* | **ap-type** { **type** *type-name* | **type-id** *type-id* } }

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Procedure (2/2)

1. Create an HSB service and enter the HSB service view.

   [AC] **hsb-service** *service-index*

2. Configure IP addresses and port numbers for establishing an HSB channel.

   [AC] **service-ip-port local-ip** { *local-ipv4-address*| *local-ipv6-address*} **peer-ip** { *peer-ipv4-address*| *peer-ipv6-address*} **local-data-port** *local-port* **peer-data-port** *peer-port*

3. Bind the WLAN service to the HSB group.

   [AC] **hsb-service-type ap hsb-group** *group-index*

4. Bind the DHCP service to the HSB group.

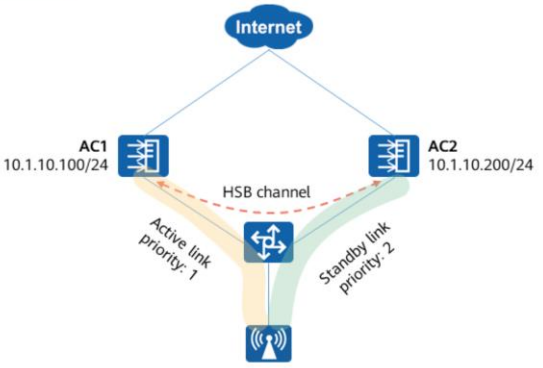
   [AC] **hsb-service-type dhcp hsb-group** *group-index*

5. Bind the NAC service to the HSB group.

   [AC] **hsb-service-type access-user hsb-service** *service-index*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (1/3)



- Configure dual-link HSB for AC1 and AC2. Specify AC1 as the active device with the priority of 1, and AC2 as the standby device with the priority of 2.
- Implement the HSB function.

Configuration on AC1

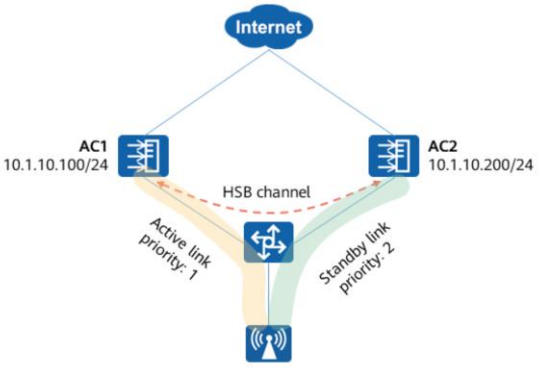
```
[AC1] wlan
[AC1-wlan-view] ac protect enable
[AC1-wlan-view] ac protect protect-ac 10.1.10.200 priority 1
```

```
[AC1] hsb-service 0
[AC1-hsb-service-0] service-ip-port local-ip 10.1.10.100 peer-ip 10.1.10.200
local-data-port 10241 peer-data-port 10241
[AC1-hsb-service-0] quit
```

```
[AC1] hsb-service-type access-user hsb-group 0
[AC1] hsb-service-type dhcp hsb-group 0
[AC1] hsb-service-type ap hsb-group 0
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (2/3)



- Configure dual-link HSB for AC1 and AC2. Specify AC1 as the active device with the priority of 1, and AC2 as the standby device with the priority of 2.
- Implement the HSB function.

Configuration on AC2

```
[AC2] wlan
[AC2-wlan-view] ac protect enable
[AC2-wlan-view] ac protect protect-ac 10.1.10.100 priority 2
```

```
[AC2] hsb-service 0
[AC2-hsb-service-0] service-ip-port local-ip 10.1.10.200 peer-ip 10.1.10.100
local-data-port 10241 peer-data-port 10241
[AC2-hsb-service-0] quit
```

```
[AC2] hsb-service-type access-user hsb-group 0
[AC2] hsb-service-type dhcp hsb-group 0
[AC2] hsb-service-type ap hsb-group 0
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (3/3)

- Check the dual-link HSB configuration on AC1.

```
[AC1] display ac protect
------------------------------------------------------------
Protect state          : enable
Protect AC IPv4        : 10.1.10.200
Protect AC IPv6        : -
Priority               : 0
Protect restore        : enable
...
```

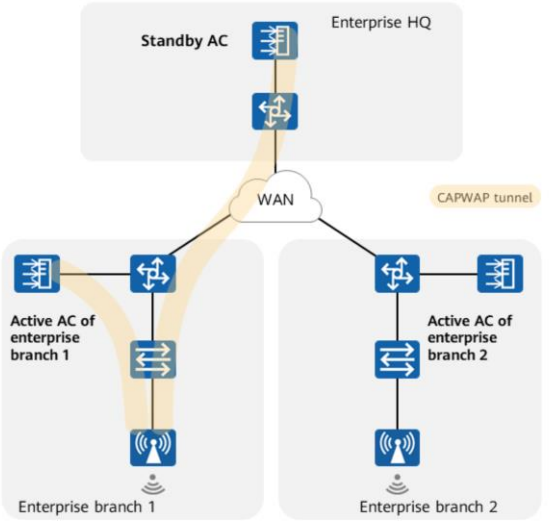
- Check the HSB service on AC1.

```
[AC1] display hsb-service 0
Hot Standby Service Information:
------------------------------------------------------------
 Local IP Address       : 10.1.10.100
 Peer IP Address        : 10.1.10.200
 Source Port            : 10241
 Destination Port       : 10241
 Keep Alive Times       : 5
 Keep Alive Interval    : 3
 Service State          : Connected
 Service Batch Modules  : AP
                          Access-user
                          DHCP
...
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

VRRP HSB > Dual-Link HSB > N+1 Backup

AC Reliability: N+1 Backup



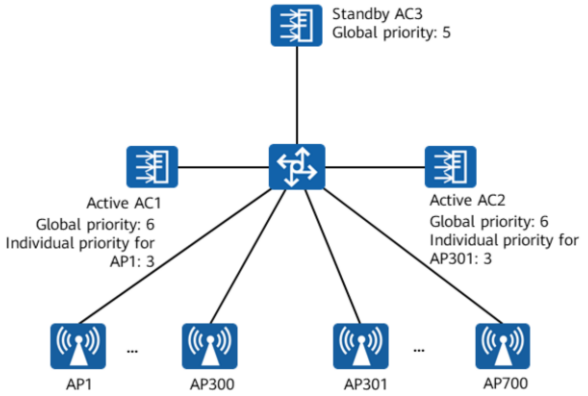
- N+1 backup uses one AC to provide backup services for multiple active ACs on an AC + Fit AP network.
- When the network runs properly, an AP sets up a CAPWAP link only with the active AC to which it belongs.
- If the active AC or the CAPWAP link fails, the standby AC replaces the active AC to manage the AP and establishes a CAPWAP link with the AP to provide services.
- Active/Standby switchover and switchback are supported.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- When the CAPWAP tunnel between an AP and the active AC is disconnected, the AP attempts to establish a CAPWAP tunnel with the standby AC. After the new CAPWAP tunnel is established, the AP restarts and obtains configurations from the standby AC. During this process, services are affected.

N+1 Backup: Selecting Active and Standby ACs



- In the Discovery phase, after the AP discovers ACs, it selects the AC with the highest priority as the active AC and connects to it.
- An AC has two types of priorities:
  - Global priority: AC priority configured for all APs. The default value is 0, and the maximum value is 7. A smaller value indicates a higher priority.
  - Individual priority: AC priority configured for a single AP or APs in a specified AP group. No default value is available.
- Global AC priority < Individual AC priority for an AP

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved. HUAWEI

- When the AC receives a Discovery Request packet from an AP, if no individual priority is configured for the AP, the AC sends a Discovery Response packet carrying the global priority to the AP.
- If the AC has an individual priority for the AP, the AC sends a Discovery Response packet carrying the individual priority to the AP.
- Configure proper priorities on the active and standby ACs to control access of APs on the two ACs.
- The AP selects the active AC based on the following rules:
  - Check primary ACs on the AP. If there is only one primary AC, the AP selects it as the active AC. If there are multiple primary ACs, the AP selects the AC with the lowest load as the active AC. If the loads are the same, the AP selects the AC with the smallest IP address as the active AC.
  - Compare AC loads, that is, numbers of access APs and STAs. The AP selects the AC with the lowest load as the active AC. The number of allowed APs is compared ahead of the number of allowed STAs. When the numbers of allowed APs are the same on ACs, the AP selects the AC that can connect more STAs as the active AC.

Configuration Procedure (1/2)

1. Create an AP system profile and enter the AP system profile view.

   [AC-wlan-view] **ap-system-profile name** *profile-name*

2. Configure an IP address for the primary AC.

   [AC-wlan-ap-system-prof-huawei] **primary-access { ip-address** *ip-address* | **ipv6-address** *ipv6-address* }

3. Configure an IP address for the backup AC.

   [AC-wlan-ap-system-prof-huawei] **backup-access { ip-address** *ip-address* | **ipv6-address** *ipv6-address* **}**

4. Bind the AP system profile to an AP group.

   [AC-wlan-ap-group-huawei] **ap-system-profile** *profile-name*

5. Bind the AP system profile to an AP.

   [AC-wlan-ap-0] **ap-system-profile** *profile-name*

6. Restart APs to make the dual-link HSB function take effect.

   [AC-wlan-view] **ap-reset { all | ap-name** *ap-name* | **ap-mac** *ap-mac* | **ap-id** *ap-id* | **ap-group** *ap-group* | **ap-type { type** *type-name* | **type-id** *type-id* **} }**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Procedure (2/2)

1. Enable global revertive switching.

   [AC-wlan-view] **undo ac protect restore disable**

2. Configure the CAPWAP heartbeat interval and the number of CAPWAP heartbeat detections.

   [AC] **capwap echo { interval** *interval-value* **| times** *times-value* **}**

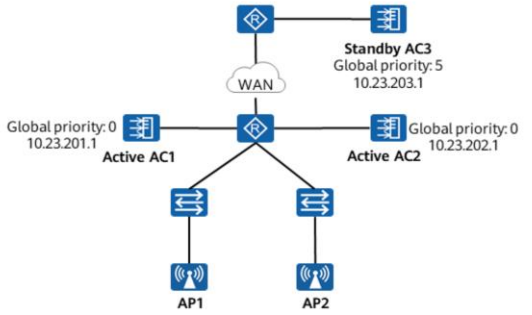
3. Enable N+1 backup.

   [AC-wlan-view] **undo ac protect enable**

   The **ac protect enable** command enables dual-link HSB globally and disables N+1 backup.
   The **undo ac protect enable** command disables dual-link HSB globally and enables N+1 backup.
   By default, dual-link HSB is disabled globally, and N+1 backup is enabled.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (1/5)



- Each AC has been configured to communicate with other network devices.
- Configure AC1 as the active AC for AP1 and AC2 as the active AC for AP2. On the active ACs, configure the active and standby AC information.
- Configure AC3 as the standby AC for AP1 and AP2, and configure two AP groups and basic WLAN services on AC3. Ensure that service configurations on AC3 are the same as those on AC1 and AC2.
- Configure N+1 backup on the active and standby ACs. After the configuration is completed, restart all APs.

Configuration on AC1

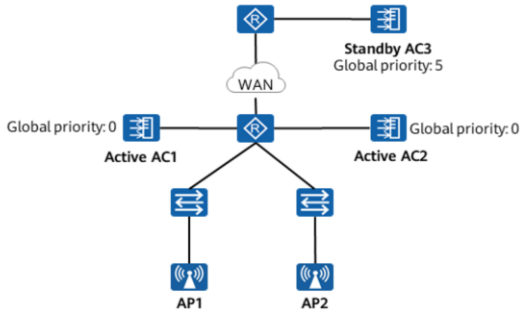
```
[AC1-wlan-view] ap-system-profile name ap-system1
[AC1-wlan-ap-system-prof-ap-system1] primary-access ip-address 10.23.201.1
[AC1-wlan-ap-system-prof-ap-system1] backup-access ip-address 10.23.203.1
[AC1-wlan-ap-system-prof-ap-system1] quit
[AC1-wlan-view] ap-group name ap-group1
[AC1-wlan-ap-group-ap-group1] ap-system-profile ap-system
[AC1-wlan-ap-group-ap-group1] quit
```

Configuration on AC2

```
[AC2-wlan-view] ap-system-profile name ap-system2
[AC2-wlan-ap-system-prof-ap-system2] primary-access ip-address 10.23.202.1
[AC2-wlan-ap-system-prof-ap-system2] backup-access ip-address 10.23.203.1
[AC2-wlan-ap-system-prof-ap-system2] quit
[AC2-wlan-view] ap-group name ap-group2
[AC2-wlan-ap-group-ap-group2] ap-system-profile ap-system2
[AC2-wlan-ap-group-ap-group2] quit
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.
HUAWEI

Configuration Example (2/5)



- Each AC has been configured to communicate with other network devices.
- Configure AC1 as the active AC for AP1 and AC2 as the active AC for AP2. On the active ACs, configure the active and standby AC information.
- Configure AC3 as the standby AC for AP1 and AP2, and configure two AP groups and basic WLAN services on AC3. Ensure that service configurations on AC3 are the same as those on AC1 and AC2.
- Configure N+1 backup on the active ACs first and then on the standby AC. After the configuration is completed, restart all the APs.

Configuration on AC3

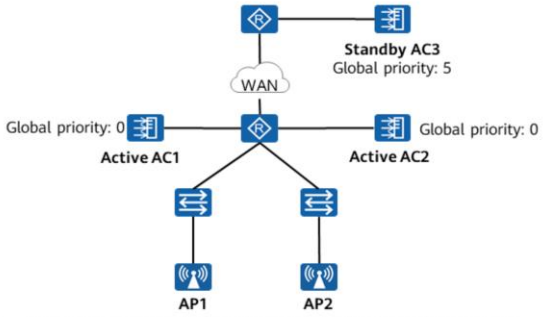
```
[AC3-wlan-view] ap-system-profile name ap-system1
[AC3-wlan-ap-system-prof-ap-system1] primary-access ip-address 10.23.201.1
[AC3-wlan-ap-system-prof-ap-system1] backup-access ip-address 10.23.203.1
[AC3-wlan-ap-system-prof-ap-system1] quit
```

```
[AC3-wlan-view] ap-system-profile name ap-system2
[AC3-wlan-ap-system-prof-ap-system2] primary-access ip-address 10.23.202.1
[AC3-wlan-ap-system-prof-ap-system2] backup-access ip-address 10.23.203.1
[AC3-wlan-ap-system-prof-ap-system2] quit
```

```
[AC3-wlan-view] ap-group name ap-group1
[AC3-wlan-ap-group-ap-group1] ap-system-profile ap-system1
[AC3-wlan-ap-group-ap-group1] quit
[AC3-wlan-view] ap-group name ap-group2
[AC3-wlan-ap-group-ap-group2] ap-system-profile ap-system2
[AC3-wlan-ap-group-ap-group2] quit
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (3/5)



- Each AC has been configured to communicate with other network devices.
- Configure AC1 as the active AC for AP1 and AC2 as the active AC for AP2. On the active ACs, configure the active and standby AC information.
- Configure AC3 as the standby AC for AP1 and AP2, and configure two AP groups and basic WLAN services on AC3. Ensure that service configurations on AC3 are the same as those on AC1 and AC2.
- Configure N+1 backup on the active ACs first and then on the standby AC. After the configuration is completed, restart all the APs.

Configurations on the ACs

[AC1-wlan-view] **undo ac protect enable**
Info: Backup function has already disabled.
[AC1-wlan-view] **ap-reset all**
Warning: Reset AP(s), continue?[Y/N]: **y**

[AC2-wlan-view] **undo ac protect enable**
Info: Backup function has already disabled.
[AC2-wlan-view] **ap-reset all**
Warning: Reset AP(s), continue?[Y/N]: **y**

[AC3-wlan-view] **undo ac protect restore disable**
Info: Protect restore has already enabled.
[AC3-wlan-view] **undo ac protect enable**
Info: Backup function has already disabled.
[AC3-wlan-view] **ap-reset all**
Warning: Reset AP(s), continue?[Y/N]: **y**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- By default, global revertive switchover is enabled. The system displays an Info message if you run the **undo ac protect restore disable** command.
- By default, N+1 backup is enabled. The system displays an Info message if you run the **undo ac protect enable** command. Run the **ap-reset all** command on the active AC to restart all APs. After the APs are restarted, N+1 backup starts to take effect.

Configuration Example (4/5)

- Check N+1 backup information on AC1 (active).

```
[AC1] display ac protect
------------------------------------------------------------------
Protect state       : disable
Protect AC IPv4     : -
Protect AC IPv6     : -
Priority            : 0
Protect restore     : enable
```

```
[AC1] display ap-system-profile name ap-system
------------------------------------------------------------------------------
AC priority             : -
Protect AC IP address   : -
Primary AC              : 10.23.201.1
Backup AC               : 10.23.203.1
```

- Check N+1 backup information on AC2 (active).

```
[AC2] display ac protect
------------------------------------------------------------------
Protect state       : disable
Protect AC IPv4     : -
Protect AC IPv6     : -
Priority            : 0
Protect restore     : enable
```

```
[AC2] display ap-system-profile name ap-system
------------------------------------------------------------------------------
AC priority             : -
Protect AC IP address   : -
Primary AC              : 10.23.202.1
Backup AC               : 10.23.203.1
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Configuration Example (5/5)

- Check N+1 backup information on AC3 (standby).

```
[AC3] display ac protect
------------------------------------------------------------------
Protect state       : disable
Protect AC IPv4     : -
Protect AC IPv6     : -
Priority            : 0
Protect restore     : enable
...
```

```
[AC3-wlan-view] display ap-system-profile name ap-system1
------------------------------------------------------------------------------
AC priority                 : -
Protect AC IP address       : -
Primary AC                  : 10.23.201.1
Backup AC                   : 10.23.203.1
...
```

```
[AC3-wlan-view] display ap-system-profile name ap-system2
------------------------------------------------------------------------------
AC priority                 : -
Protect AC IP address       : -
Primary AC                  : 10.23.202.1
Backup AC                   : 10.23.203.1
...
```


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Summary: AC Reliability Technologies

Item	VRRP HSB	Dual-Link HSB	N+1 Backup
Switching speed	The switchover speed is fast, with little impact on services. The configuration of the VRRP preemption delay implements a faster switchover than other backup modes.	The AP status switchover is slow and occurs only when CAPWAP link disconnection timeout is detected. After the AP status is switched, STAs do not need to go offline and online again.	The AP status switchover is slow and occurs only when CAPWAP link disconnection timeout is detected. APs and STAs need to go online again, and services are interrupted for a short period of time.
Deployment of active and standby ACs at different places	Not recommended	Supported	Supported
Constraints	The models and software versions of the active and standby ACs must be the same. One standby AC can provide backup for only one active AC.	The models and software versions of the active and standby ACs must be the same. One standby AC can provide backup for only one active AC.	The software versions of the active and standby ACs must be the same. No constraint is placed on the AC model. A standby AC can provide backup services for multiple active ACs, which reduces device purchase costs.
Applicability	Scenarios that require high reliability, without the need for AC deployment at different places	Scenarios that require high reliability and AC deployment at different places	Scenarios with low reliability but high cost control requirements

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

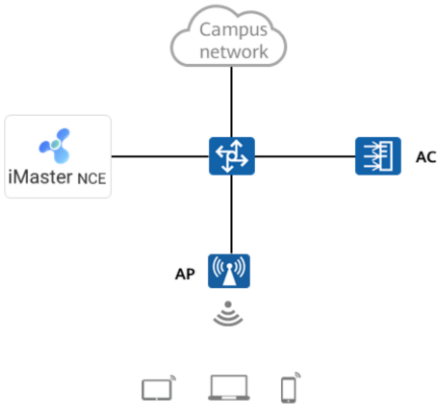
1. Large-Scale WLAN Networking Overview
2. VLAN Pool
3. DHCP
4. Roaming
5. Reliability
6. **NAC**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

NAC Overview

- Network Access Control (NAC) is an end-to-end security technology, which ensures network security by authenticating clients and users who attempt to access a network.



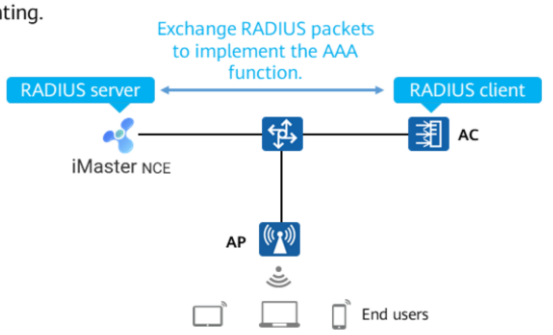
NAC and AAA work together to implement access authentication.

- NAC:
  - Enables interaction between users and access devices.
  - Controls users' access modes (802.1X, MAC address authentication, or Portal authentication), access parameters, and timers.
  - Ensures secure and stable connections between authorized users and access devices.
- AAA:
  - Enables interaction between access devices and authentication servers.
  - Controls the access rights of access users through authentication, authorization, and accounting.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

RADIUS Overview

- AAA can be implemented using multiple protocols. RADIUS is the most frequently used one in actual applications.
- RADIUS is a distributed protocol that uses the client/server model, which protects a network from unauthorized access. It is often used in network environments that require high security and allow remote user access.
- RADIUS defines the UDP-based RADIUS packet format and transmission mechanism, and specifies UDP ports 1812 and 1813 respectively for authentication and accounting.
- RADIUS has the following features:
  - Client/Server model
  - Secure message exchange mechanism
  - Fine scalability



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

802.1X Authentication

- 802.1X, an IEEE authentication standard for users' network access control, is mainly used to address the authentication and security problems on the Ethernet.
- An 802.1X authentication system uses a typical client/server architecture and consists of three entities: supplicant, authenticator, and authentication server.
- An authentication server is usually a RADIUS server that provides authentication, authorization, and accounting services to supplicants.
- 802.1X authentication is recommended for authenticating employees in medium- and large-sized enterprises.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- An 802.1X authentication system uses the Extensible Authentication Protocol (EAP) to enable information exchange among the supplicant, authenticator, and authentication server. Common 802.1X authentication protocols include the Protected Extensible Authentication Protocol (PEAP) and the Transport Layer Security (TLS). Their differences are as follows:
  - PEAP: The administrator assigns a user name and a password to a user. The user enters the user name and password for authentication when accessing the WLAN.
  - TLS: Users use certificates for authentication. This authentication mode is generally used together with enterprise apps, such as Huawei EasyAccess.
- 802.1X authentication is recommended for authenticating employees in medium- and large-sized enterprises.

MAC Address Authentication

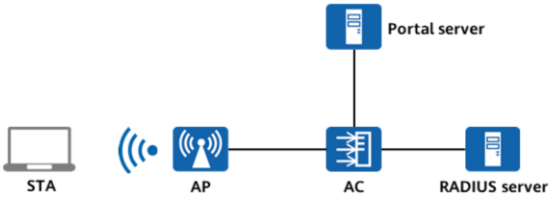


- MAC address authentication controls network access rights of a user based on the MAC address. The user does not need to install any client software.
- The access device starts authenticating a user when detecting the user's MAC address for the first time on the interface where MAC address authentication has been enabled.
- During the authentication process, the user does not need to enter the user name or password.
- MAC address authentication is usually used for dumb terminals (such as printers) or works with the authentication server to implement MAC address-prioritized Portal authentication. After a user is authenticated for the first time, the user can access the network again without authentication within a specified period.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Portal Authentication

- Portal authentication is also called web authentication, which uses a browser as the authentication client and does not require independent authentication client.
- Before a user accesses the Internet, user authentication is required on the Portal page. The network resources are not available to the user until the authentication succeeds. In addition, service providers can develop services on the Portal page, for example, placing advertisements.
- Portal authentication is recommended for authenticating visitors of large- and medium-sized enterprises, business exhibitions, and public places.
- Common Portal authentication modes are as follows:
  - User name and password authentication: The front-end administrator applies for a temporary account for a visitor. The visitor then uses the temporary account for authentication.
  - SMS authentication: Visitors are authenticated using verification codes sent to their mobile phones.

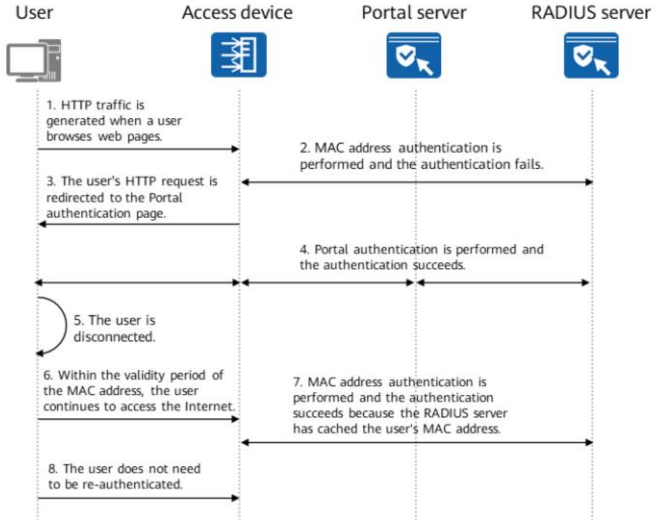


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

HUAWEI

- Definition
  - Portal authentication is also called web authentication. Generally, Portal authentication websites are referred to as web portals. Before a user accesses the Internet, user authentication is required on the Portal page. If the authentication fails, the user can access only specified network resources. The user can access other network resources only after the authentication succeeds.
- Advantages
  - Ease of use: In most cases, Portal authentication does not require additional client software and allows users to be directly authenticated on the web page.
  - Convenient operations: Business development can be achieved on the Portal page such as advertisements push and enterprise publicity.
  - Mature technology: Portal authentication has been widely used in networks of carriers, fast food chains, hotels, and schools.
  - Flexible deployment: Portal authentication implements access control at the access layer or at the ingress of critical data.
  - Flexible user management: Users can be authenticated based on the combination of user names and any one of VLANs, IP addresses, and MAC addresses.

MAC Address-Prioritized Portal Authentication



Technology background

After the Portal authentication succeeds, the user needs to enter the user name and password again to reconnect to the network if the user disconnects from it. User experience is poor.

MAC address-prioritized Portal authentication

- MAC address-prioritized Portal authentication allows disconnected users who passed Portal authentication to reconnect to the network within a certain period of time after the disconnection, without entering the user name and password, as long as they pass MAC address authentication.
- To use this function, you need to configure MAC address authentication and Portal authentication on the access device, and enable MAC address-prioritized Portal authentication and set the MAC address validity period on the authentication server.

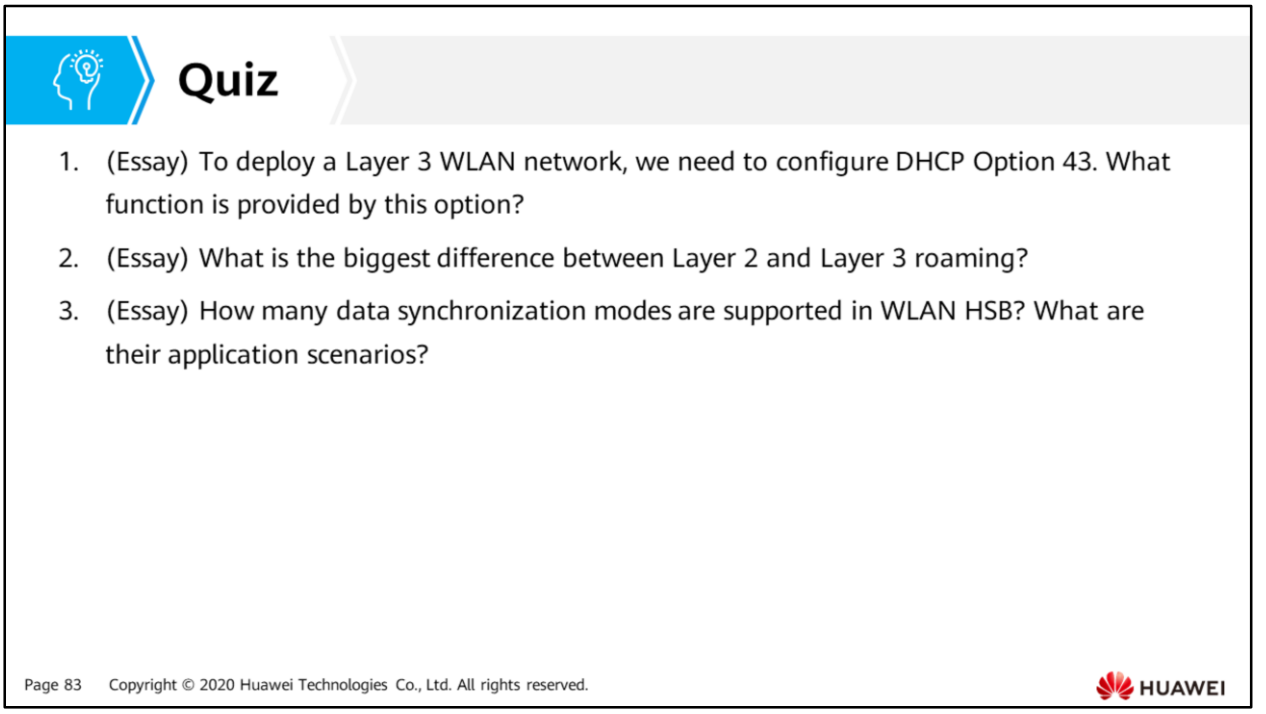
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Comparison Among Three Authentication Modes

- NAC provides three authentication modes: 802.1X authentication, MAC address authentication, and Portal authentication. The three authentication modes are implemented differently and are applicable to different scenarios. In practice, you can use a proper authentication mode or multiple authentication modes based on scenarios. The combination of authentication modes depends on device specifications.

Item	802.1X Authentication	MAC Address Authentication	Portal Authentication
Application Scenario	New network with concentrated users and high security requirements	Access authentication of dumb terminals such as printers and fax machines	Scenario where users are sparsely distributed or move freely
Client	Required	Not required	Not required
Advantage	High security	No client required	Flexible deployment
Disadvantage	Inflexible deployment	MAC address registration required, making management complex	Low security

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



1. DHCP broadcast request packets sent by an AP are Layer 2 packets and cannot be transmitted over Layer 3 networks. Therefore, the AP cannot discover the AC on a Layer 3 network in broadcast mode. Option 43 must be configured to advertise the AC's IP address. Otherwise, the AP cannot obtain the AC's IP address and will fail to go online.
2. The difference is that the SSIDs of the APs associated with STAs before and after roaming belong to different VLANs.
    - Layer 2 roaming allows STAs to roam within the same subnet.
    - Layer 3 roaming allows STAs to roam between different subnets.
3. Data can be backed up in batches, in real time, or periodically.
    - Batch backup: After HSB is configured, the active device synchronizes existing session entries to the standby device at a time.
    - Real-time backup: The active device backs up new entries or entry changes to the standby device in a timely manner.
    - Periodic backup: The standby device checks whether its existing session entries are the same as those on the active device every 30 minutes. If so, the standby device synchronizes the session entries from the active device.

Summary

- Large-scale WLAN networking allows users to easily and securely access a wireless network and move freely within the Wi-Fi coverage area, eliminating the restrictions of wired networks.
- In this course, we have learned large-scale WLAN networking technologies, including the large-scale WLAN architecture, key technologies, roaming, reliability, and NAC technologies.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com



Introduction to Enterprise Datacom Solutions

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Foreword

- In 1946, the first electronic computer was invented, opening the door to the information age for mankind and also laying the foundation for the development of information and communications technology (ICT).
- The ever-changing ICT landscape keeps transforming our lifestyles. In the past few years, new ICT represented by the Internet of Things (IoT), big data, cloud computing, and artificial intelligence (AI) has gradually become an important engine driving social and economic transformation and increasingly penetrated into all industries.
- Data communications networks ("datacom networks") are the infrastructure for data interactions and also the cornerstone for building a digital world. The explosive growth of data traffic and storage, driven by new ICT, poses greater challenges to datacom networks than ever before. To address these challenges, major vendors in the industry constantly optimize their network solutions to meet digital transformation requirements.
- Enterprise networks can be classified into campus networks, WLANs, data center networks, and WANs, depending on application purposes. This course describes the datacom network classifications, architectures, typical application scenarios, trends, and challenges in each domain and Huawei solutions in such domains.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Objectives

- Upon completion of this course, you will be able to:
  - Describe the functions of datacom networks in the digital world.
  - Describe the scenario classifications and basic concepts of datacom network solutions.
  - Describe Huawei's campus network, WLAN, data center network, and WAN solutions.

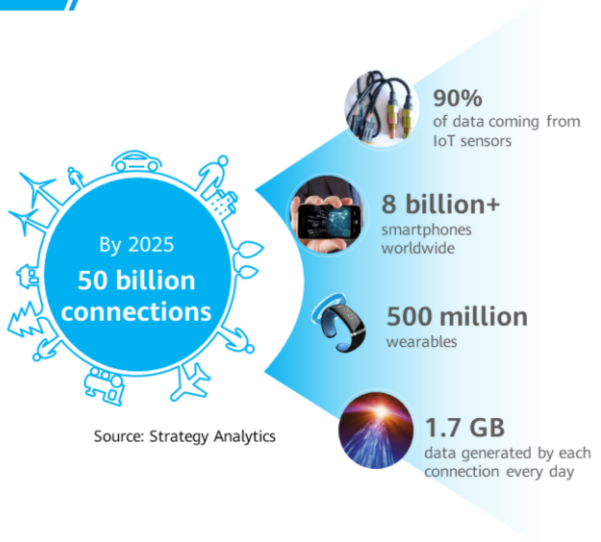
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. **Enterprise Datacom Network Panorama**
2. Enterprise Datacom Network Solutions

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Connectivity of Everything Builds a Digital World

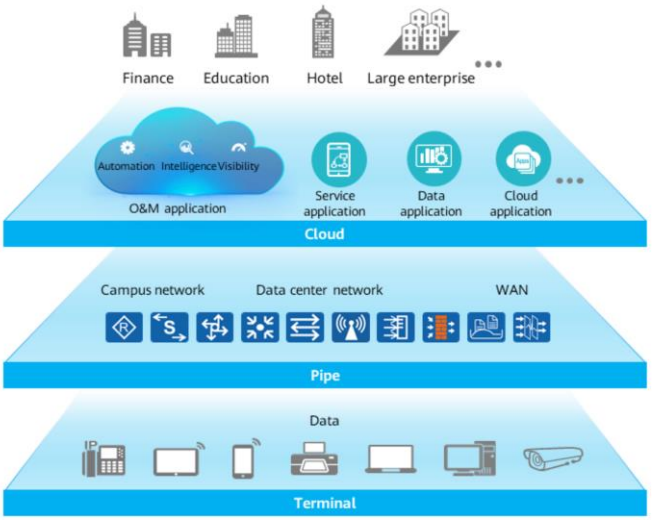


Connectivity of everything:

- People to people, people to things, and things to things are connected.
- The essence is data connectivity and computing.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Datacom Networks Are "Pipes" for Connectivity of Everything



Computing

Supplies computing power to pipes and terminals

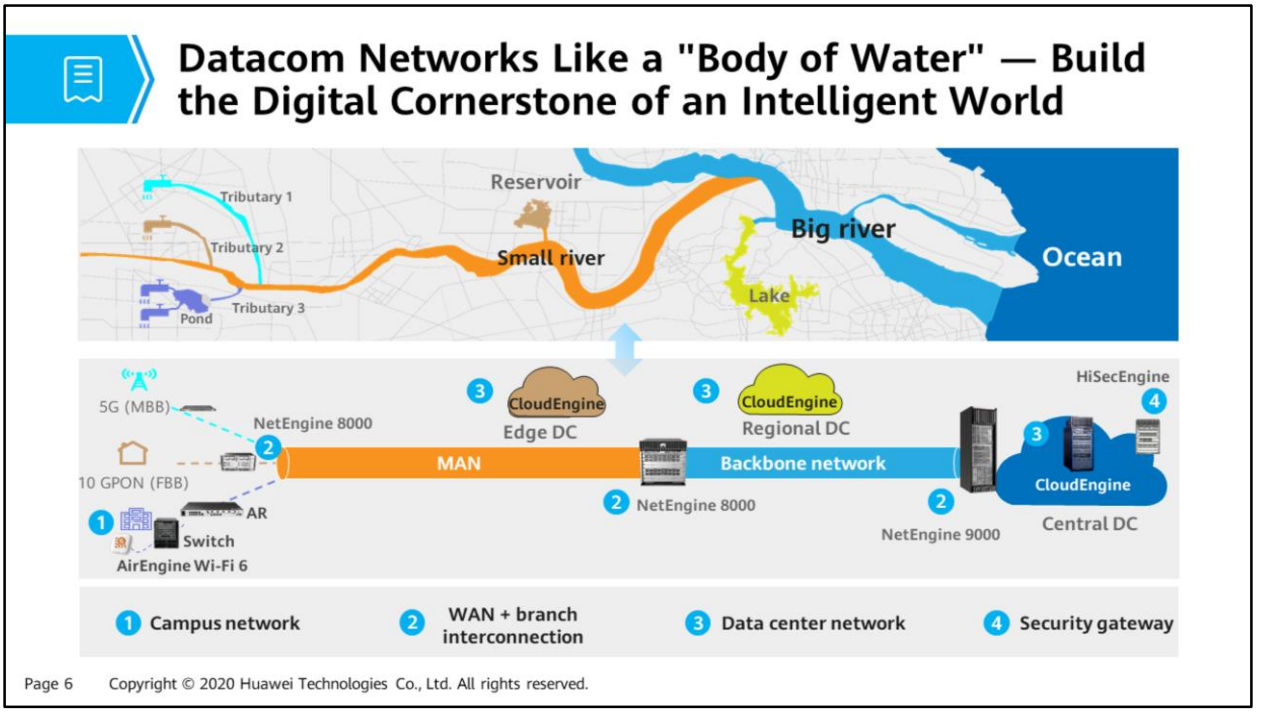
Connectivity

- Provides connections for data interactions between terminals and the cloud
- Senses applications and services
- Achieves simplicity, intelligence, ultra-broadband, and automation

Awareness

Provides service data and displays service results

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Datacom networks are like a body of water consisting of rivers, lakes, oceans, and others. The datacom industry is the digital cornerstone for building an intelligent world. Many datacom beginners may not have a comprehensive understanding of the datacom industry. We can think of the datacom industry as a body of water. Through this analogy, you can feel what the datacom is and how important it is.
    1. The datacom industry is a truly fully-connected industry in the connectivity field. Datacom networks are available at each network layer, just like a fully-connected body of water.
    2. 5G access and home broadband access are like tributaries, and campus networks are like ponds.
    3. The MAN is like a small river, and the backbone network is like a big river.
    4. The central DC is like an ocean, the regional DC is like a large lake, and the edge DC is like a reservoir.
    5. Huawei provides best-in-class services for customers in the fields of campus networks, WANs and branch networks, data center networks, and network security. Behind these are Huawei's four-engine products and solutions, namely, AirEngine (for campus network solutions), NetEngine (for MAN and backbone network solutions), CloudEngine (for DCN solutions), and HiSecEngine (for security solutions).

Tributary 1: Campus Network



- Campus is a broad concept. It can be said that "a city, except roads, is composed by campuses". 90% of activities in a city happen in campuses, and most data is aggregated from campuses.
- A campus network is an internal network of an enterprise or organization. It is related to WAN interconnection and data centers.
- The main purpose of a campus network is to operate the main business of an enterprise more efficiently.
- WLANs are widely used on campus networks. Nowadays, more campus networks prefer to use wireless.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Tributary 2: WLAN



- A WLAN is a wireless local area network (LAN) constructed by using wireless technologies, such as Wi-Fi, infrared (IR), Bluetooth, and ZigBee.
- Wi-Fi is a WLAN technology based on the IEEE 802.11 standard.
- The most well-known standards that contribute to Wi-Fi evolution are 802.11b, 802.11a, 802.11g, 802.11n, and 802.11ac. In 2018, the Wi-Fi Alliance simplified Wi-Fi naming conventions and renamed the latest 802.11ax standard Wi-Fi 6.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Lake & Ocean: Data Center Network



- A data center is used to transmit, display, compute, and store massive amounts of data.
- A data center is a complex set of facilities, including buildings. It includes not only computer systems, but also other supporting equipment (e.g. communications and storage systems), environmental control equipment, monitoring equipment, and various security facilities.
- A data center network provides internal and external data center interconnection.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

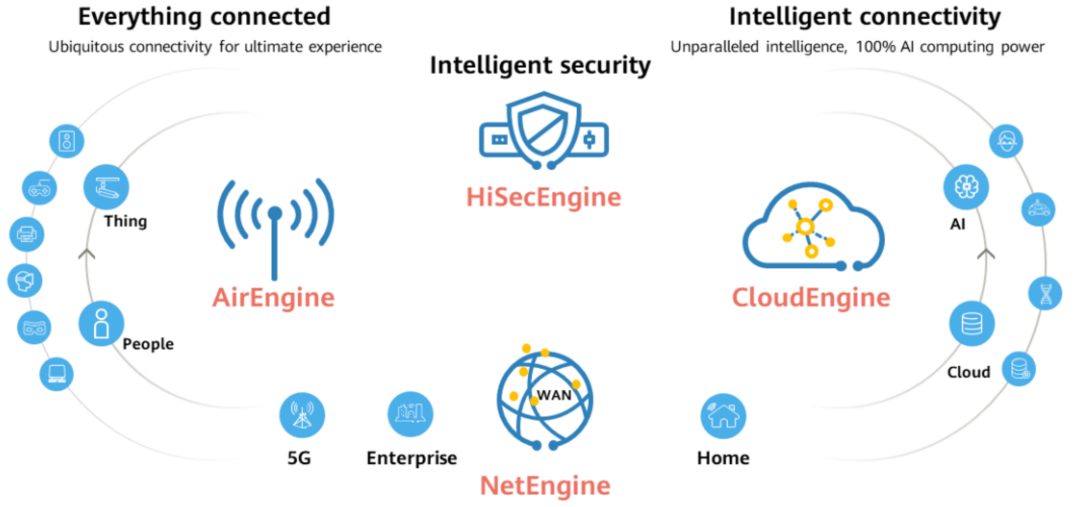
River: WAN



- A wide area network (WAN) is a remote network that connects multiple LANs or metropolitan area networks (MANs) in different areas for communication.
- WANs typically cover large geographical areas. They connect multiple regions, cities, and countries or span several continents while providing long-distance communication to form an international remote network.
- WANs are used to interconnect campus networks or data center networks.

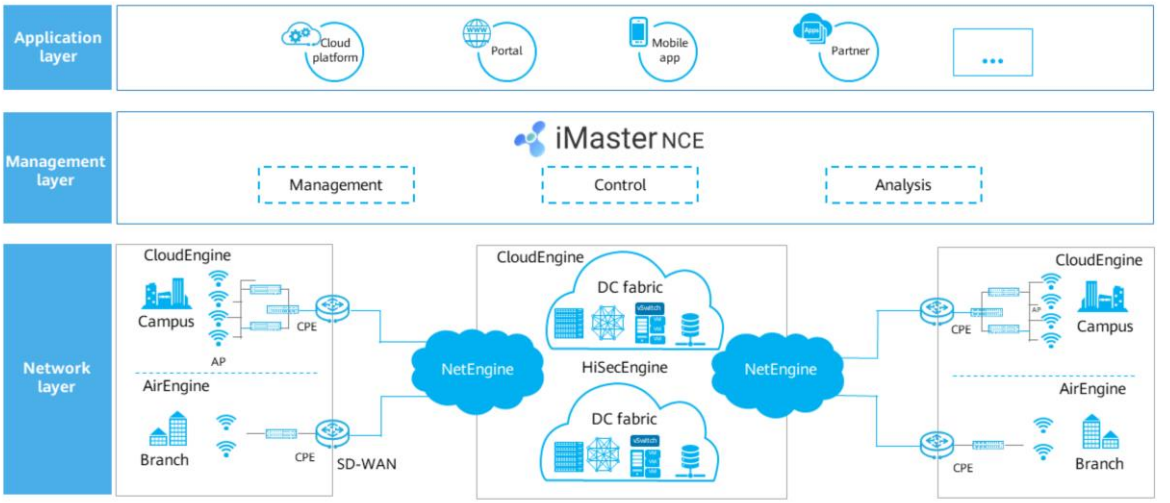
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Huawei's Four Engines, Ideal for Building Datacom Networks



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Overview of Huawei Enterprise Datacom Solutions



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

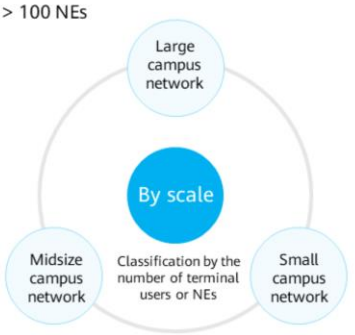
Contents

1. Enterprise Datacom Network Panorama
2. **Enterprise Datacom Network Solutions**
   - **Campus Network**
   - WLAN
   - DCN
   - Bearer WAN
   - SD-WAN

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Campus Network Classification

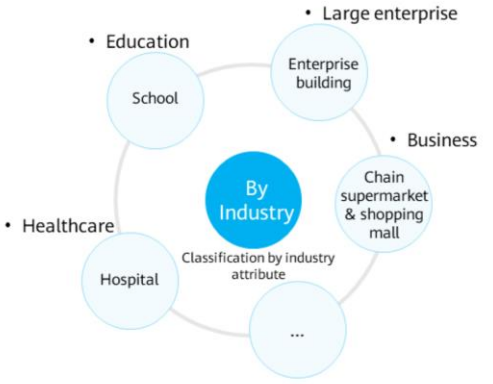
- > 2000 terminal users
- > 100 NEs



- 200 to 2000 terminal users
- 25 to 100 NEs

- < 200 terminal users
- < 25 NEs

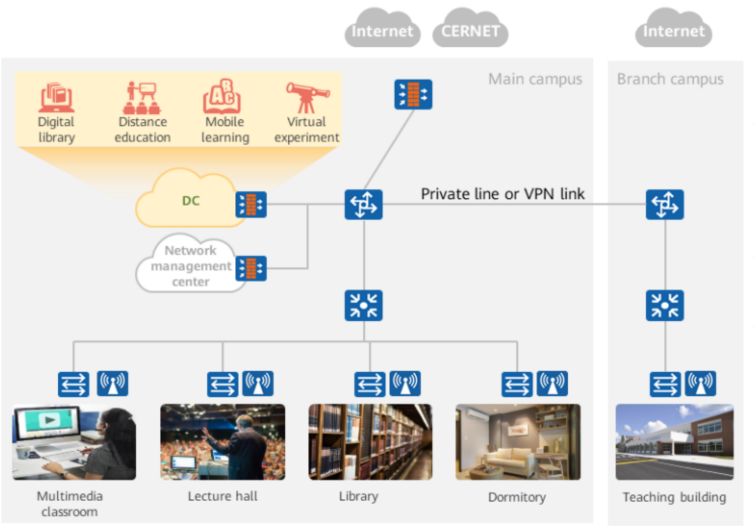
Networks of different scales have different requirements and pain points.



To meet the requirements of different industries, we should design campus network architectures based on the industries they serve. The ultimate goal is to develop campus network solutions with evident industry attributes.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Typical Campus Network Scenario: Higher Education Campus Network



- A school campus network is a computer network that provides teaching, research, and comprehensive information services for teachers and students, as well as family members and visitors.
- A higher education campus network refers to the campus network for universities/colleges.
- A higher education campus network is generally divided into dormitory, living, teaching, and public areas. It provides wired and wireless network access services, taking universities/colleges to the digital era and improving talent cultivation and fostering more innovations.

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- **CERNET** refers to the China Education and Research Network. The CERNET project is funded by the Chinese government and directly managed by the Chinese Ministry of Education. It is a nationwide education and research computer network constructed and operated by Tsinghua University and the other leading Chinese universities.

Characteristics of Campus Networks in the Digital Era

Ubiquitous connectivity



- Anytime, anywhere access
- High-quality service support

On-demand services



- Quick service deployment and adjustment
- Rapid rollout of value-added applications

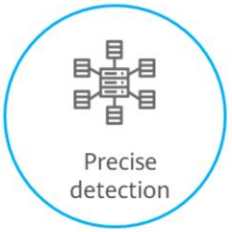
Intelligent and trustworthy



- Automated fault identification and predictive optimization
- Precise threat disposal and proactive threat defense

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Challenges Faced by Campus Network O&M



Traditional O&M is based on SNMP, which collects data in minutes. Once a fault occurs, **it is impossible to obtain data of the fault occurrence time in real time.**

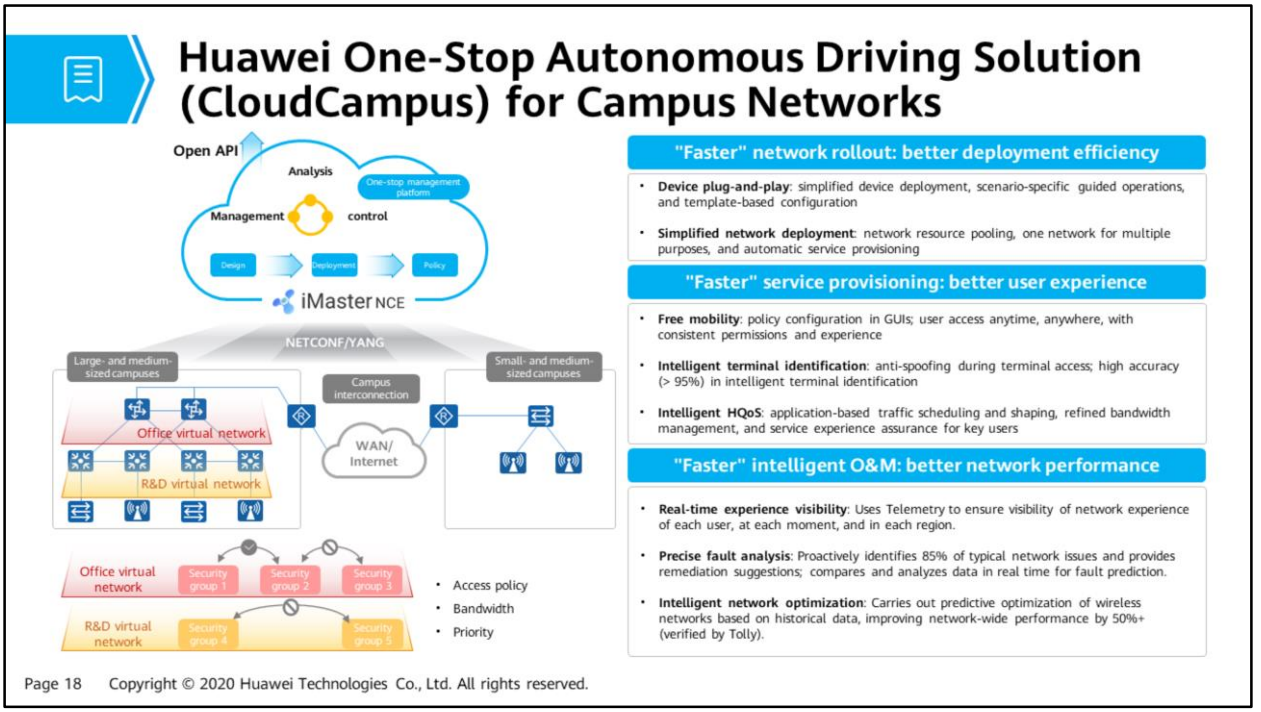


Traditional O&M monitors device indicators only. Although indicators may be normal, user experience is poor. **There is no user and network correlation analysis.**



Traditional O&M cannot identify network issues until users make complaints. **It is impossible to proactively identify and analyze issues.**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- Huawei's CloudCampus solution is dedicated to building an ultra-broadband, intelligent, simplified, secure, and open campus network that aligns with service intents. This new network can provide enterprises with real-time insights into and quick responses to network and service requirements, enabling them to seize transient business opportunities.
- The CloudCampus Solution is a one-stop autonomous driving solution for campus networks.

Solution Component 1: Intent-Driven Campus Network Hardware Products

CloudEngine switches



CloudEngine S12700E: new core for campus networks in the Wi-Fi 6 era

CloudEngine S6730-H: full-featured 10GE routing switch

CloudEngine S5732-H: enhanced GE/multi-GE/hybrid optical-electrical switch

CloudEngine S5735-S: standard gigabit access switch

CloudEngine S5735-S: simplified gigabit access switch

AirEngine Wi-Fi 6 APs



8760-X1-PRO 6760-X1/X1E 5760-51 5760-22W 6760R-51/51E 8760R-X1/X1E

AirEngine 8760-X1-PRO: Wi-Fi 6 indoor flagship AP

AirEngine 6760-X1/X1E: Wi-Fi 6 indoor high-end AP

AirEngine 8760R-X1/X1E: Wi-Fi 6 outdoor AP

AirEngine 9700D-M + 5760-22WD: Wi-Fi 6 agile distributed AP and RU

AirEngine 5760-22W: Wi-Fi 6 wall plate AP



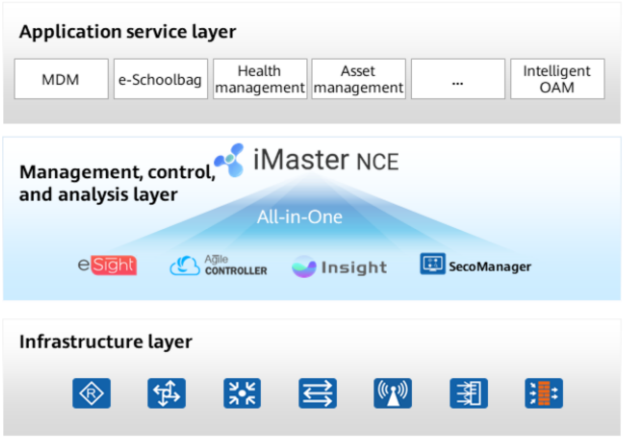
HiSecEngine AI firewalls



NetEngine AR routers

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 2: iMaster NCE-Campus



Automation + intelligence	• SDN-based automated service configuration and deployment • AI-powered intelligent analysis, prediction, and troubleshooting
Management + control + analysis	• Unified data foundation • Fault detection, location, and handling, all at one stop
Planning + construction + maintenance + optimization	• Full lifecycle management • Simulation, verification, monitoring, and optimization

iMaster NCE-Campus: autonomous driving network management and control system for campus networks

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 3: iMaster NCE-CampusInsight

Real-time experience visualization



- **Per region**: Uses a 7-dimensional evaluation system to visually display the network status and user experience of the entire network or each region.
- **Per user**: Displays the full-journey network experience (who, when, to which AP, experience, and issue) of each user in real time, and supports fault tracing.
- **Per application**: Implements real-time voice and video application experience awareness, quick and intelligent demarcation of faulty devices, and poor-QoE root cause analysis

Minute-level fault demarcation



- **Proactive issue identification**: The AI algorithm continuously trained by 200,000+ Huawei terminals proactively identifies 85% of potential network issues.
- **Minute-level fault locating**: The fault inference engine is used to demarcate and identify root causes of faults in minutes and provide effective remediation suggestions.
- **Intelligent fault prediction**: AI is used to dynamically generate baselines based on historical data and predict possible faults by comparing and analyzing the baselines with real-time data.

Intelligent network optimization



- **Real-time simulation feedback**: Evaluates wireless network channel conflicts in real time based on neighbor and RF information of devices deployed on the floor and provides optimization suggestions.
- **Predictive optimization**: Based on historical data analysis, identifies edge APs, predicts AP load trends, performs predictive optimization on wireless networks, and views the gain comparison before and after the optimization. The network performance is improved by 50%+, as verified by Tolly.

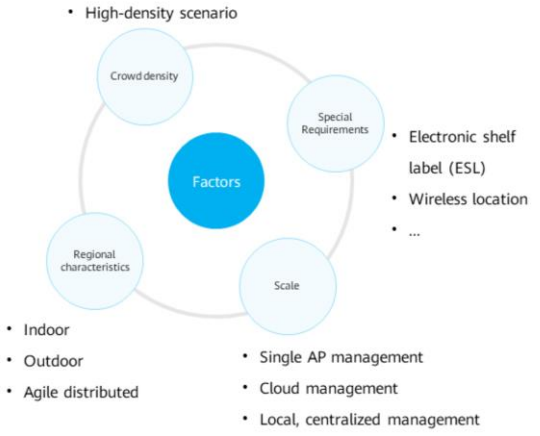
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Enterprise Datacom Network Panorama
2. **Enterprise Datacom Network Solutions**
   - Campus Network
   - **WLAN**
   - DCN
   - Bearer WAN
   - SD-WAN

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

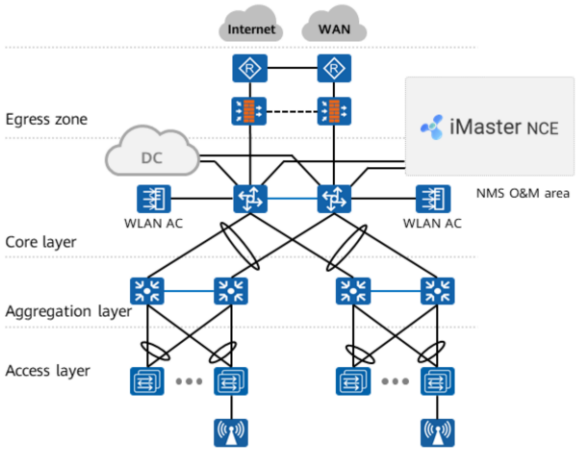
WLAN Scenarios



- **Scale**: The management complexity, requirements, and methods vary depending on the networking scale, which significantly affects the networking.
- **Crowd density**: In high-density scenarios, users are concentrated, which drastically affects the bandwidth per capita and has high requirements on WLAN deployment and planning.
- **Regional characteristics**: In outdoor scenarios, long-distance signal coverage is a typical requirement. This poses special requirements on antennas and high requirements on device reliability and stability.
- **Special requirements**: In the digital era, an increasing number of value-added services (VASs) are popularized. Wireless technologies such as ESL, location, and asset management have higher requirements on WLAN compatibility.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Typical WLAN Networking Scenario: Large Campus (Standalone WLAN AC)

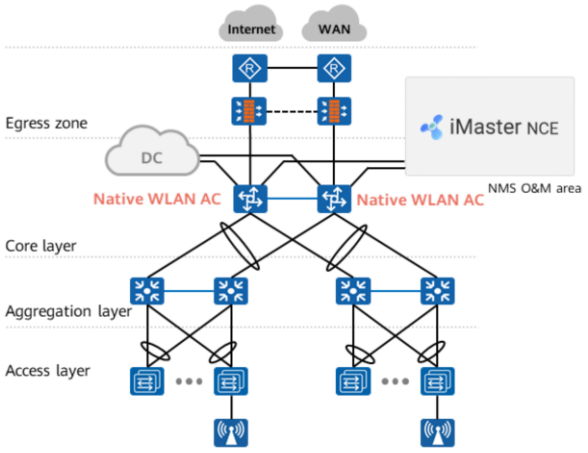


Solution Description

1. If a wired campus network has been deployed and a wireless network needs to be deployed independently, or the wireless network scale is large, you are advised to deploy a standalone WLAN AC.
2. On a large campus network, the WLAN AC is usually connected to the aggregation or core switch in off-line mode.
3. To minimize changes to the existing wired network and facilitate centralized management and control of the WLAN AC, tunnel forwarding is recommended. To improve WLAN AC reliability, VRRP hot standby is typically deployed in the standalone WLAN AC solution.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Typical WLAN Networking Scenario: Large Campus (Wired and Wireless Convergence)

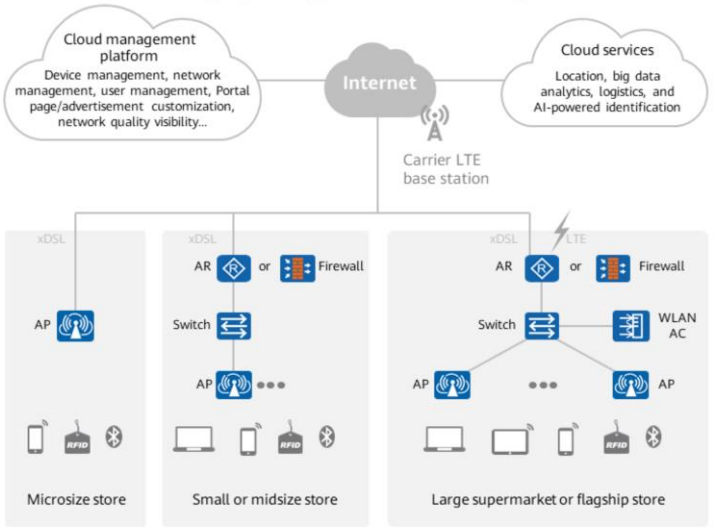


Solution Description

1. To uniformly manage and configure wired and wireless access devices so as to minimize management costs, you are advised to use the native WLAN AC solution.
2. The Ethernet network processor (ENP) series native WLAN AC is the core of the native WLAN AC solution. In this solution, the WLAN AC function is integrated on switches. Special service cards are installed on the switches so that the switches can manage both wired and wireless access devices.
3. The native WLAN AC solution enables wired and wireless users to access the network simultaneously, implementing unified management of wired and wireless users. This solution uses reliability technologies (for example, stacking and Eth-Trunk) of switches to implement device-level and link-level redundancy backup.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

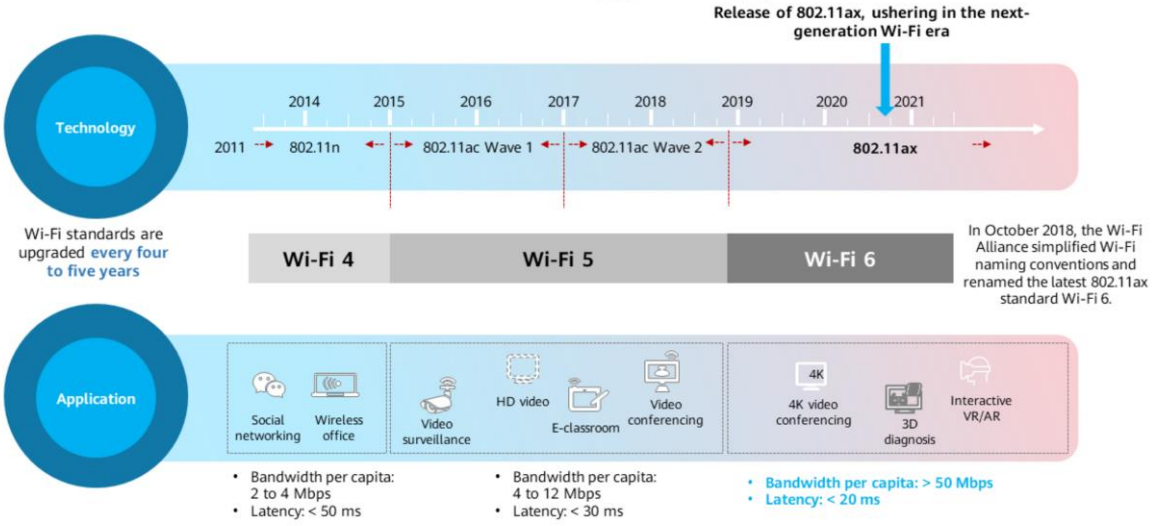
Typical WLAN Networking Scenario: Chain Shopping Mall/Supermarket



- Campus networks of shopping malls and supermarkets enable digital office and digital consumption experience space. For example:
  - Wi-Fi access of guests
  - Digital billboard
  - Smart shopping guide
  - ESL
- Market competition and consumer demand evolution drive digital transformation of offline retail services. In the form of "new retail", chain stores and supermarkets focus on:
  - Shopping experience
  - Operational efficiency

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

The Wi-Fi 6 Era Has Come...
Two Drivers: Technology + Application



Wi-Fi standards are upgraded **every four to five years**

In October 2018, the Wi-Fi Alliance simplified Wi-Fi naming conventions and renamed the latest 802.11ax standard Wi-Fi 6.

- Bandwidth per capita: 2 to 4 Mbps
- Latency: < 50 ms

- Bandwidth per capita: 4 to 12 Mbps
- Latency: < 30 ms

- **Bandwidth per capita: > 50 Mbps**
- **Latency: < 20 ms**

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Wi-Fi 6 Enables Higher Performance

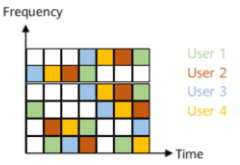
High Bandwidth



1024-QAM
8x8 MU-MIMO

Speed: up to 9.6 Gbps
4x higher bandwidth

High Concurrency

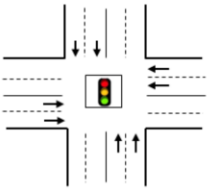


UL/DL OFDMA

UL/DL MU-MIMO

- 1024 STAs per AP
- 4x higher user concurrency

Low Latency



OFDMA

Spatial reuse

- Service latency: 20 ms
- Average latency: 30%↓

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

3 Challenges Facing WLAN O&M



Challenge 1: Proactive issue identification

- Network faults can be detected only after user complaints are received.
- Potential issues that affect user experience cannot be identified.

Challenge 3: User experience measurement

- Poor user experience while device indicators are normal
- Lack of user and network association analysis





Challenge 2: Issue locating and analysis

- No convenient fault tracing method after a network fault occurs
- Lack of key data generated when a fault occurs

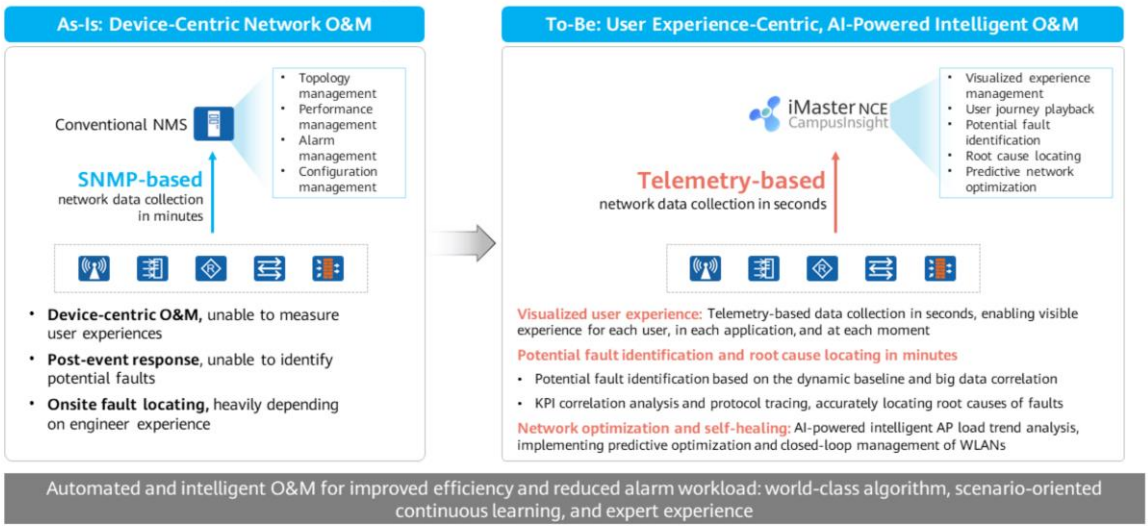
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 1: AirEngine Wi-Fi 6 Diverse-Supply New Products



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 2: iMaster NCE-CampusInsight



Automated and intelligent O&M for improved efficiency and reduced alarm workload: world-class algorithm, scenario-oriented continuous learning, and expert experience

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

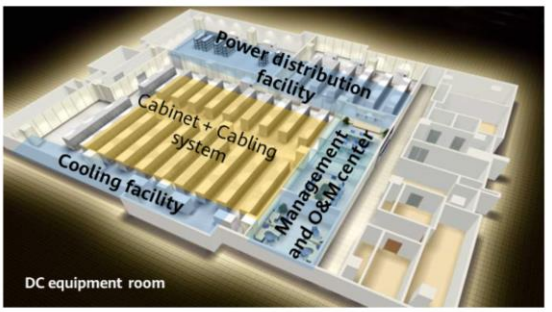
1. Enterprise Datacom Network Panorama
2. **Enterprise Datacom Network Solutions**
   - Campus Network
   - WLAN
   - **DCN**
   - Bearer WAN
   - SD-WAN

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

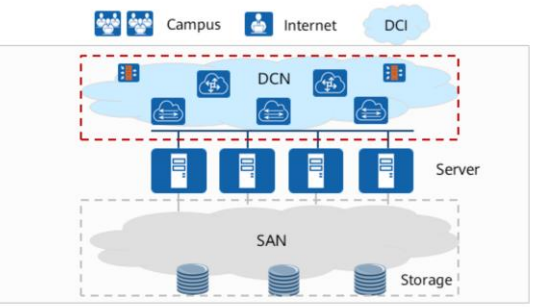
Introduction to a DC

- A DC is a complex set of facilities, including equipment rooms. It not only includes the computer system and devices related to it (for example, communication and storage system), but also contains the redundant data communication connections, environmental control device, monitoring devices, and various security devices.
- DCs use the hierarchical architecture. Typically, L1 refers to the DC infrastructure, and L2 refers to the ICT equipment in the DC.

DC L1 view (infrastructure)



DC L2 view (ICT device)



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Components in a DC

DC devices are classified into computing devices, storage devices, and network devices (excluding the DC infrastructure).

1 Computing device

- Server



- FusionServer Pro intelligent server
- TaiShan server
- Atlas AI computing platform

2 Storage device

- Storage array
- Cloud storage



- OceanStor Dorado all-flash storage
- OceanStor V5 hybrid flash storage
- FusionStorage cloud storage

3 Network device

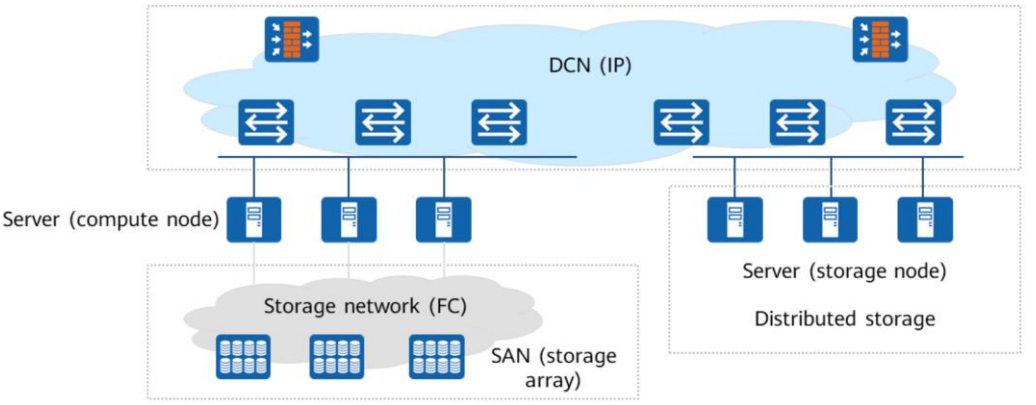
- Router
- Switch
- Firewall
- Load balancer
- IDS/IPS
- • •

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- This slide uses Huawei DC products as an example.

Typical DC Networking

A typical DC networking includes a DCN, a storage area network (SAN), and servers.

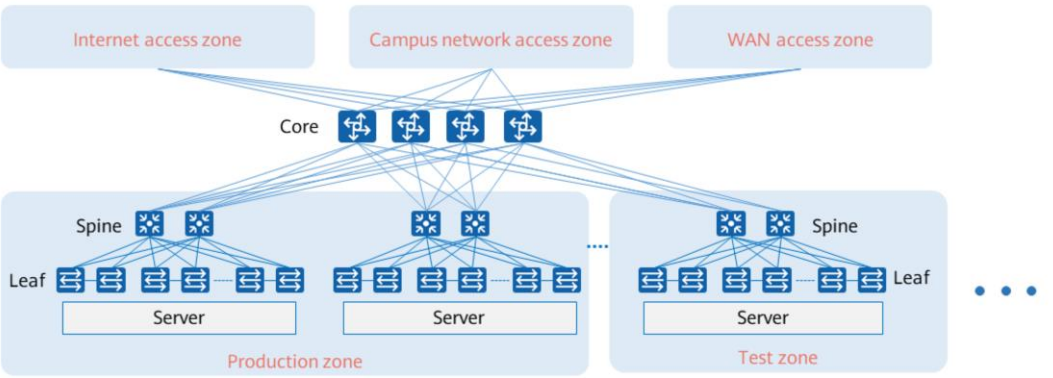


 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- DCM: provides interconnection between computing units in a DC and between computing units in a DC and egresses.
- SAN: consists of storage arrays and FC switches and provides block storage. The storage network that uses the FC protocol is called FC SAN, and the storage network that uses the IP protocol is called IP SAN.
- Distributed storage: The deployment mode of distributed storage is different from that of disk arrays. Data is stored on multiple independent servers (storage nodes) in a distributed manner. It is also used for cloud storage.
- Server (compute node): provides computing services.

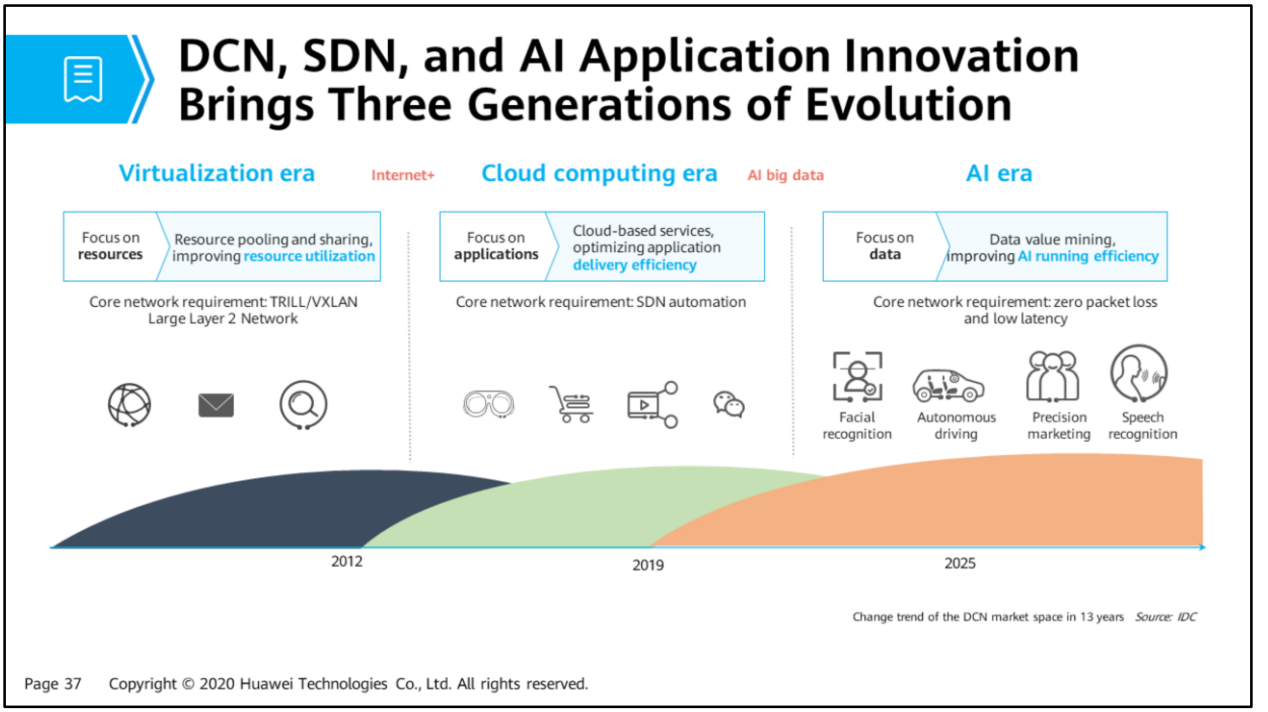
DCN

- A DCN provides interconnection within a DC and interconnection with internal and external egresses of a DC.
- A DCN consists of a series of network devices and is divided into different zones based on service functions.



 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

- There is no fixed zone division for DCNs. Enterprises in different industries are divided into different zones. For example, a financial DC is divided into production zone 1, production zone 2, test zone 1, test zone 2, big data zone, and operation and management zone.
- In this example:
  - Internet access zone: is used to transmit traffic of access to the Internet.
  - Campus network access zone: is used to transmit traffic of access to the enterprise campus network.
  - WAN access zone: connects to the WAN built by an enterprise. Remote zones include DCs and campuses in other cities.
  - Production zone: connects to the production network.
  - Test zone: connects the test network.



- The AI era focuses on data and mines data value to improve AI running efficiency, so AI requires low latency of DCNs.

New Service Characteristics Pose New Challenges to Networks

Rapid construction of ultra-large DCNs
Requiring quick network construction

Improving AI computing power through container-based GPU invoking
Requiring association between containers and networks for rapid login and logout

Many network changes caused by frequent service changes
Requiring intelligent network evaluation and verification

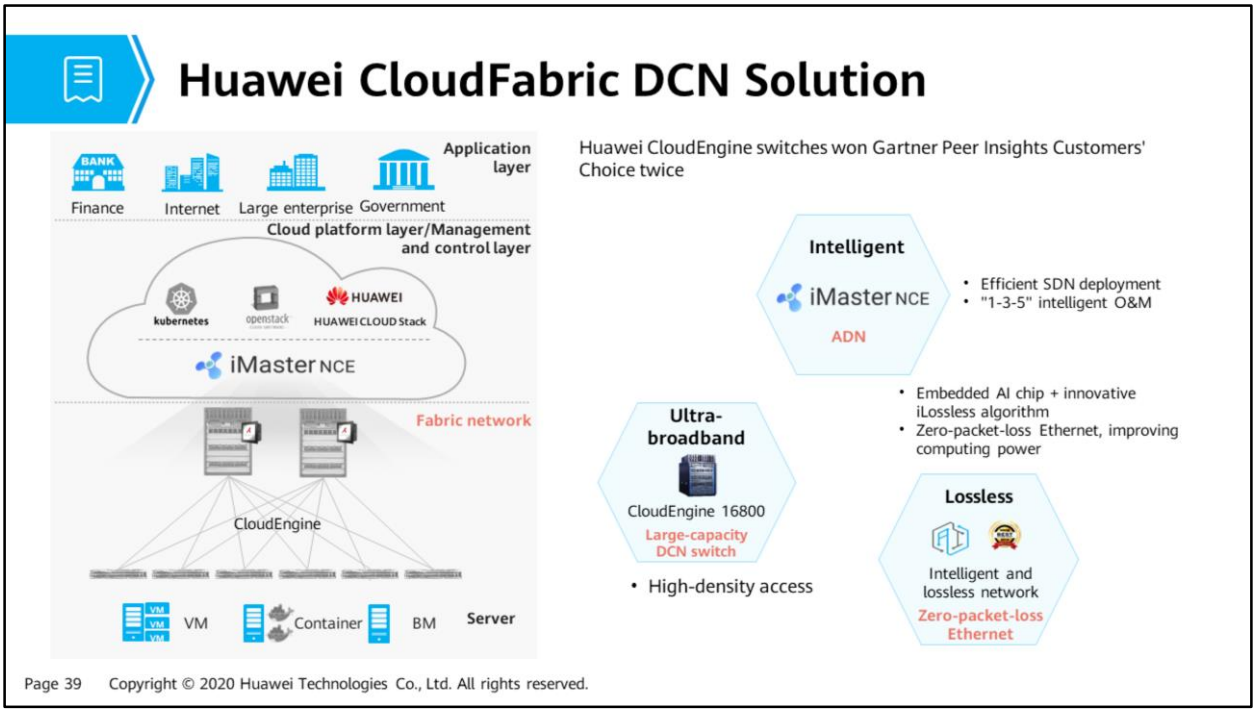






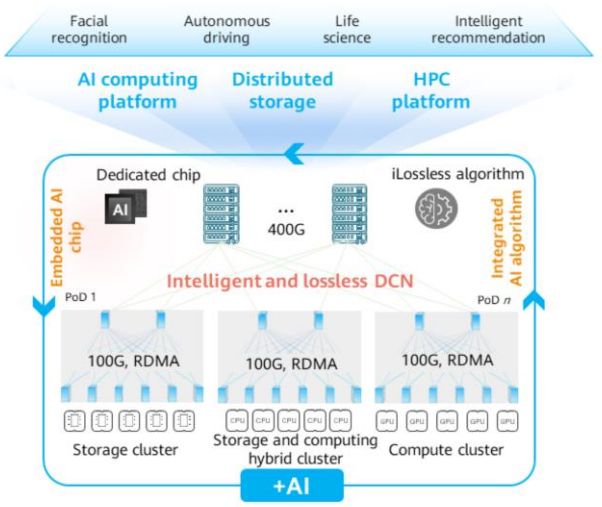


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



- * 1-3-5: Faults can be detected within 1 minute, located within 3 minutes, and rectified within 5 minutes.

Huawei Intelligent and Lossless DCN



Full convergence, three-network integration in a DC

NIC: RoCE iNIC integration

Network: Integration of computing, storage, and data networks, reducing the TCO

Zero packet loss, accelerating RDMA communication

Computing: Improving AI training efficiency

Storage: Improving IOPS performance of distributed storage

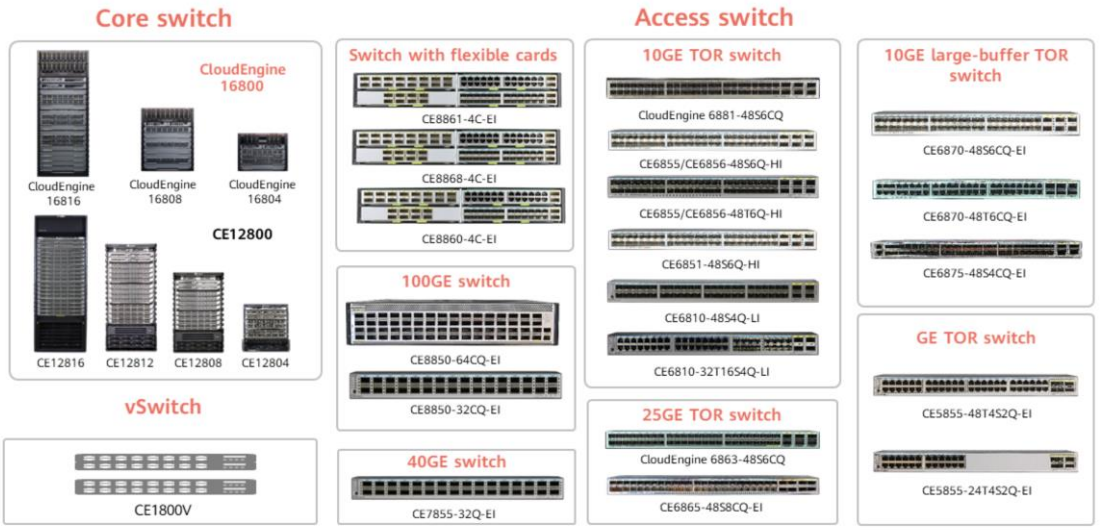
High bandwidth and 400GE network evolution

Bandwidth: 25GE to 400GE

Scale: Small-scale to large-scale

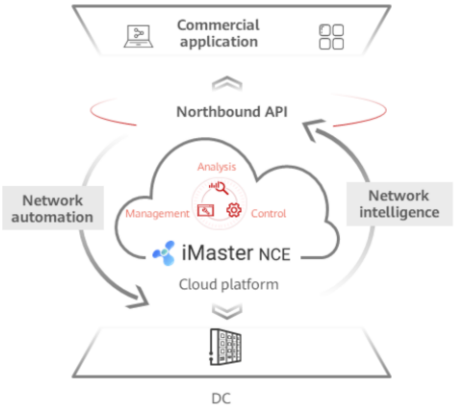
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 1: Huawei DC Switches



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 2: iMaster NCE-Fabric



- Provides full lifecycle DCN management and simplified and automated management experience.
- Abstracts network resources and services in the northbound direction and adapts to various devices and networks in the southbound direction.

Zero-waiting deployment through E2E automated network deployment

- Ultra-high-speed network provisioning: GUI-based drag-and-drop operations
- Fast container rollout: 10K containers per minute

Zero-error configuration through pre-evaluation of change risks

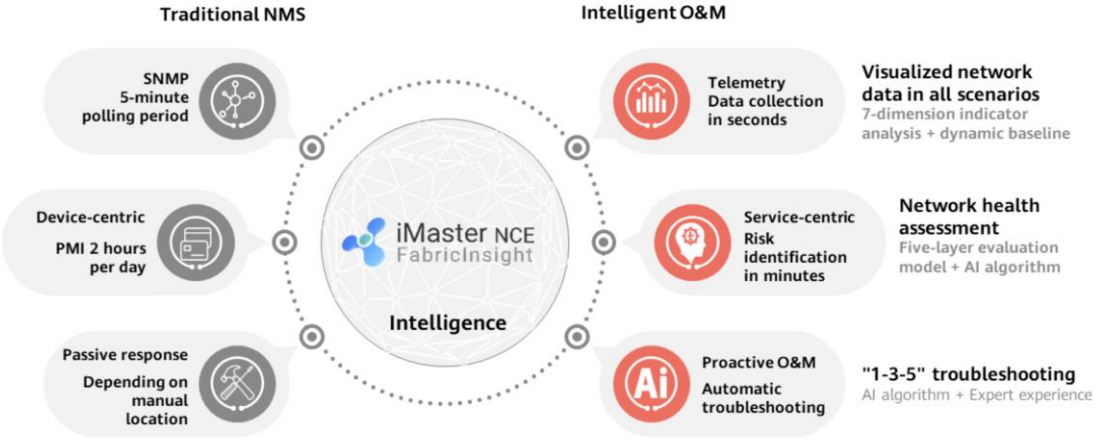
- Pre-event simulation: Before network deployment, use formal verification algorithms to simulate the live network configuration plane, evaluate the impact of the configuration changes on the network, and evaluate whether live network resources are sufficient.
- Post-event verification: Verify the connectivity, interfaces, and routes of the underlay network on the configuration plane.

Zero-interruption through intelligent recovery of typical faults

- "1-3-5" troubleshooting: Typical faults of 75 categories are detected within 1 minute, located within 3 minutes, and rectified within 5 minutes.
- Network health: The network health is comprehensively evaluated based on service experience to proactively predict potential faults.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 3: iMaster NCE-FabricInsight



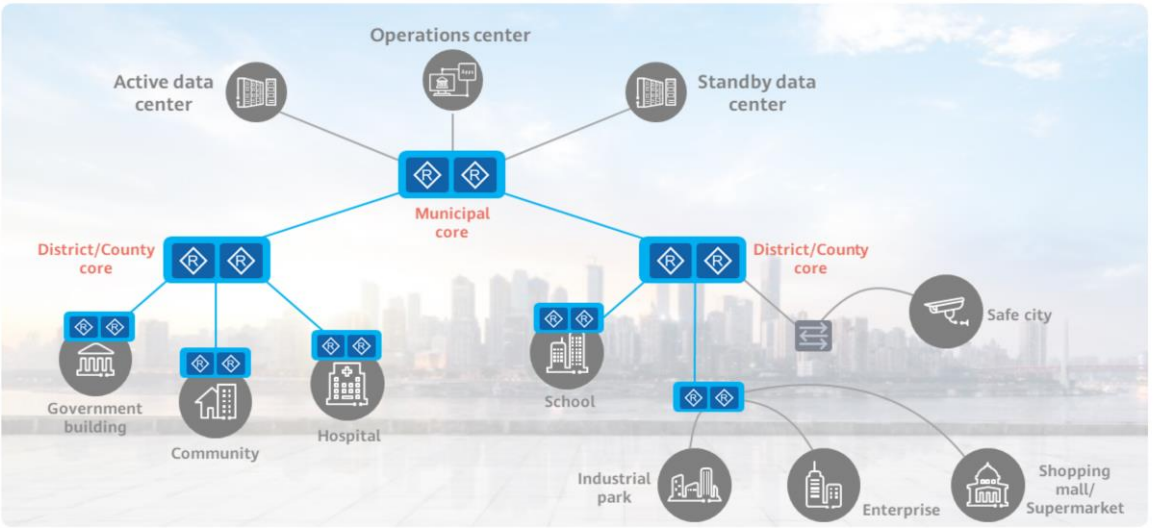
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Enterprise Datacom Network Panorama
2. **Enterprise Datacom Network Solutions**
   - Campus Network
   - WLAN
   - DCN
   - **Bearer WAN**
   - SD-WAN

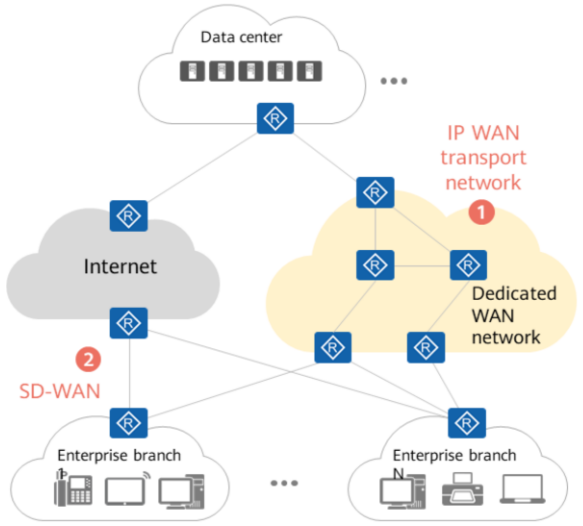
 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

WAN from a Perspective of Smart City Communications Networks



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

IP WAN Transport Network and Enterprise WAN



IP WAN transport network

Scenario description:
- It implements long-haul interconnection between sites and provides dedicated, reliable, ultra-broadband, and QoS-guaranteed private line services.

Typical scenarios:
- Education metro network: implements interconnection between colleges.
- Financial WAN backbone: implements interconnection between branches and data centers.

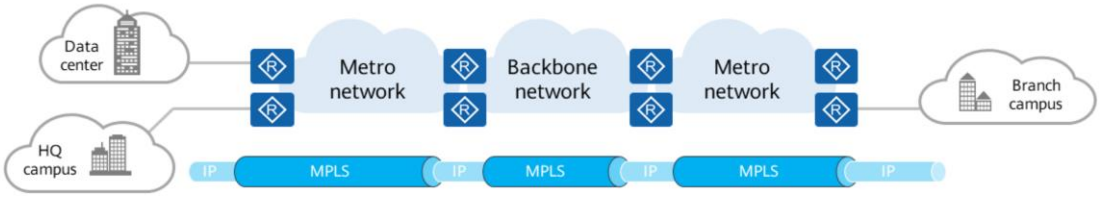
SD-WAN

Scenario description:
- It provides WAN access for enterprise branches, flexibly defines network service models based on communication requirements between branches or between branches and data centers, ensures key service experience, and properly uses WAN egress resources.

Typical scenario:
- WAN access for financial branches.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Traditional IP WAN Transport Network



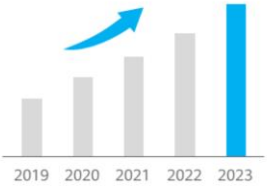
- A WAN is a long-haul communication network that connects LANs or MANs in different areas.
- In terms of space, WANs are usually deployed across regions. The construction of WAN transport networks is difficult, costly, and time-consuming.
- Service deployment efficiency, network congestion, and latency are important factors that need to be considered for WAN transport networks.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Challenges Faced by Traditional IP WAN Transport Networks

Rapidly growing network traffic

WAN traffic increasing rapidly year by year



How can large-capacity and cost-effective converged transport networks be built?

Differentiated connections for various industries

Smart finance, smart grid, smart city



How can differentiated connection services with guaranteed SLAs be provided?

Automatic and intelligent O&M

Service quality visualization, rapid troubleshooting



How can proactive O&M be implemented, towards autonomous driving network?

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Huawei IP WAN Transport Network Solution

Scenario

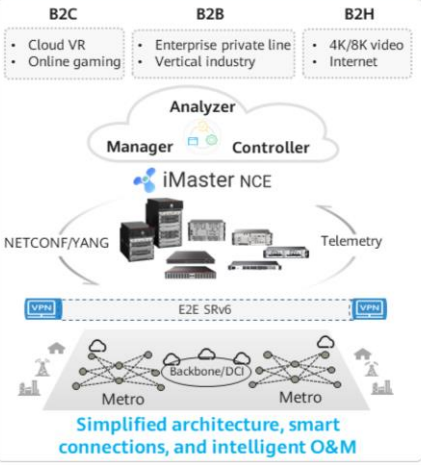
Government
- National backbone network
- e-Government converged transport network

Electric power
- SDH-to-IP reconstruction
- PCM bearer

Finance
- One-hop connection to the cloud, fast service provisioning
- Inter-DC traffic optimization

ISP
- Metro network reconstruction
- User access, Layer 3 to edge

Target Network Architecture



Simplified architecture, smart connections, and intelligent O&M

Solution

New platform Large capacity, all-service NetEngine 8000 series



- 14.4T/slot, industry's highest-density 400G
- Unified platform for all scenarios and 5-in-1 (broadband service/private line/data center egress/international egress/BNG)

New protocol SLAs guaranteed, and latency committed

SRv6

- TTM shortened and latency committed

New pipe FlexE-based slicing, guaranteeing bandwidth



- Zero preemption between FlexE slices, guaranteeing bandwidth
- Large-network slicing granularity (1G v.s. 5G)

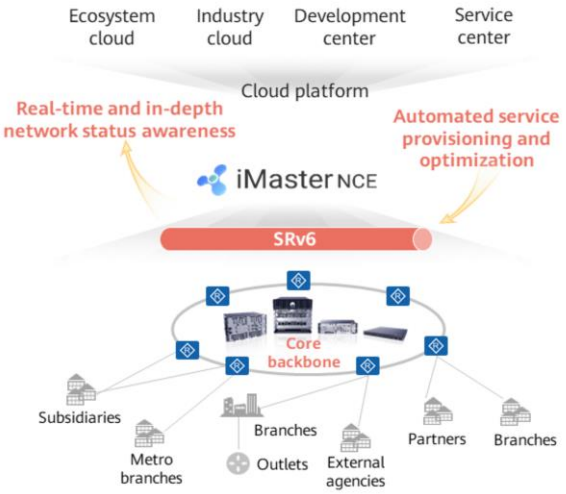
New O&M AI-powered intelligent O&M, visualized service quality

iMaster NCE

- iMaster NCE+iFIT, hop-by-hop packet loss detection solution
- Ti-LFA protocol, E2E switching within 50 ms

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

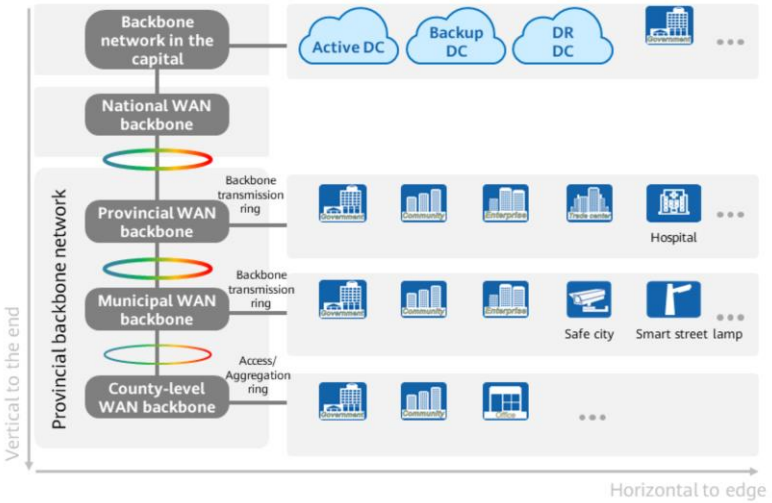
Typical Application Scenarios of the Bearer WAN Solution: Finance



1. **Rapid service rollout:** Services can be quickly provisioned across the HQ and branches. SRv6 can be automatically provisioned within minutes.
2. **AI-based intelligent O&M:** iMaster NCE+SRv6 identifies applications and tenants and implements intelligent traffic steering based on conditions such as latency and bandwidth to guarantee SLAs.
3. **Dynamic traffic optimization:** iMaster NCE dynamically optimizes DCI network traffic, improving the bandwidth utilization of links between multi-active DCs.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

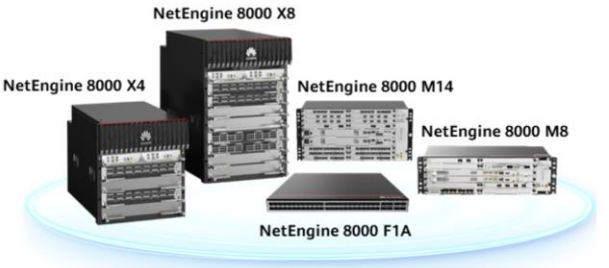
Typical Application Scenarios of the Bearer WAN Solution: e-Government



1. **SRv6/EVPN:** The e-Government network covers a wide range and is large in scale. SRv6 simplifies network protocols, service provisioning, and O&M.
2. **Network slicing:** FlexE-based slicing implements hard isolation between different government departments, ensuring service quality and security for powerful departments.
3. **Full-series service routers:** They support a high capacity of 14.4 Tbps/slot and GE/10GE/100GE full-rate interfaces.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 1: NetEngine 8000



New platform

14.4T/slot

Multi-service transport, supporting SR, PE, DC gateway, and BRAS

New protocol

Full NetEngine 8000 series support for SRv6

Cross-domain seamless connection and application-level SLA assurance

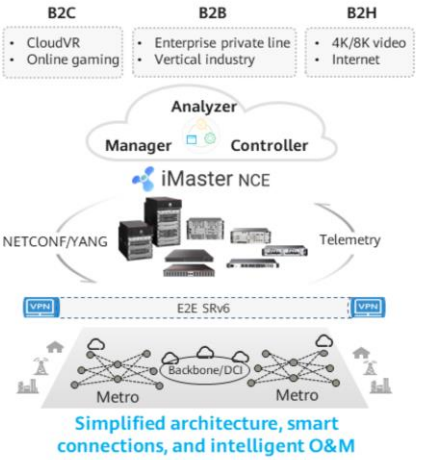
New O&M

AI-boosted intelligent O&M

Automatic service provisioning, real-time SLA awareness, and minute-level fault locating

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 2: iMaster NCE-IP



Simplified architecture, smart connections, and intelligent O&M

Management

- **NE management:** topology, alarms, configuration, and inventory
- **Management of traditional services:** static L2VPN and static L3VPN

Control

- **Centralized path computation:** centralized path computation based on multi-dimensional constraints
- **Logical topology:** cost, latency, bandwidth, and usage
- **Network optimization:** PCEP- and IP-based traffic adjustment and optimization

Analytic

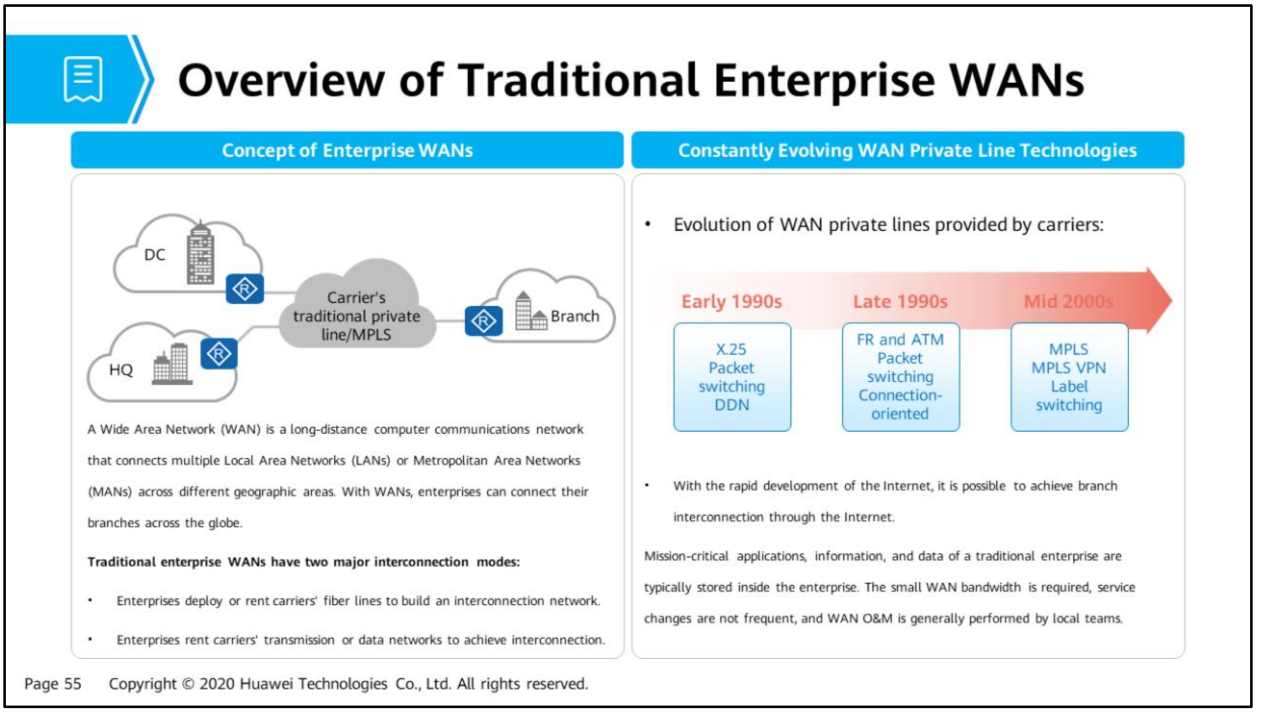
- **Basic network analysis:** performance, traffic, and quality presentation and analysis
- **Prediction and analysis:** traffic, fault, and exception prediction

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Contents

1. Enterprise Datacom Network Panorama
2. **Enterprise Datacom Network Solutions**
   - Campus Network
   - WLAN
   - DCN
   - Bearer WAN
   - **SD-WAN**

 Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



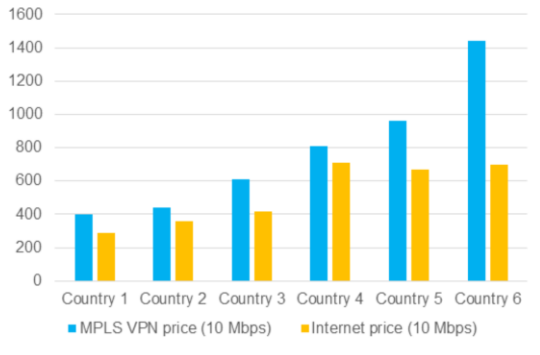
- A Wide Area Network (WAN), generally provided by a carrier, is a long-distance computer communications network that connects multiple Local Area Networks (LANs) or Metropolitan Area Networks (MANs) across different geographic areas. A typical WAN covers distances of tens to tens of thousands of kilometers. It spans a large geographic area such as across cities, regions, or countries. Through WANs, enterprises can set up an interconnection network for their branches worldwide, facilitating their daily operations.
- Traditional enterprise WANs have two major interconnection modes:
  - The first mode is that enterprises deploy or rent carriers' fiber lines to build an interconnection network.
  - In the second mode, enterprises rent carriers' transmission or data networks to achieve interconnection.
- Generally, only enterprises with strong financial strengths prefer the first mode. Most enterprises tend to use the second mode, that is, rent the lines or networks provided by carriers to build their own WANs. With the rapid development of the Internet, it is possible to achieve branch interconnection through the Internet. However, the Internet has certain weaknesses, such as low reliability and a lack of end-to-end quality assurance. That's why large enterprises generally do not simply rely on the Internet to construct WANs for multi-branch interconnections. The Internet is often used as a remote access mode for traveling employees or as a backup solution for branch interconnections. Only some non-mission-critical services are carried via the Internet.
- Enterprises use the WANs provided by carriers to interconnect their branches, headquarters, and data centers (DCs) across different geographic areas. Mission-critical applications, information, and data of a traditional enterprise are typically stored inside the enterprise. The small WAN bandwidth is required, service changes are not frequent, and WAN O&M is generally performed by local teams. This traditional enterprise WAN architecture has long been playing an important role in enterprise branch

interconnections and meets the service requirements of enterprise users.

Challenges Facing Traditional Enterprise WANs

High Private Line Costs

Sharp increase of WAN traffic vs. high price of traditional private lines

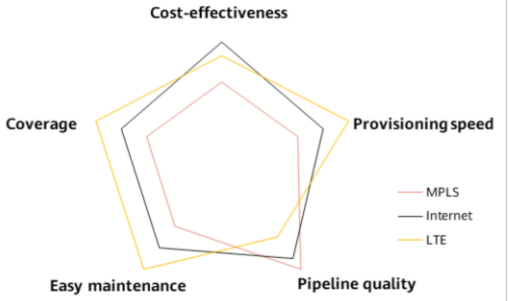


Source: TeleGeography © 2017 PriMetrica, Inc.

Long Private Line Construction Period

Time-consuming private line construction and huge difficulties obtaining traditional private lines

(Apply for private lines -> Allocate resources -> Conduct construction)

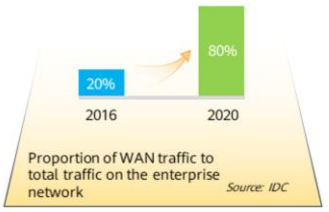


Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Development of Enterprise WAN Technologies

Migration of Enterprise Services to the Cloud

Enterprise services are migrated to the cloud, and enterprise egress traffic increases sharply.





Intelligence Cloudification Video Mobility

Emergence of Software Defined Networking (SDN)

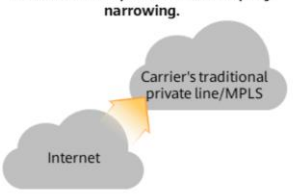
Unified management and centralized configuration



SDN introduces a new role — the centralized controller, which can not only understand the administrator's requirements on the network, but also fully manage and configure the physical network, implementing application-oriented automatic deployment and quick provisioning of the network.

High-Speed Development of the Internet

The gap between the quality of the Internet and traditional private lines is rapidly narrowing.

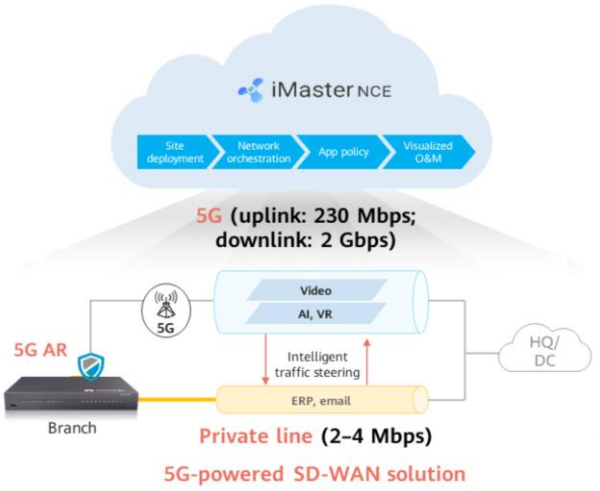


The coverage and network performance of Internet are continuously improving, and the gap between the quality of the Internet and traditional private lines is narrowing rapidly. As such, more and more enterprises and organizations are using the Internet for interconnection.

With the rapid development of enterprise WAN-related businesses and technologies, the new Software-Defined WAN (SD-WAN) solution emerges.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Huawei Intent-Driven SD-WAN: Accelerating Cloud Transformation of Enterprise Services



5G uplink: full lineup of 5G-capable NetEngine AR6000s

Large bandwidth: uplink: 230 Mbps; downlink: 2 Gbps
All network generations: 5G, 4G, and 3G
Dual architectures: non-standalone (NSA) and standalone (SA)

High performance: congestion-free forwarding

The forwarding architecture meets SD-WAN development requirements in the next five years.

High-quality network: Flexible, reliable, and secure enterprise interconnection

Forwarding-control separation architecture, and on-demand orchestration of 20+ networking models
Enhanced proactive defense capabilities of CPEs, offering E2E security

Optimal experience: application-based intelligent traffic steering, ensuring experience of key applications

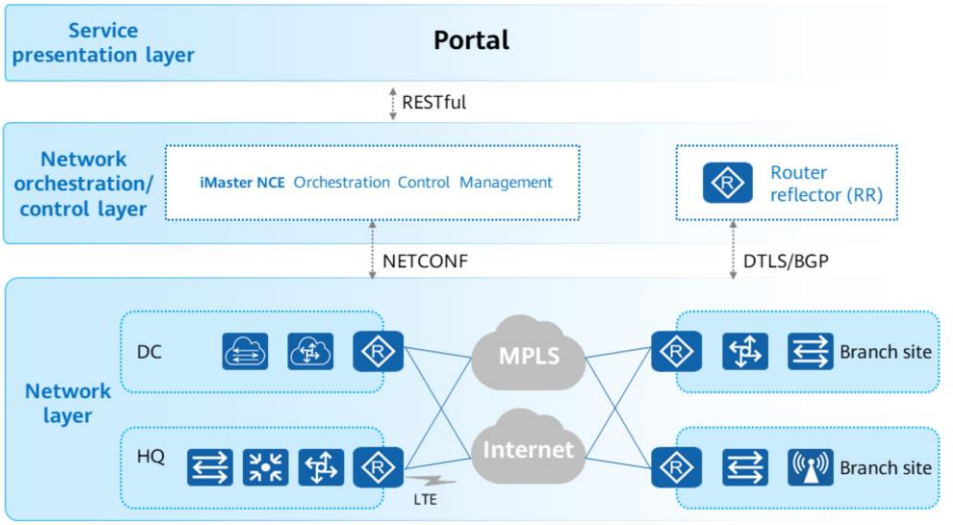
Application-based intelligent traffic steering: on-demand scheduling via 5G or fiber
Adaptive forward error correction (A-FEC): no freeze frame or artifact even at 20% packet loss

Simplified O&M: full-process automation and plug-and-play

Multiple ZTP modes: deployment of branch networks in minutes
Visualized application, branch, device, and link status: centralized management and simplified O&M

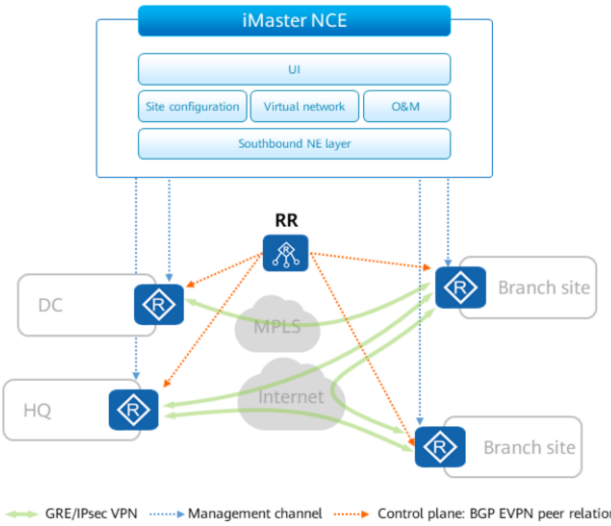
Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Overall Architecture of the SD-WAN Solution



Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

SD-WAN — Powered EVPN Interconnection Solution

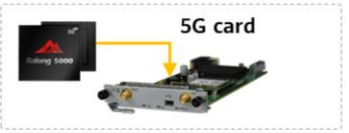


- The WAN is abstracted and modeled to decouple upper-layer network services from lower-layer networks, implementing network automation.
- An independent control plane is deployed to separate network forwarding and control planes, thereby implementing centralized network control.
- Centralized network monitoring and visualization enable centralized management of WANs in an E2E manner, implementing intelligent O&M.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 1: Full Lineup of NetEngine AR Routers

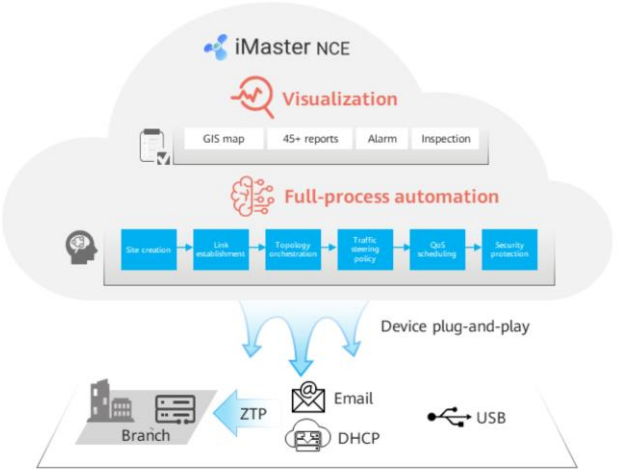
Full lineup of 5G-capable NetEngine AR6000s: Build 5G high-speed network egresses for branches of enterprises of all sizes



NetEngine AR6120 series	NetEngine AR6140 series	NetEngine AR6280 series	NetEngine AR6300 series
			
10GE uplinks & multi-service	Extensive interfaces, dual power supplies	3x↑ performance, dual power supplies	3x↑ performance, dual SRUs, and dual power supplies
Small/Midsize branch	*Midsize branch*	*HQ/Large branch*	*HQ/Large branch*

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.

Solution Component 2: iMaster NCE-WAN



Visualized O&M

- GIS map-based network monitoring and visualized application, link, site, and network status
- Automatic network inspection and precise alarm notification by email

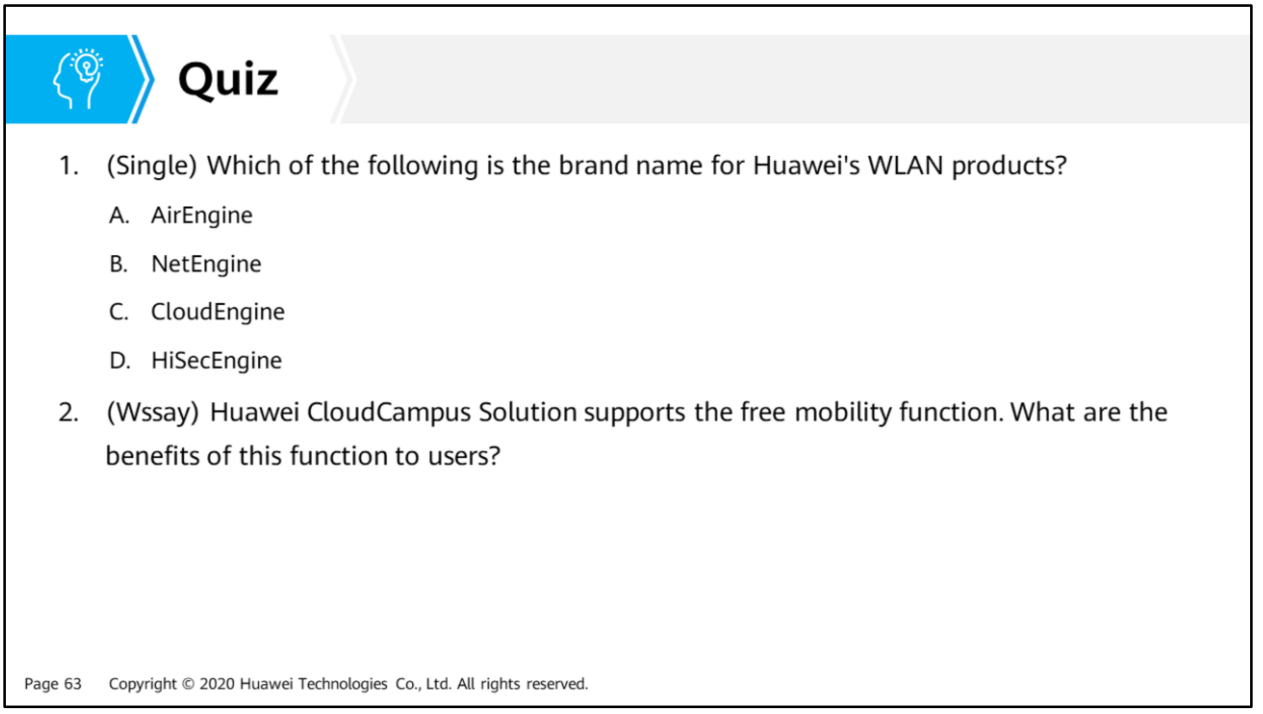
Full-process automation

- Template- and process-based site and network topology configuration
- Application-centric routing, QoS, and security policies

ZTP in multiple scenarios

- Various ZTP deployment modes and device plug-and-play

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.



1. A
2. Free mobility enables GUI-based policy configuration and allows users to access the network anytime and anywhere, with consistent roaming permissions and experience.

Summary

- This course systematically introduces the panorama of datacom networks in the digital world.
- After completing this course, you will be able to understand the concepts and significance of the campus network, WLAN, data center network, and WAN, as well as the basic architecture and applications of the preceding networks in typical industries.
- After completing this course, you will also be able to systematically understand Huawei's enterprise datacom solutions, including the campus network solution, WLAN solution, data center network solution, Bearer WAN solution, and SD-WAN solution.

Copyright © 2020 Huawei Technologies Co., Ltd. All rights reserved.


Thank You
www.huawei.com